TRAVEL PREDICTION WITH MODELS OF INDIVIDUAL CHOICE BEHAVIOR

by

FRANK SANFORD KOPPELMAN

S.B., Massachusetts Institute of Technology
(1959)
M.B.A., Harvard University , Graduate School of Business Administration
(1961)

Submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology
(June, 1975)

Signature of Author . . . .
Department of Civil Engineering, May 16, 1975

Certified by. . . . .
Thesis Supervisor

Accepted by. . . . .
for the Department of Civil Engineering

ABSTRACT

TRAVEL PREDICTION WITH MODELS OF INDIVIDUAL CHOICE BEHAVIOR

by

FRANK SANFORD KOPPELMAN

Submitted to the Department of Civil Engineering on May 16, 1975, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

This study is concerned with the prediction of travel behavior. Planning process needs require predictions of travel flows at a level of aggregation which is relevant to the policies under study and the precision with which they have been formulated. In contrast with the need for aggregate predictions based on an aggregate description of socio-economic characteristics and the attributes of the proposed transportation system, travel behavior theory is postulated at the level of the individual or other behavioral unit. Group behavior, which is the object of prediction, is determined by the travel choices of the individual behavioral units. Procedures for aggregate prediction should reflect this relationship.

The concept of an aggregated prediction model is developed for this purpose. This model differs from other aggregate prediction models by its explicit dependence on relationships which describe individual choice behavior. The aggregated prediction model contains three components. These are the disaggregate choice model, the distribution of independent variables, and an aggregation procedure. The theoretically correct aggregation procedure is the summing or averaging of individual choice probabilities to obtain aggregate volume or share predictions. The difficulty of providing the data required to evaluate the choice probabilities for the entire prediction group motivates the search for alternative aggregation procedures.

The focus of this study is to identify the errors in prediction contributed by use of these alternative aggregation procedures and to place this error in perspective with other sources of error in prediction.

Theoretical and simulation analyses identify the structural pattern of error introduced by the aggregation procedure. The sensitivity of this error to the distribution of net utilities between pairs of alternatives is described.

The empirical study confirms the significance of the aggregation error. It also indicates the potential for reducing aggregation error by the classification of prediction groups in accordance with differences in available choices. The empirical study also indicates that there is substantial error in other components of the prediction process. These errors are large enough to overwhelm the effect of error due to aggregation in the prediction situation studied.

This study concludes that it is feasible to use disaggregate choice models to predict aggregate travel demand. The required predictions can be based on information about independent variables which is not significantly different from that required by commonly used models. Situations where substantial bias may occur can be avoided by construction of aggregate prediction groups which are reasonably homogeneous. When these siguations cannot be avoided aggregation error may be significantly reduced by classification or by incremental prediction.

Thesis Supervisor: Moshe E. Ben-Akiva

Title: Assistant Professor of Civil Engineering

ACKNOWLEDGEMENTS

I take this opportunity to acknowledge the value of studying and working with the faculty and students of the Transportation Systems Division of the Department of Civil Engineering at M.I.T.

In particular, I wish to acknowledge the invaluable guidance of Professor Marvin L. Manheim, who encouraged me to undertake advanced study in transportation and provided stimulus and support during the years of study and research.

Professor Moshe E. Ben-Akiva, my thesis advisor, contributed in important ways to the development of this research study and to its execution. Professor Ben-Akiva's insights and advice were of great value in helping to refine concepts and resolve important research issues. I particularly wish to thank him for the long and stimulating discussions prior to and during this study.

Professors Paul Roberts, Marvin Manheim, and Wayne Pecknold and Gerald Kraft, members of my thesis committee, contributed extensively to this research through the many ideas, suggestions, criticism and encouragement which they gave me.

I wish to thank Dr. Kirtland C. Mead of the MITRE Corporation, Professor Charles Manski of Carnegie-Mellon University and Professor Antti Talvitie of the University of Oklahoma for their ideas, suggestions and criticisms.

Edward Weiner of the Office of the Secretary and David Gendell of the Federal Highway Administration, U.S. Department of Transportation, offered helpful suggestions on the structuring of the research approach.

To my friends and colleagues at M.I.T., particularly Uzi Landau, Steven Lerman, and Len Sherman, I express my thanks for the hours of discussions during which many key ideas were developed and refined or discarded, as appropriate.

The Metropolitan Washington Council of Governments, R.H. Pratt Associates and Cambridge Systematics are acknowledged for supplying the data used in the empirical study.

Preliminary development of the research was supported under contract DOT-OS-30120 of the U.S. Department of Transportation, Program for University Research for "Experiments to Clarify Priorities in Urban Travel Forecasting Demand and Development", and the State of California Contract 13924 for research in "Community and Environmental Values in Transportation Planning". The primary research was conducted under Contract DOT-OS-50001 with the U.S. Department of Transportation, Office of the Secretary, for "Development of an Aggregate Model of Urbanized Area Travel Behavior".

I also wish to thank the people who helped in the preparation of the final document. Becky Muller undertook the major load and responded with enthusiasm and energy. Marianne Lehners, Patricia Otis, Carol Walb and Imadiel Ariel provided valuable help.

And to my wife, Marianne, my profound gratitude for sharing freely and fully in the strain, sacrifice and satisfaction of these years.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION AND SUMMARY

1.1 Background and Purpose of the Study

A critical element in the transportation planning process is the prediction of the future performance and external impacts of the transportation system for each of the available plan alternatives. Prediction should provide adequate information about each outcome so that the groups concerned can make informed choices among the available alternatives. To be used effectively the predictions obtained must distinguish between different policies with respect to travel flows, system performance and external impacts. Evaluation needs require that these forecasts be at a level of aggregation which is relevant to the policies under study and the precision with which they have been formulated.

In contrast with the need for aggregate predictions based on an aggregate descriptions of the socio-economic characteristics and attributes of proposed transportation 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 determined by the travel choices of the individual behavioral units. Procedures for the prediction of aggregate demand should reflect this relationship. Such aggregation as is necessary may be accomplished either implicitly as is done in the application of models which have been estimated on aggregate data or explicitly as must be done in the use of disaggregate choice models.

A basic requirement of the resultant models is that they attain consistency between the aggregate predictions and the predictions which would be obtained by use of models of individual choice behavior based on a complete description of the characteristics of each behavioral unit and the travel choices available to it. This requirement, establishes a direct linkage between the aggregate prediction model and the underlying model of individual choice behavior.

The purpose of this study is to:

- Investigate the issues involved in predicting aggregate travel demand consistent with disaggregate travel choice behavior, and
- Identify, test and evaluate alternative procedures for predicting aggregate travel demand based on models of disaggregate travel choice behavior.
- Describe the structure of errors in prediction contributed by different aggregation procedures and place this error in perspective with other sources of error in prediction.

1.2 Travel Demand Prediction

As a result of the requirements of the 1962 Federal Aid Highway Act and subsequent federal legislation,over 200 urbanized regions in the United States are required to make and periodically update transportation studies and plans. Most of these studies are based on a prediction methodology developed in the 1950's which relates demand for travel to exogenous growth in population, employment and income and attempts to predict the distribution of this travel over alternative future transportation systems (Martin et al, 1961; Bureau of Public Roads, 1970). These models suffer from a variety of weaknesses including:

- Inadequate level of service sensitivity,
- An emphasis on, and bias towards, the prediction of travel on highway links,
- Inability to consider a wide range of non-capital alternatives, and
- High cost of gathering the data required to calibrate the models used.

The structure and orientation of these models limits their usefulness for addressing the wide range of policy issues facing transportation planners and the range of alternatives which are being considered to address these issues (Brand,1972; Manheim,1972; Stopher & Lisco,1970).

Recently, attempts have been made to overcome these limitations by analysis of disaggregate - individual or household - data. Disaggregate analysis relates the observed choice of the behavioral unit* to its own characteristics and the attributes of choices which it faces. Because they do not contain any aggregation scheme, these models can be used at different levels of aggregation, in different places and at different times.

Disaggregate models were initially developed as research tools. The primary objective of these analyses of disaggregate choice behavior was to improve the understanding of travel behavior. Brand (1973) pointed out the need to base travel models on a "well specified structural or behavioral representation of the decision process." He also argued that there was "a way to go in getting models based on individual choice behavior into the field."

The gap between the need for aggregate forecasts and the models of individual travel behavior can be bridged by an explicit aggregation procedure. To assure behavioral consistency, such a procedure must be designed so that aggregate predictions of travel behavior are based on the structure of the disaggregate choice model, and aggregate forecasts of the distribution of independent variables.

Conceptually, there are two alternative approaches to the development of aggregate prediction models which are consistent with underlying

---

*The behavioral unit is the social unit which makes the travel decision. Depending on the choice involved, the behavioral unit may be an individual, a household or some other group engaged in cooperative decision making.

disaggregate behavior. The first approach is to aggregate the model structure itself to obtain a consistent aggregate structure which can be estimated at the aggregate level. When the underlying choice model is non-linear and the aggregate groups are not homogeneous, a consistent aggregated function will include parameters of the choice model and of the distribution of independent variables. The requirements on the structure of the underlying choice model and the distribution of variables from which a consistent estimable aggregate relationship can be derived are extremely restrictive. To date this has been accomplished in the case of a binary probit choice function with a multivariate normal distribution of independent variables (McFadden and Reid, 1975).*

The second approach is to explicitly aggregate an estimated disaggregate choice model to obtain aggregate predictions of travel demand. If the characteristics of every individual or household and all the alternatives which they face could be predicted, it would be conceptually simple to make probabilistic predictions of choice behavior for each individual and aggregate these predictions over the prediction group to obtain the aggregate travel demand prediction. Generally, this disaggregate information is not available. Thus, the requirement for the aggregated model is to predict aggregate travel behavior based on

---

*Although this approach provides a mechanism for estimating a behaviorally consistent model when only aggregate data are available, it is preferable in terms of data collection and estimation efficiency to estimate the corresponding disaggregate function and predict with an aggregated structure as also suggested by the same authors.

an aggregate description of population and transportation system characteristics.

The advantages which can be gained by use of disaggregate choice models as a basis for the prediction of aggregate demand are significant. Among these are:

- Greater predictive validity,
- Applicability to a wider range of areas, over a longer time span,
- Improved sensitivity to changes in transportation services characteristics,
- Applicability at different levels of aggregation, and
- Increased efficiency in the use of data.

For all of these reasons aggregate models should be based on the analysis of individual travel choice behavior.

The analysis to be undertaken explicitly investigates the issues involved in using disaggregate choice models for the prediction of aggregate travel behavior. Procedures for obtaining aggregate predictions based on disaggregate choice models are described. These procedures are evaluated in terms of the "aggregation error" which they introduce to the aggregate prediction.

1.3 Structure of Aggregated Prediction Models

Any model which predicts the travel behavior of groups of behavioral units may be described as an aggregate prediction model. The central characteristic of such models is their ability to predict travel demand at some level of aggregation based on input variables which are also aggregate, usually at the same level as the desired predictions.

The definitional distinction of an aggregated travel demand model is its basis in disaggregate choice behavior explicitly aggregated to the desired level of prediction. The aggregated model consists of (Figure 1.1):

- A disaggregate travel choice model,
- The distribution of independent variables, and
- An aggregation procedure which operates on these two elements to obtain aggregate predictions.

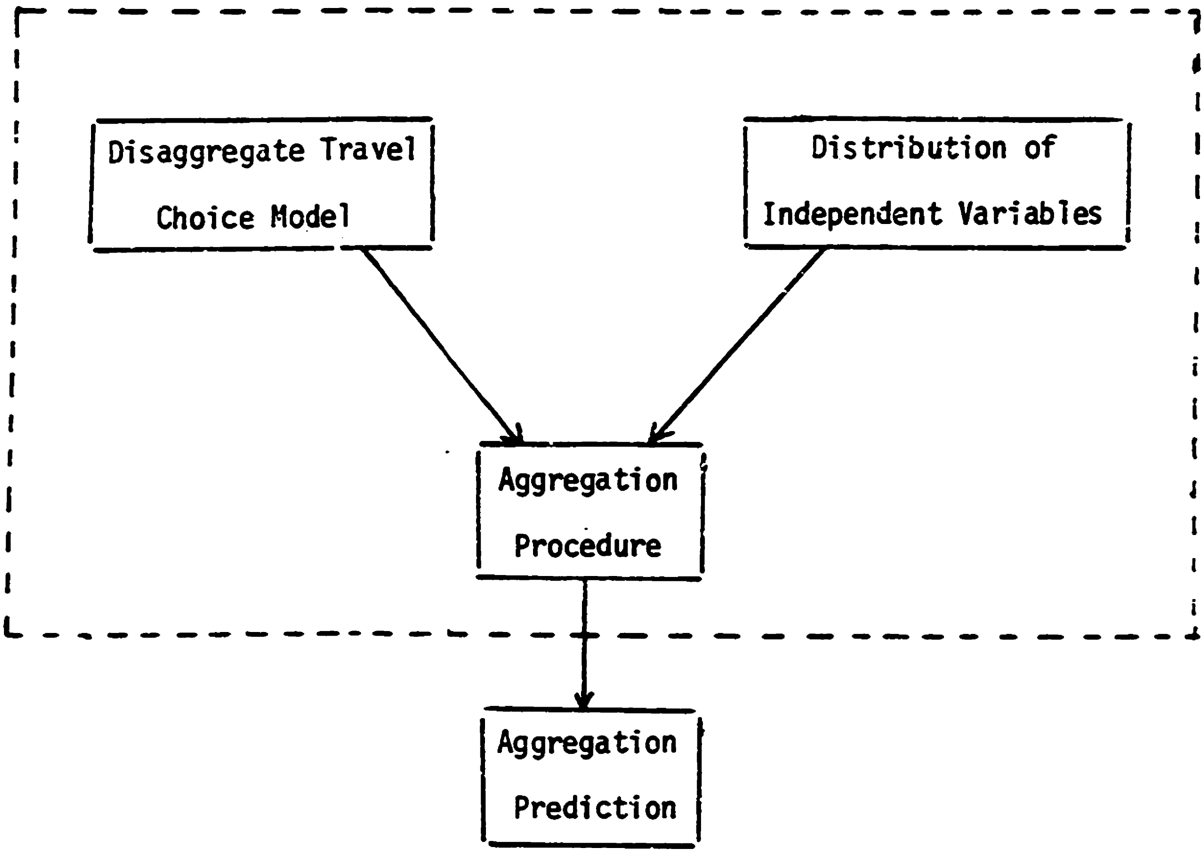
The disaggregate choice model relates the probability of a behavioral unit choosing an alternative out of the set of available alternatives to the characteristics of the behavioral unit and the attributes of the alternatives available to it.

The distribution of independent variables describes the distribution of the socio-economic characteristics and attributes of transportation alternatives for the aggregate group for which the prediction is to be made.

The aggregation procedure operates on the disaggregate choice model

FIGURE 1.1

Aggregated Prediction Model Structure



and the representation of the distribution of choice variables to produce aggregate predictions. Such procedures may result in a mathematical reformulation of the model elements to obtain an aggregate relationship based on parameters of the choice model and the variable distribution (McFadden and Reid, 1975; Talvitie, 1973) or they may be numerical computation procedures.

Different aggregation procedures are characterized by their computational formulation and their input data requirements. They range from the explicit summation or averaging procedure, referred to earlier as theoretically consistent, to the naive procedure, which requires only average values of variables for the aggregate group. All of the procedures other than explicit summation or averaging are approximate and introduce error into the aggregate predictions. An important focus of this study is to identify the structure of the errors introduced by different aggregation procedures and place them in perspective with other sources of error in the prediction process.

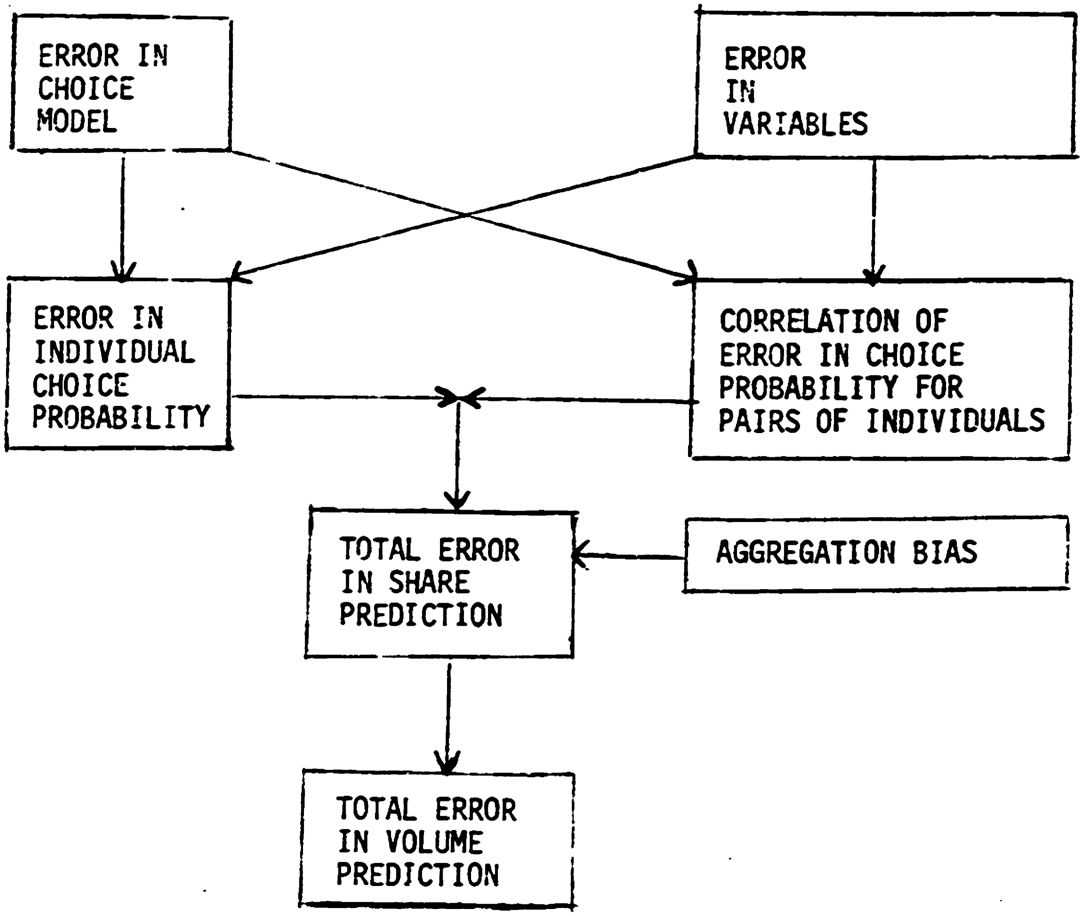
1.4 Errors in Prediction with Aggregated Models

Errors in aggregate predictions result from errors in the disaggregate choice model, errors in the description of the distribution of independent variables and errors in the aggregation procedure.

These errors interact with each other and are propagated through the prediction process to produce errors in the aggregate predictions (Figure 1.2, Propagation of Error to Share and Volume Prediction). The errors include stochastic errors and bias errors. Stochastic and bias errors enter into the choice model and variable distribution description as a result of specification errors, errors in variables, and errors in the parameter estimation procedure. These errors are propagated through the model and contribute errors to the final predictions. The stochastic errors in the parameters of both the choice model and the variable distribution description can be estimated as part of the parameter estimation process. The aggregation procedure also introduces errors into the aggregate prediction. These errors are structural in nature and introduce bias into aggregate prediction. Aggregate procedures are deterministic and do not normally introduce stochastic error into the predictions obtained although they do propagate stochastic error from other sources to the aggregate prediction. When the elements which determine the direction and magnitude of aggregation bias are not known or observable, the aggregation error may appear to be stochastic.

FIGURE 1.2

PROPAGATION OF ERROR TO SHARE AND VOLUME PREDICTION



The stochastic error in the prediction of choice probabilities is directly determined by the stochastic error in choice model parameters and choice variables. The magnitude of the prediction error is related to the magnitude of these errors. The corresponding stochastic error in the prediction of aggregate shares depends on the aggregation procedure used. When the complete enumeration or summation procedure described earlier as theoretically consistent is used the variance in aggregate share is determined by the variance in individual choice prediction, the covariance between the choice predictions for pairs of individuals and the number of individuals in the aggregate group. If there is no covariance between predictions for pairs of individuals, the variance in aggregate share declines to zero as the number of observations increases. However, when there is covariance between the individual choice predictions (some positive covariance generally exists due to the common error in parameters), the variance in aggregate shares declines to the average of the covariances as the number of observations increases. Alternative aggregation procedures propagate the variance in the model parameters and variable estimates differently.

1.5 Bias Errors of Alternative Aggregation Procedures

Mathematical and simulation analysis of a selected set of aggregation procedures provides insight into their expected performance. These analyses are conducted to obtain an understanding of the effect of aggregation error only. They use the enumeration procedure as a reference point and assume that the data required for each of the procedures is perfectly available. Any differences in share estimates by an approximate aggregation procedure and the enumeration procedure are due to aggregation bias. The mathematical analysis examines the aggregation bias of the naive, statistical differentials, and classification procedures. The simulation analysis examines these procedures plus numerical integration procedure with assumed normal and assumed uniform distributions and classification with two and three classes.

The key results of these analyses are:

- The aggregation error of each procedure is sensitive to the location of the utility distribution on the choice function. In general, maximum aggregation bias occurs in the area of maximum curvature of the choice function.
- The magnitude of bias for each aggregation procedure and each location on the choice function increases monotonically with increases in the variance of the distribution of net utility values in the aggregate group. The sensitivity of aggregation bias to changes in the variance is different for each aggregation procedure.

- The shape of the distribution affects the structure of the aggregation bias of each procedure. When the distribution is symmetric, all of the procedures examined have zero aggregation bias when the share for each alternative is equal to the reciprical of the number of alternatives. When the distributions are skewed this zero bias point is shifted to a different location on the choice function. The maximum bias location for each of the methods is also shifted when the distribution is skewed.

The results of these analyses and, in particular, comparison of maximum and average absolute values of aggregation bias for each of the procedures suggests a general order of ranking of the procedures in terms of their aggregation bias. The ranking for any particular situation described by the average net utility values and the distribution of net utilities may differ from this general ordering. When the distribution of net utility values in the aggregate group is approximately symmetric the procedures of integration with normal or uniform distribution and classification into three classes all have similar and low aggregation bias. Classification with two classes has the next higher aggregation bias followed by the statistical differentials and naive procedures. When the distribution of net utilities in the aggregate group is skewed,* the classification procedure with three classes has

*The differences in ranking, and the general differences in bias error, depend on the shape of the distribution. The simulation analysis on which the rankings are based consider one symmetric distribution and one skewed distribution.

the least aggregation bias. Integration with normal or uniform distribution and classification with two groups have slightly higher and very similar levels of aggregation bias. These are followed by the statistical differentials and naive procedure. When the distributions are skewed there are numerous instances where a procedure, which generally has higher aggregation bias than another procedure, has lower aggregation bias than that procedure. Specifically, there are some situations in which the naive procedure, which normally has the greatest aggregation bias, has the least aggregation bias. Such reversals occur for limited ranges of situations.

1.6 Disaggregation of Errors in Prediction Analysis

Different sources of errors in prediction were identified earlier. It is useful to disaggregate total prediction errors by source of error. This can be done for three major categories of error. These are:

- Specification error due to the application of the model in an area or situation different from the one in which it was estimated,
- Errors in the choice model or disaggregate predictions due to errors either in the model parameters or in the values of the choice variables,
- Aggregation error which is due to using an approximate procedure to transform the disaggregate relationship to provide aggregate predictions.

The errors from these different sources can be identified by comparison of predictions made by different procedures. In particular, the aggregation bias of any procedure can be isolated from other sources of error by comparing the predictions by that procedure to the predictions by the complete enumeration procedure.

Errors in the choice model may be isolated by comparing the predictions by the enumeration procedure, which has no aggregation error, against the observed shares. However, this comparison is clouded due to the fact that the observed shares are an approximate representation of the actual shares in the aggregate population. In some cases, the variance of the observed share estimate of aggregate shares is greater than the variance of the prediction by the models being used.

It is impossible to completely isolate possible errors due to the transfer of the model to a data set different from that on which it was estimated. Some insight into errors of transferability can be obtained by comparison of model error for predictions of different data sets as compared to predictions on the data set on which the model was originally estimated using the complete enumeration procedure in both cases.

Total errors due to errors from all of the sources discussed are obtained by comparing predictions by each of the aggregation procedures against the observed shares. Once again account must be taken of the error in the observed shares as a representation of the actual shares.

To obtain statistically useful comparisons between the predictions of alternative procedures it is necessary to test these predictions in multiple prediction situations. The errors from these individual situations can be summarized by use error measures which reduce the essential information to a small number of indices.

The error measures selected for this study consist of a basic error measure which is the difference between the prediction and the base against which it is being compared divided by the magnitude of the prediction expressed as the percentage error of the value of the prediction. These error measures are combined to produce a weighted (by size of prediction) root mean square error measure.

1.7 Empirical Analysis

The empirical analysis identifies the magnitude of aggregation error of different aggregation procedures and places this error into perspective with errors from other sources. To accomplish these objectives, the analysis considers both the aggregation error and the total error in prediction with different aggregation procedures.

The situation studied is prediction of mode share for the work trip by breadwinners in the Washington metropolitan area commuting to a workplace in the central business district by one of three modes: drive alone, shared ride, or transit ride.

The data for the analysis were collected by the Washington Council of Governments as part of their 1968 Home Interview Survey and extended to include level of service data by R.H. Pratt Associates. A merged file of household, tripmaker and level of service information created by Cambridge Systematics Incorporated (1975) was reduced to approximtely 2100 observations on households from which breadwinners made home based work trips to the central business district by one of the modes described.

The disaggregate observations in the initial data set are assigned to 45 residence districts. These districts are divided into three groups, Groups one and two in Washington and Maryland are differentiated by a random selection procedure. Group three consists of districts entirely within Virginia.

A disaggregate choice model is estimated for a subset of 874 of these households in 17 districts in Group one. The choice model includes level of service data for the alternative modes and the socio-economic characteristics of the worker and the household of which he is a member. All of the parameters are significant and have the expected sign.

An estimate of the variance in predictions in individual choice probability is made using representative values of the choice variables, assumptions about the error in the choice variables, and the errors in the parameter estimates determined as part of the model estimation procedure. Based on these estimates, the average standard error in prediction is approximately 21% of the corresponding probability value.

Predictions are made for districts in each of the three groups described earlier, for all districts grouped together and for three higher levels of aggregation.

The prediction errors are examined for four different aggregation procedures. The procedures are the enumeration procedure, the naive procedure, the statistical differentials procedure, and classification in three groups. In each case the procedure is modified to account for the fact that some members of the aggregate groups do not have the complete choice set available to them. That is, individuals who do not have a drivers license or who reside in a household which does not have a car available are assumed not to have the drive alone mode available.

Predictions are made for both the conditions which prevailed during the period in which the data set was collected and for three different sets of travel service changes. In each case the aggregation error of the naive, statistical differentials and classification procedures is determined by comparing these predictions to those by the enumeration procedure. For all of the cases considered the aggregation error was least for the classification procedure and, with one exception, second least for the naive procedure and greatest for the statistical differentials procedure. The magnitude of aggregation error for each procedure was similar for the different prediction situations and the different levels of aggregation. The orders of magnitude of error for the different procedures, in terms of the weighted root mean square error per unit of prediction, are approximately 3% of the value of the predictions for the classification procedure, 8% for the naive procedure, and between 10 and 14% for the statistical differentials procedure.

Total error is estimated for the conditions which prevailed during the data collection period by comparing the predictions by all four procedures to the observed shares. The measured total error* is related to the level of aggregation. There is a substantial reduction in total error as the number of observations in the prediction groups is increased. The measured error for the complete enumeration procedure, which includes no aggregation error, is 28% for groups with an average of 50 observations, 20% for groups with an average of 200 observations, and 15% for groups

---

*Measured total error includes errors in predictions and errors in observed shares.

with an average of 500 observations. This reduction with increasing average number of observation results in part from reduced error in the share prediction and in part from reduced error in the observed shares. At all three levels of aggregation the errors from the choice model and observed shares are great enough to overwhelm the aggregation errors of the naive and classification procedures. That is, the measured total errors for these procedures are approximately equal to those for the complete enumeration procedure. The statistical differentials procedure, however, has higher errors than the other procedures at all levels of aggregation.

1.8 Conclusions

An aggregated travel demand prediction structure provides a framework for predicting aggregate travel demand based on the behavioral structure of the disaggregate choice model. The results of the analysis indicate that it is feasible to use such a structure for the prediction of aggregate choice shares.

Analysis of aggregation bias using both mathematical and simulation techniques provides evidence of potentially large magnitudes of bias by different aggregation procedures for a limited range of choice situations. These analyses suggest a ranking of procedures in order of magnitude of error. The specific ranking of procedures is different for different distributions of net utilities between alternatives. In general, procedures of integration with assumed distribution and procedures of classification have low aggregation bias, followed by the statistical differentials and naive procedures. For each procedure, the aggregation bias is sensitive to the mean value of the net utility between pairs of alternatives and the variance and shape of the distribution of net utilities between pairs of alternatives.

The results of the empirical study indicate that aggregate predictions of choice shares based on a disaggregate choice structure using the naive or classification procedures have total error of between 15 and 30% for different size prediction groups. When these errors are adjusted for the error in observed shares the true prediction error will be lower.

Based on these results it is concluded that aggregated prediction models may be used with relatively simple aggregation procedures to provide aggregate choice share predictions. In the application of such procedures particular care should be taken to avoid, where possible, aggregation over groups which have significant differences in choice sets or are otherwise very dissimilar. Further studies should be undertaken to verify and extend these results.

Such studies should be used to:

- Verify these results in the context of different travel choices and different areas,
- Extend these results to multi-dimensional choice situations, and
- Make an explicit comparison between the prediction results using aggregated prediction models and conventional models estimated with aggregate data.

Additional studies should focus on analysis of the sources and propagation of errors in the prediction process and experimental application of the aggregated prediction models to case studies of important transportation issues.

The results obtained confirm both the desireability and feasibility of using disaggregate choice models for the prediction of aggregate travel demand.

1.9 Structure of the Report

Chapter II describes the reasons for constructing aggregated prediction models based on disaggregate choice behavior to predict aggregate travel demand. The structure and components of an aggregated prediction model framework are presented with a discussion of the implications of using such a framework in prediction.

Chapter III presents a mathematical analysis of aggregation bias using the naive, statistical differentials and classification procedures. The analysis considers the bias in share and share sensitivity predictions. The analysis indicates the importance of the distribution of values of choice influencing variables and net utilities between pairs of alternatives in binary and multiple choice situations. The effect of the location on the choice function, variance and shape of the net utility distributions is described. The expected lower bias of both the classification or statistical differentials procedure as compared to the naive procedure is identified.

Chapter IV describes a simulation analysis of aggregation bias which extends the results of the mathematical analysis to a wider range of methods for the binary choice case. The affect of difference in shape and variance of the distribution of net utilities is also examined.

Chapter V describes the way in which errors enter the model formulation and prediction process and are propagated through that process to create errors in prediction. An approach is described to be used in identifying the total error in prediction by various procedures and to disaggregate that error into its major source components. A set of parametric error measures is developed and specific measures are selected for use in the empirical analysis.

Chapter VI reports the results of the empirical analysis. Predictions made by alternative procedures are compared to the observed shares and theoretically based enumeration procedure to identify the errors associated with them. The observed errors are disaggregated into errors due to aggregation and errors from other sources.

Chapter VII presents the overall results and conclusions of the study, suggests the implications of these results for further developments of the transportation systems analysis process and points out useful directions for future research.

CHAPTER II

AGGREGATE PREDICTION MODELS

2.1 Introduction

A model which predicts the travel behavior of groups of behavioral units - individuals or households - may be considered an aggregate model of travel demand. The primary definitional characteristic of such models is that they predict the aggregate travel behavior of groups of behavioral units. It is always possible to summarize detailed predictions to any higher level of aggregation desired by explicit addition of more detailed results. Therefore, an important characteristic of an aggregate model is its ability to obtain the desired aggregate predictions using less than completely detailed, or aggregate, input data.

Travel behavior theory is postulated at the individual level. It relates the choice behavior of individuals to their own characteristics and the attributes of the choice alternatives available to them. For example, a trip maker's choice of mode to work depends on the number of automobiles which he or his family has available and the trip time by different modes from his residence to his workplace rather than some average number of automobiles available to families in his community and average trip time from this area to the area of his workplace.

The central issue in the development and evaluation of aggregate prediction models is a determination of the conditions, if any, for which such models provide predictions which are consistent with the use of more detailed information and a model structure which represents the choice behavior of the individuals who are making the travel choice decision. Consistency, in this context, implies that the difference in predictions of the aggregate procedure is small enough not to affect the results of the analysis of the problem at hand (Green, 1964). That is, the test of consistency is not one of perfectly matching the predictions of a "preferred" set of models over all possible situations, but rather one of approximate agreement within some decision tolerance over the range of situations in which the models are actually used.

Historically, aggregate travel demand models have been developed without explicit recognition of the dependence of the travel process on causal relationships at the level of the individual traveller. Rather, these models have been based on the analysis of empirical relationships between aggregate measures of travel behavior, socio-economic characteristics and transportation system characteristics with aggregations based on geographic proximity (Martin, et al, 1961; Bureau of Public Roads, 1970). Aggregation of travel, traveler and system description data prior to analysis, by reducing the variation in observed behavior, results in the loss of most of the behavioral sensitivity in trip making(Fleet and Robertson, 1968; McCarthy, 1969). In addition, these models are implicitly dependent on the distribution of characteristics within the aggregate groups used for estimation and are thereby specific to the time, place and level of aggregation at which they are calibrated (Kassoff and Deutschman, 1969; Kannel and Heathington, 1973). Aggregate travel demand model systems have been widely criticized for their failure to represent theoretically expected relationships between socio-economic and transportation service characteristics and predicted travel behavior (Manheim, 1970; Stopher and Lisco, 1970; De Donnea, 1971; Brand, 1972; CRA, 1972). In part, the weaknesses of these models are due to the failure to recognize the relationship between empirical aggregate observations and theories of travel behavior which are inherently disaggregate and a corresponding inability to construct a credible theory of travel behavior at the aggregate level.*

The purpose of this chapter is to describe a structural framework for aggregate prediction based on and consistent with disaggregate travel behavior models. Section 2.2 describes the basis for reliance on disaggregate relationships for travel prediction. Section 2.3 describes the problems of consistency and estimation efficiency in attempts to use models estimated with aggregate data. Section 2.4 defines a structural framework for aggregated prediction models based on disaggregate behavioral relationships. Section 2.5 describes criteria which may be used for the evaluation of aggregated prediction models. The components of an aggregated prediction model are described in Sections 2.6, 2.7 and 2.8. These are the disaggregate choice model, the distribution of independent variables and the aggregation procedures, respectively. Section 2.9 describes the aggregated prediction structure in the multidimensional choice context.

*For a discussion of the problem of integrating micro theory and macro representation of economic behavior, see Theil (1955), Allen (1956) or Green (1964).

2.2 The Role of Disaggregate Relationships in the Development of Aggregate Prediction Models

The purpose of travel prediction models is to predict the behavior of groups of individuals - travelers or potential travelers. The aggregate behavior of such groups of individuals is the sum of the behavior of the individuals in the group.

Disaggregate models of travel demand are based on theories of individual choice behavior. These models describe the probability of an individual choosing an alternative from a set of available alternatives as a function of his characteristics and the attributes of the available alternatives. Aggregate travel demand, the volume or share of individuals in a group to make a specific choice is the sum or average of the choice probabilities of the individuals in the group as follows:

$$V_i = \sum_{t \in T} P_t(i:A) \tag{2.1a}$$

$$S_i = \sum_{t \in T} P_t(i:A)/T \tag{2.1b}$$

where: V_i ...is the expected volume to choose alternative i,

S_i ...is the expected share to choose alternative i,

$P_t(i:A)$...is the estimated probability that individual t will choose alternative i, and

T ...indicates the set of individuals in the prediction and their total number.

This formulation explicitly represents aggregate demand as the sum of individual demands which is consistent with the definition of aggregate demand. This formulation gives consistent estimates of aggregate travel behavior at any level or type of aggregation required for a particular study. However, use of this explicit representation requires adequate information to evaluate the choice probability for each individual in the prediction group.

The analysis of travel behavior at the level of the individual is always preferred on theoretical grounds because of its correspondence with the actual behavioral choice process. Because they do not contain any aggregation structure, these models are potentially transferable to different levels of aggregation, different places (Atherton, 1975) and different times (Kannel and Heathington, 1973).

Attempts to use disaggregate models as a basis for prediction have been made in recent years (PMM, 1973; Tahir and Hovind, 1973; Ben-Akiva and Richards, 1975; Liou et al, 1975;Difiglio and Reed,1975). These prediction applications are primarily concerned with travel mode choice behavior.

The advantages which may be gained by use of models based on disaggregate choice behavior rather than aggregate correlative relationships are considerable. Some of these a.e:

1. Predictive validity: the relationships which are constructed through disaggregate analysis represent behavioral relationships rather than the correlative relationships generally observed in analysis of aggregate data and therefore are more re-

liable for prediction of travel behavior under changed conditions.

2. Applicability to more areas over a longer time span: individually based behavioral relationships should be transferable to a wide range of area types as opposed to aggregate models which are sensitive to the scale of analysis and the existing distribution of characteristics in the aggregate groups on which they are based.
3. Realistic sensitivity to changes in transportation service characteristics: due to its behavioral rather than correlative base, disaggregate analysis provides a model which can more reliably predict the influence of changes in system characteristics than aggregate models. Furthermore, individual or household observations generally include a wide range of service characteristics and should be valid over a wide range of future conditions.
4. Potential for use at alternative levels of aggregation: a major advantage of the disaggregate analysis is that it may be applied at any level of aggregation which is useful to the policy planner and that it will produce consistent results at different levels of aggregation. Aggregate based models do not have this capability.
5. Increased efficiency in the use of data: a disaggregate analysis may be based upon relatively small data sets drawn from one or

a few areas.

For all of these reasons, aggregate models should be based on the disaggregate analysis of travel choice behavior.

The gap between the need for aggregate forecasts and the general availability by familiar prediction methods of aggregate independent variables,versus the model of travel behavior at the disaggregate level, must be bridged by an explicit aggregation procedure. Using such a procedure, aggregate predictions of travel behavior can be made based upon the structure of the disaggregate model, forecasts of the socio-economic characteristics of the population and a description of the transportation services to be provided.

2.3 Aggregate Models of Travel Behavior - Problems of Consistency and Estimation Efficiency

An inconsistency exists between the theory of travel behavior expressed in terms of the individual or other behavioral unit and the dominant empirical approach to the analysis of travel behavior of the last two decades. Early developments in travel demand modelling were motivated by the need to obtain aggregate predictions for use in planning capital investment in new transportation facilities. The models developed were based on estimated relationships between aggregate observed travel and aggregate socio-economic characteristics. These models generally failed to consider, or only partially considered, the effect of level of service characteristics on travel demand.

Theories of travel behavior, to the extent they existed, were developed by analysis of the observed aggregate relationship and often by analogy with other physical phenomena. These theories relied on mass statistical relationships which were effective in fitting the observed data, but were of limited value in explaining the underlying causal relationships or predicting the effects of alternative investment and/or operating strategies on system performance, travel flows and external impacts.

More recently, disaggregate models of travel behavior have been developed based on theories of individual choice behavior. These models were initially used as research tools to improve understanding of the way in which socio-economic characteristics and level of service attributes influenced travel choice behavior (Lave, 1969; Lisco, 1967; Stopher and Lavender, 1972; etc.). The gains in understanding obtained both through the theoretical developments and model estimation and analyses provide a basic structure from which analogous reasoning might be used to improve the specification of aggregate relationships. The explicit representation of aggregate travel demand based on models of individual choice is a simple summation of travel choice over groups of individuals to obtain estimates of total travel demand.

The open question is whether it is possible to create an aggregate model which is consistent with the underlying disaggregate relationship and which is in a form which can be estimated. The development of a consistent aggregate functional relationship depends on the structure of the disaggregate choice relationships and the distribution of independent variables. It has been shown (Theil, 1955; Green, 1964) that, if the disaggregate relationship is linear or the members of each prediction group are homogeneous with respect to the choice influencing variable and taste preferences, the aggregate function has identical form and parameters as the disaggregate function. In this case consistent aggregate predictions are obtained by using average values of the independent variables.

However, in the present case, where the disaggregate function is non-linear and the prediction groups are generally not homogeneous, the consistency requirement can only be satisfied by a model which includes parameters of the distribution of these variables as well as parameters of the choice function. The selection and method of inclusion of these parameters should be based on the structure of the distributions and of the choice models which represent the behavior of the aggregate group under study. The direct method of obtaining the aggregate function is by integration of the choice function over the distribution of the independent variables (Kanafani, 1972). However, the conditions which must be placed on the structure of the choice model and the distribution of independent variables to obtain a closed form function are highly restrictive. McFadden and Reid (1975) demonstrated that a disaggregate probit binary choice function can be integrated over multivariate normally distributed independent variables to obtain an aggregate binary share probit model. They also describe an iterative procedure which can be used to estimate the aggregate model. It appears that this result can be extended to multiple choices. However, problems of estimating the multiple choice probit function,combined with the iterative procedure which must be used to estimate the aggregate model, mitigate against general use of this approach.*

Thus, if aggregate predictions are to be made which are consistent with underlying disaggregate choice models, an aggregate prediction procedure must be developed which is explicitly based on the disaggregate choice model. The structure of such aggregated prediction models

*Prediction with a multivariate probit structure is also potentially very costly for choice sets with more than four or five alternatives.

is described in the next section.

2.4 Structure of an Aggregated Prediction Model

An aggregated prediction model is one which explicitly incorporates disaggregate behavioral relationships in an overall structure which allows aggregation of either the model itself or of its mathematical relationships to obtain aggregate predictions of travel behavior. A model structure for aggregated prediction (Figure 2.1, Aggregated Prediction Model Structure) consists of three components. These are:

- 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.

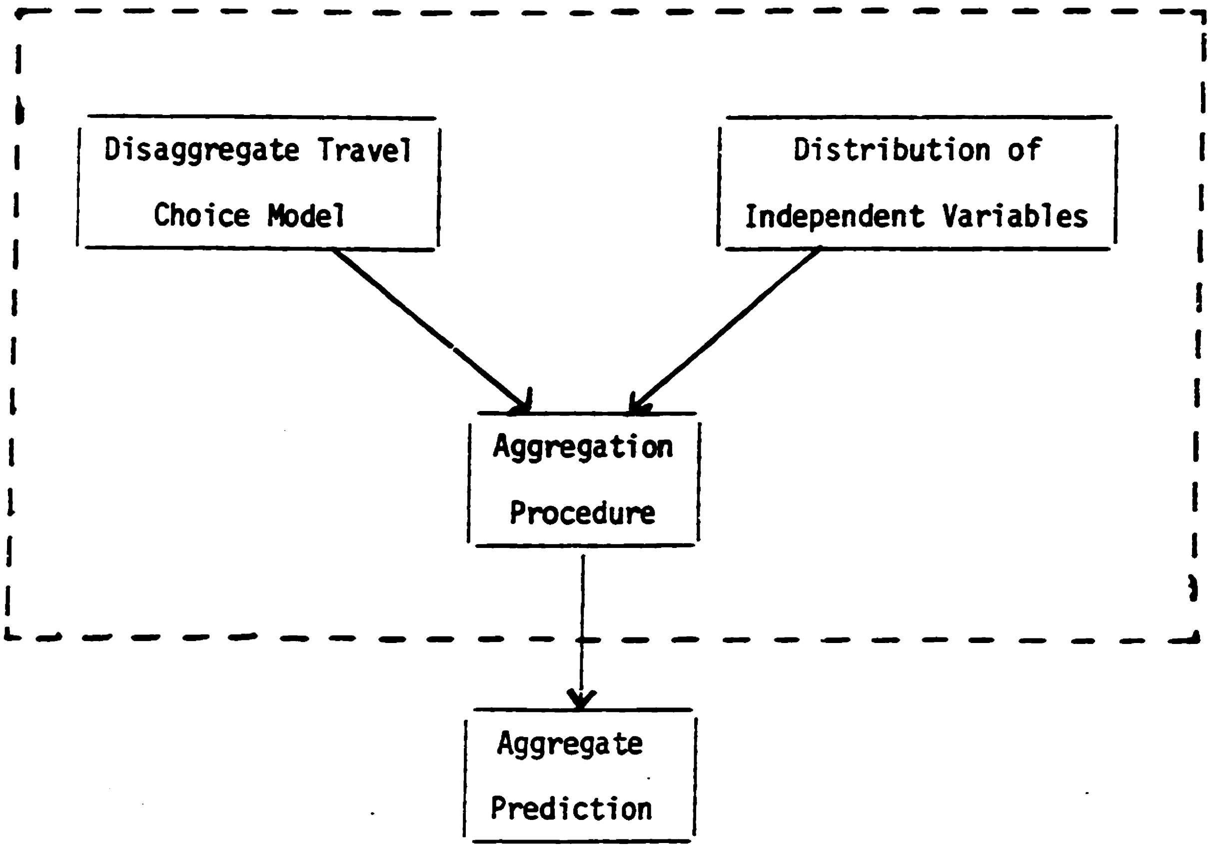
The <u>disaggregate choice model</u> relates the probability of a behavioral unit choosing an alternative out of a set of available alternatives as a function of the attributes of the alternatives and the characteristics of the individual or behavioral unit. This is expressed, mathematically, by

$$P_t(i:A_t) = f^i(X_{jt}, \forall j \in A_t) \tag{2.2}$$

where $P_t(i:A_t)$...is the probability of individual t to choose alternative i of the set of alternatives available to him, A_t

FIGURE 2.1

Aggregated Prediction Model Structure



$f^i(X_{jt}, \forall\ j \in A_t)$...identifes the specification of the choice model structure, and

X_{jt} ...is the vector of attributes of alternative j for individual t.

The independent variables are disaggregate variables which describe the individual faced with a travel choice and the attributes of the alternatives available to that individual.

The distribution of independent variables can be represented by a joint density function which describes the distribution in the aggregate group of all of the variables which influence choice behavior:

$$g(X_{jt}, \forall\ j \in A_t) \tag{2.3}$$

where $g(X_{jt}, \forall\ j \in A_t)$...represents the distribution in the aggregate prediction group of the variables which influence individual choice behavior.

The joint density function may describe a multivariate continuous distribution, a multivariate discrete distribution or some combination of discrete and continuous distributions. In the extreme case it may represent the exact or predicted characteristics of all of the individuals in the aggregate group and the attributes of the alternatives which they individually have available to them. In a more general way the distribution is some method of describing the frequency of occurrence in the prediction group of specific values of the choice influencing characteristics.

The aggregation procedure operates on the disaggregate behavioral structure and the representation of the distribution of independent variables to produce aggregate predictions. A wide range of procedures can be used for this purpose. They may result in a mathematical reformulation of the model elements or may operate numerically. The theoretically consistent procedure described earlier (Equation 2.1) computes the volume or share of aggregate travel directly by summing or averaging individual choice probabilities. The difficulty of satisfying the extreme data requirements of this formulation motivates the search for aggregation procedures which have less detailed information requirements.

The structure of an aggregated prediction model constructed from the disaggregate model makes explicit the dependence of aggregate travel demand on:

- The individual's response to the attributes of the alternatives which he faces (the disaggregate choice function), and
- The distribution of individuals according to their characteristics and the attributes of the alternatives they face.

This explicit representation is the basis of two major advantages of explicitly aggregated prediction models based on disaggregate analysis over aggregate forecasting models based on correlative analysis of aggregate data. These are:

- Improved sensitivity to changes in individual behavior due to changes in travel service or other environmental attributes, including policy controlled variables, and

- Sensitivity to changes in the distribution of the characteristics of, or faced by, the population.

The structure of the aggregated model requires the development of an independent variables distribution model to be employed in conjunction with the disaggregate choice mode. This adds a potentially complicating dimension to the application of disaggregate models. However, it is always possible to develop simple models of distribution based on assumptions which are similar to the implicit "no change" assumptions which are embodied in existing models based on aggregate relationships. Explicit representation of the distribution assumptions makes it possible to develop and use improved representations of the distribution of independent variables.

The next section describes criteria to be considered in the evaluation and selection of aggregated prediction models for use in particular situations.

2.5 Criteria for Evaluation of Aggregated Prediction Models

It is useful to identify and describe criteria which may be used to characterize and evaluate alternative aggregated prediction models. The proposed criteria are grouped in two categories which describe the performance characteristics and operational requirements of different aggregated prediction models. The performance characteristics describe the satisfaction of prediction requirements of planning situations in which the model will be used. The operational requirements describe the demands which the model places on analytic resources. Tradeoffs will have to be made between improved performance and reduction of the operatioanl requirements of different models.

The performance criteria include (1) the prediction accuracy of the model in the range of situations for which it will be used, and (2) the sensitivity of the model to changes in policies for which predictions are required. These criteria are interrelated and are dependent on the conformity of the aggregated model with the theoretical relationship between aggregate demand and disaggregate choice behavior. Prediction accuracy can be evaluated in part by comparison of model predictions with observed behavior. Testing beyond the range of available observations must be based on conformity with the underlying theoretical structure.

Sensitivity of the aggregated prediction model to relevant policy alternatives is determined by the specification of the underlying disaggregate choice model and the aggregation procedure used. The range of policy alternatives which may be studied has increased dramatically in recent years. This increase results from increased recognition of the range of options which are available to government officials for modifying transportation supply characteristics. The ability to model the impact of such changes depends on the proper specification of the disaggregate choice model and also of the related supply models.* The range of policy options to be considered should include capital investment, operating and pricing alternatives. The level of service characteristics included in the travel choice model must be selected to allow sensitivity to changes in these policies. As a group, these level of service characteristics should describe the factors which influence individual choice behavior.

The operational requirements category includes criteria which describe the ease of use and resource cost of different aggregated prediction procedures. These criteria represent the practical issues associated with using different procedures such as input data requirements, computational requirements and staff support requirements. Different aggregation methods have different requirements for knowledge of the distribution characteristics which influence the disaggregate choice behavior. The degree of information required for use with different

*Predictions of travel volumes and level of service characteristics are the product of a supply-demand equilibration process. The models described herein are the demand component of the more general model structure. Proper transformation of supply decisions into level of service characteristics is the function of the supply model component.

aggregation procedures varies from complete knowledge of all the independent variables for every behavioral unit to knowledge of the mean value of the independent variables for the aggregate group of behavioral units.

The degree of satisfaction of these different criteria is interrelated. In general, procedures which give the most accurate and most theoretically consistent predictions are computationally demanding and have substantial data requirements. Alternatively, the least demanding models with respect to computational and data requirements generally give the least accurate and least theoretically consistent predictions. This tradeoff must be evaluated in selecting an aggregated prediction procedure for a specific application.

2.6 The Disaggregate Choice Model

The central characteristic of disaggregate choice models is the explicit representation of the choice process as probabilistic rather than deterministic. The probabilistic structure of disaggregate choice models is derived from the assumption that the utility which the individual asribes to an alternative is a random variable. If the utility of each alternative to an individual were fixed and known, the prediction process would consist of comparing the utilities of the available alternatives and assigning the individual to the alternative with the highest utility. However, when the utility is viewed as a random variable the choice probability of an alternative is:

$$P_t(i:A_t) = \text{Prob}\ (U_{it} \geq U_{jt}\ \forall\ j \in A_t) \tag{2.4}$$

where U_{it}...represents the utility of alternative i to individual t.

That is, the probability of choosing an alternative from the available set is identically the probability that its utility is greater than the utility of any other available alternative.

The functional form of the choice model is determined by the selection of a specific joint distribution of the utility functions and analytically solving the probability expression (Equation 2.4). This is done by expressing the utility in terms of a systematic portion and a random portion

$$U_{it} = h(V_{it}, \ e_{it}) \tag{2.5}$$

where V_{it}...is the systematic portion of the utility function,

e_{it}...is the random portion of the utility function, and

$h(V_{it}, \ e_{it})$...expresses the relationship between the systematic and random portions.

The most common assumption is that the disturbance term is additive so that

$$U_{it} = V_{it} + e_{it} \tag{2.6}$$

Substituting this expression in Equation 2.4 gives

$$\begin{aligned} P_t(i:A_t) &= \text{Prob } (V_{it} + e_{it} \geq V_{jt} + e_{jt}, \forall \ j \in A_t) \\ &= \text{Prob } (e_{jt} - e_{it} \geq V_{it} - V_{jt}, \forall \ j \in A_t) \end{aligned} \tag{2.7}$$

which can be solved by integrating over the distribution of the random terms (CRA, 1972).

A generalized probabilistic choice function is represented in terms of the systematic portion of the utility function by

$$P_t(i:A_t) = f^i(V_{jt}, \ \forall \ _j \in A_t) \tag{2.8}$$

where $f^i(V_{jt}, \forall \ _j \in A_t)$...expresses the function of the systematic portion of utilities in a particular form for prediction of the choice probability for the i^{th} alternative.

Specific functional forms result from different assumptions about the distribution of the random element of the utility functions. The functional forms resulting from various distributional assumptions are described by CRA (1972).

The systematic portion of the utility, V_{it}, is a function of the attributes of alternative i (prices and service characteristics) and the characteristics of individual t (income and other characteristics affecting tests). Ben-Akiva (1973) described alternative characteristic variables as generic (applies to all alternatives) or alternative specific (applies to a subset of alternatives). Alternative specific variables include alternative preference variables which have a value of one for a given alternative and of zero for all other alternatives. Socio-economic variables enter into the model in a meaningful way when they are transformed to have alternative specific values.*

The general probabilistic choice function may be expressed in terms of the vectors of variables which influence the utility of each alternative by:

$$P_t(i:A_t) = f^i(X_{jt}, \ \forall_j \in A_t) \tag{2.9}$$

where $f^i(X_{jt}, \ \forall_j \in A_t)$...is now interpreted to represent the model structure and parameters, and

X_{jt}...is the vector of variables which influence the utility of alternative j to individual t.

*If socio-economic variables are not alternative specific, they do not influence the difference between alternative utilities and therefore do not affect choice probabilities.

The first developed disaggregate travel behavior models described individual mode choice behavior (Warner, 1962; McGillivray, 1969; Lave, 1969; Stopher, 1969). More recently, disaggregate models have been developed for a wider range of travel decisions including mode choice, time of day, destination and frequency in a general model structure (CRA, 1972). Models are currently being developed to describe simultaneous choice along two or more of these dimensions (Ben-Akiva, 1973; Lerman and Ben-Akiva, 1975; Adler and Ben-Akiva, 1975; Lerman, 1975).

When the choice set includes multiple choices, the structural relationship between choices can be identified as independent, recursive or simultaneous (Ben-Akiva, 1973).

The joint probability is represented by

$$P_t(i,j : AB) \tag{2.10}$$

which is the probability of individual t jointly choosing i and j from the set of alternatives AB. If the i,j choices are independent, the joint choice probability may be written

$$P_t(i,j : AB) = P_t(i{:}A)\cdot P_t(j{:}B). \tag{2.11}$$

If the joint choice is recursive, i conditional on j, the joint choice probability may be written

$$P_t(i,j : AB) = P_t(i{:}A/j)\cdot P_t(j{:}B) \tag{2.12}$$

where $P_t(i:A/j)$...expresses the probability of choosing j of the set B given that i has already been chosen.

If the choices are simultaneous, the joint choice function (Equation 2.10) cannot be simplified.

The choice structure of the disaggregate model is determined on theoretical grounds. The representation of different joint choice structures in the aggregated prediction model is described in Section 2.9.

2.7 The Distribution of Independent Variables

The development of aggregate predictions of travel behavior consistent with the underlying behavioral structure requires representation of the distribution of independent variables.

The distribution of variables may be represented in one of four ways. These are:

1. Enumeration which describes the distribution by actual or estimated values of the variables for each disaggregate unit. Complete enumeration exists when data is available for every member of the aggregate prediction group. Partial enumeration exists when data is available for a subset of the prediction group. Partial enumeration is based on a sample of the group members selected to represent the true distribution in the total group.
2. Density functions represent the distribution of variables in terms of the frequency of different variable values in the prediction group. These distributions may be based on theoretical and/or empirical analyses which provide guidelines for selection of both the structure of the distributions and their parameters. Density functions may be discrete or continuous depending on the nature of the variable. For example, the number of household members and cars available are discrete variables while income and travel time are continuous variables. The distributional

structure may attempt to fully represent the complexity of the true distributional structure or may include simplifying assumptions about the shape of the individual distributions and their interactions.

3. Distribution moments describe the distribution in terms of its moments and cross moments which provide information about the spread and shape of individual distributions and their interactions. The successive addition of higher order moments increases the refinement of the distributional representation.

4. Classification describes the assignment of members of the prediction group to relatively homogeneous subgroups. The assignment is based on the members' values for one or a set of variables. A wide range of information refinement can be represented depending on the fineness of the classifications and the number of characteristics used.

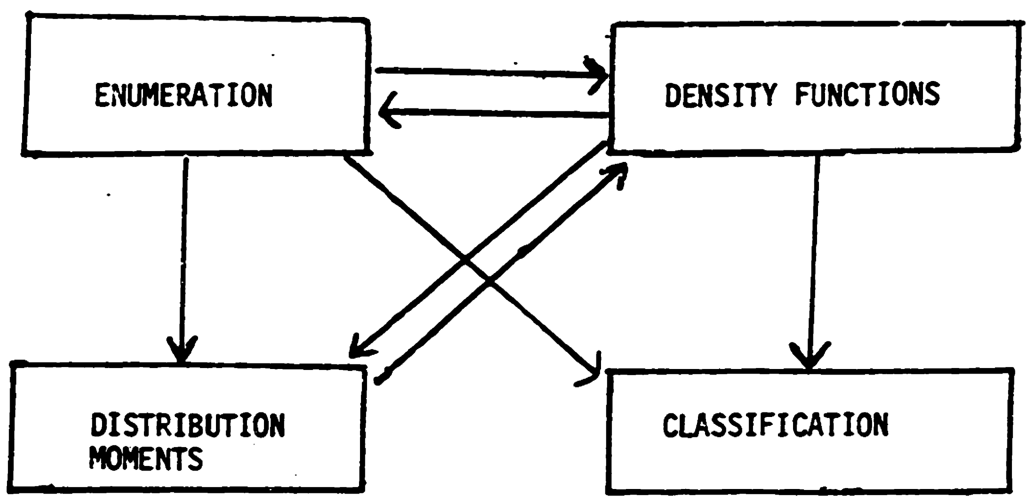
These different methods of representing the distribution of independent variables provide a range of information about the actual distributions. Enumeration provides maximum detail about the distribution of variables in each prediction group. Density functions, which have less detail with respect to a single group take advantage of outside information, either theory or related empirical analysis, to provide some structure which may describe the distribution in the prediction group more accurately than a random sampling of the same population.

The different representations of variable distributions are interrelated. Figure 2.2, Relationships Among Methods of Representing the Distribution of Independent Variables, indicates (by the arrows shown) the possibilities for transforming information from one representation to another. Enumeration, if based on a large enough sample, can be used to estimate parameters of a density function or distribution moments and can be used as a basis for structuring a classification schedule. Density functions may be used to generate a pseudo enumeration through use of a Monte Carlo procedure. Density functions can be used to determine distribution moments or as a basis for classification. Distribution moments also may be used to identify density functions through use of a system of density functions (Pearson, 1895) or by augmenting them with an assumed distributional form. Classifications, unless they are extremely fine, are not usable for generating any of the other distributional representations.

Different methods may be used to forecast the required distributions. The simplest distribution forecast procedure is to project the existing distribution in a zone unchanged over the period of interest except for already planned or in process changes which can be explicitly identified. This assumption is best for short-term predictions. For example, the near term effect of a change in public transit service can be based upon the existing distribution of household and highway service characteristics with changes in transit service characteristics only.

FIGURE 2.2

RELATIONSHIPS AMONG METHODS OF REPRESENTING
THE DISTRIBUTION OF INDEPENDENT VARIABLES



Another approach is to relate parameters of distributions or their moments to a small number of indices, such as mean values, which are commonly predicted. These relationships can be used to develop a distribution structure based on existing prediction capabilities. These relationships can also be modified over a period of time to reflect assumptions about changes in the distribution of socio-economic characteristics or policy shifts designed to change the distribution of service characteristics.

The distribution of spatially related transport service characteristics facing household is interrelated with the distribution of socio-economic or household characteristics. This suggests the need to develop joint distributions over household and spatial characteristics. This step is much simplified if an assumption of independence can be justified between household and spatial distribution, or if the dependence can be simply specified.

However, it is more reasonable to argue that the distribution of household and spatial characteristics are related through the household location choice process. This relationship can be modelled by first forecasting the region wide distribution of household characteristics to geographic locations as part of a residential location choice model which explicitly accounts for geographic, neighborhood and transportation service characteristics (Lerman, 1975).

This situation can be described by a recursive residential location-mode choice structure. In this case households which are identified by the workplace of the breadwinner, first select a residential location and then choose a mode to work based on the residential location.* The variable distribution required in this case is the distribution for households which have made a prior choice of residential location. In this case the distribution of variables for the mode choice decision is derived from the total CBD worker distribution by:

$$g(X|j) = \frac{P_t(j|X)\ g\ (X)}{\int_X P_t(j|X)g(X)dX} \qquad (2.13)$$

where $g(X|j)$...is the distribution of variables given a prior choice of residential location, j,

$P_t(j|X)$...is the probability of choosing a residential location j by a household faced with the variables, X, and

$g(X)$ ...is the distribution of variables for all households who have a member working in the central business district.

That is the relative frequency of occurrence of certain sets of values in the subgroup (CBD worker households residing in a particular location) is based on the frequency of those values in the general population weighted by the probability of choosing to join the subgroup when those values occurred. In this way, the distributions which are relevant to each set of choices become linked together in a structurally consistent manner. When the choice of interest is embedded in a sequential choice

*The structure of the actual choice process is more complex. However, this representation is a reasonable description of short term behavior where the residential location is fixed by prior choice, but the transit choice is subject to revision.

process, the conditional distribution of choice variables may be predicted directly or the overall marginal distribution of choice variables may be predicted and processed through a series of choice steps. The major tradeoffs in such a decision are (1) the modelling effort involved, (2) the accuracy of the different methods of predicting the conditional distribution and (3) the value of maintaining consistency among distributions used in different stages of the modelling process.

2.8 The Aggregation Procedure

The aggregation procedure is the element of the aggregated prediction model which operates on the disaggregate choice model and the distributional representation to obtain estimates of aggregate shares or volumes. As described in Section 2.1, the aggregation procedure which represents the theoretical relationship between aggregate demand and individual choice behavior is the explicit summing or averaging of the individual choice probabilities for the behavioral units in the prediction group to obtain the aggregate volume or share estimates (Equation 2.1). However, this procedure requires more information about the distribution of independent variables than is generally available. For this reason, a range of alternative aggregation procedures are proposed. These procedures can be characterized by their data requirements and computational approach. The possible procedures cover the range from the theoretically consistent explicit summation procedure to the naive procedure which is known to be biased but has the least data and computational requirements. For convenience, the possible procedures are grouped in five categories which correspond to differences in approaching the aggregation problem. These categories are procedures of enumeration, summation-integration, statistical differentials (or moments), classification and the naive procedure. Each of these groups is discussed in turn.

Procedures of enumeration. This group includes those procedures which represent the explicit theoretical relationship between aggregate and disaggregate demand. The expected volume choosing an alternative is the sum of the individual choice probabilities for that alternative, and the expected share is the average of the disaggregate choice probabilities, that is:

$$V_i = \sum_t P_t(i:A) = \sum_t f^i(X_t)^* \tag{2.14a}$$

$$S_i = \sum_t P_t(i:A)/T = \sum_t f^i(X_t)/T \tag{2.14b}$$

This formulation requires complete knowledge of the individual characteristics and the attributes of available alternatives. The degree of information required may be reduced by use of a sample of behavioral units from the prediction group. In this case Equations 2.14 are modified to

$$V_i = \sum_t f^i(X_t) \cdot \frac{T}{T_S} \tag{2.15a}$$

$$S_i = \sum_t f^i(X_t) \cdot \frac{1}{T_S} \tag{2.15b}$$

where T_S...is the number of behavioral units in the prediction sample.

This procedure introduces a certain amount of stochastic variation and/or bias which results from the nature of the sampling process.

---

*This simplified notation replaces $f^i(X_{jt}, \forall j \in A)$ used previously.

Procedures of enumeration are important as they represent the theoretically consistent aggregation process (Equation 2.1). For this reason they can be used as a basis for the evaluation of alternative procedures.

Procedures of summation/integration. These procedures represent a weighting of the disaggregate choice probability estimates by the probability density function for the independent variables. This is done by integration or summation depending on the structure of the distribution by:

$$S_{iD} = \int_X f^i(X)\ g(X)dX \tag{2.16a}$$

or

$$S_{iD} = \sum_X f^i(X)\ g(X) \tag{2.16b}$$

where S_{iD}...is the prediction of share i by procedures of summation or integration.

If the distributional representation is precise these procedures give results which are the same as those obtained by complete enumeration (Equations 2.14). To the degree that the distributions used are approximations to the actual distribution of characteristics, the predictions obtained differ from those based on enumeration. When limited sample data is available, it may be possible to use data from different groups to determine some of the characteristics of the distribution and thereby enrich the distributional information for specific prediction groups. The computational requirements of procedures in this group are high if the integration/summation must be applied over a large number of variables. This group of procedures may use distributions determined from theoretical and empirical analyses, and/or distributions which are known not to represent the true distributions but which are assumed in order to obtain computational or other advantages.

The most advantageous distributional assumption is that the variables have the multivariate normal distributions. McFadden and Reid (1973) used this assumption with a binary choice probit model of the form

$$P_t(i:A) = \Phi\,(X_t'\,b)^* \tag{2.17}$$

with X_t assumed to be distributed multivariate normal with mean $\overline{X}$ and covariance matrix A so that $X_t'b$ is normally distributed with mean $\overline{X}'b$ and covariance $\sigma_v^2 = b'Ab$. By mathematical integration they obtained an aggregate share probit model:

$$S_{iD} = \Phi\left(\frac{\overline{X}_t'\,b}{\sqrt{1+\sigma_v^2}}\right) \tag{2.18}$$

which is commonly available in tabulated form. This representation is used in Chapter III to analyze the bias associated with the naive and statistical differential procedure. Westin (1974) used the same assumption with a binary choice logit model. The major advantage of the nor-

---

* The symbol $\Phi\,(\omega)$ implies $\int_{-\infty}^{\omega} \frac{1}{\sqrt{2\pi}} \exp[-y^2/2]\, dy$

mality assumption is the computational efficiency which is obtained by collapsing the multivariate normal distribution of independent variables to a single variable normal distribution for the utility of each choice alternative. This computational efficiency is obtained at the expense of some lost accuracy in the representation of the actual distribution.

Procedures of statistical differentials. This group of procedures expresses aggregate shares as a function of the moments of the utility distribution. The aggregate function is obtained by linearizing the disaggregate choice function by use of a Taylor series expansion and taking expectation across the aggregate prediction group (Talvitie, 1973). The share estimate in the binary choice case is given by

$$S_{iD} = f^{i}(\overline{V}) + \frac{1}{2} \left. \frac{\partial^{2} f^{i}(V)}{\partial V^{2}} \right|_{\overline{V}} \sigma_{V}^{2} + \ldots . \tag{2.19}$$

where $f^{i}(\overline{V})$ ...is the choice function in terms of the net utility between the two alternatives evaluated at the mean value of the net utility,

$\left. \frac{\partial^{2} f^{i}(V)}{\partial V^{2}} \right|_{\overline{V}}$ ...is the second derivative of the choice function with respect to net utility evaluated at mean net utility, and

σ_{V}^{2} ...is the variance of the net utility distribution in the prediction group.

The additional terms in the series consist of higher order derivates, moments and factorials. Practical issues associated with estimating higher order moments and instability of the series when the distribution is highly dispersed (Johnson and Kotz, 1969), indicate that the series should be terminated after the second term.

The corresponding equation for the multiple choice sitatuion is:

$$S_{iD} = f^i(\overline{V}) + \frac{1}{2} \sum_j \sum_k \left. \frac{\partial^2 f^i(V)}{\partial V_j \partial V_k} \right|_{\overline{V}} \mathrm{Cov}(V_j, V_k) \quad (2.20)$$

where $V, \overline{V}$ ... in this case represent the vector of utility values or mean utility values, respectively.

Procedures of classification. This group of procedures is based on (1) assigning members of the aggregate group to two or more classes, (2) using the naive method (to be described next) to predict aggregate choice shares for each class and (3) computing overall aggregate share as the weighted average of the class shares. The share prediction by the classification method is:

$$S_{iC} = \sum_G f^i(\overline{X}_G) \frac{T_G}{T} \quad (2.21)$$

where S_{iC} ... is the share predicted by the classification procedure,

$\sum_G$ ... indicates summation over all the subgroups,

$\overline{X}_G$ ... is the vector of average variable values for individuals in subgroup G,

T_G ... is the number of individuals in subgroup G, and

T ... is the number of individuals in the prediction group.

Procedures in this group are differentiated by the variables which are used for classification and the number of classes used. The selection of variables for classification should be aimed at reducing the variance of the net utility distributions. This can be accomplished by selection of the variable which contributes the greatest proportion of the total variance in the utility distribution.

Naive procedure. This category includes only one procedure which is to use the mean value of the choice influencing variables in the disaggregate choice function:

$$S_{iN} = f^i(\overline{X}) \qquad (2.22)$$

where $\overline{X}$ ... is the vector of average variable values for the prediction group.

Although this procedure is the degenerate case of the procedures of statistical differentials, it is given separate treatment for three reasons. First, the data requirements are the same as those for presently used models calibrated with aggregate data. Second, it is computationally and conceptually simple. Third, it is the method most likely to be used in the absence of recognition of the aggregation problem.

The five groups of procedures which have been identified are:

Procedures of Enumeration,

Procedures of Integration/Summation,

Procedures of Statistical Differentials,

Procedures of Classification, and

Naive Procedure.

Procedures of enumeration includes complete enumeration and partial (based on a sample) enumeration. Procedures of integration/summation includes procedures based on distributions determined by theoretical and empirical analysis and procedures based on assumed distributions. Procedures of statistical differentials includes prediction with mean and variance terms only. Procedures of classification includes a wide range of procedures differentiated by the variables or variables selected for classification and the number of classes used. The last group includes only the naive procedure.

2.9 Aggregated Prediction Structure in Multi-Dimensional Choice Context

The purpose of this section is to set forth the prediction implications of the multi-dimensional choice set of which the choice of interest is an element. The alternative choice structures identified in Section 2.6 are joint, recursive or independent. The effect of sequential prediction on the distribution of choice influencing variables for conditional choices in a sequence was described in Section 2.7 (Equation 2.13). This section integrates these concepts in the development of alternative representations of aggregate travel demand for prediction. Expressions for aggregate choice share are developed in terms of the disaggregate choice model and the variable distribution density function for different choice structures. The section concludes with a discussion of the implications of the results for the prediction of a realistic set of travel choices.

The expected aggregate shares for a two dimensional choice situation are given by:

$$S_{ij} = \int_X P_t(i,j:AB)g(X)dX \qquad (2.23)$$

where S_{ij} ... is the expected share of behavioral units to choose i and j,

$g(X)$... is the density function for the vector of choice influencing variables, and

X ... is the vector of choice influencing variables.

If the choices are made jointly the prediction of joint shares is given by Equation 2.23. Prediction of either of the marginal shares can be made by summation of the joint shares:

$$S_i = \sum_j S_{ij} \tag{2.24a}$$

$$S_j = \sum_i S_{ij} \tag{2.24b}$$

If the choice model is conditional, i dependent on j, the joint shares may be obtained directly by

$$S_{ij} = \int_X P_t(i:A/j)\, P_t(j:B)\, g(X)\, dX \tag{2.25}$$

However, it is possible to develop an alternative formulation which makes use of the conditional choice structure. To accomplish this the relationship between conditional and independent variable distributions in Equation 2.13 is rewritten to obtain:

$$g(X/j) = \frac{P_t(j:B)\, g(X)}{\int_X P_t(j:B)\, g(X)\, dx} \tag{2.26}$$

from which

$$P_t\,(j:B)\, g(X) = g(X/j) \int_X P_t(j:B)\, g(X)\, dX = g(X/j) S_j \tag{2.27}$$

so that the joint share is given by:

$$S_{ij} = \int_X P_t(i:A/j)\ g(X/j)\ S_j\ dX$$

$$= S_j \int_X P_t(i:A/J)\ g(X/j)\ dX$$

$$= S_j \cdot S_{i/j} \tag{2.28}$$

That is, if the share choosing alternative j of set B is known, the joint choice shares can be predicted using the choice probability for i and the distribution of the choice variables both conditional on choice j. The marginal share choosing j is predicted directly by:

$$S_j = \int_X P_t(j:B)\ g(X)\ dX \tag{2.29}$$

The marginal choice probability for i is obtained by summation over the joint shares which include i as shown in Equation 2.24b.

When the choices of i and j are independent, the joint choice shares can be predicted directly by:

$$S_{ij} = \int_X P_t(i:A)\ P_t(j:B)\ g(X)\ dX \tag{2.30}$$

or sequentially, using equation 2.27 by

$$S_{ij} = \int_X P_t(i:A)\ g(X/j)\ dX \cdot S_j \tag{2.31a}$$

or

$$S_{ij} = \int_X P_t(j:B)\ g(X/i)\ dX \cdot X_i \tag{2.31b}$$

In the case where the variables which influence the choice of i are independent of the variables which influence the choice of j, the conditional distributions are the same as the unconditional distribution. That is, if X can be decomposed into X_i and X_j which are independently distributed

$$g(X) = g(X_i)\ g(X_j) \tag{2.32}$$

the direct aggregate predictive equation becomes:

$$S_{ij} = \int_{X_i} \int_{X_j} P_t(i:A)\ P_t(j:B)\ g(X_i)g(X_j)\ dX_j dX_i$$

$$= \int_{X_i} P_t(i:A)\ g(X_i)\ dX_i \cdot \int_{X_i} P_t(j:B)\ g(X_j)\ dX_j$$

$$= S_i \cdot S_j \tag{2.33}$$

and the sequential prediction structure becomes

$$S_i = \int_{X_i} P_t(i:A)\ g(X_i)\ dX_i \tag{2.34}$$

$$S_{ij} = \int_{X_j} P_t(j:B)\ g(X_j)\ dX_j \cdot S_i \tag{2.35}$$

and similarly, for the j, i sequence.

The importance of this result is that at the aggregate level, the prediction of expected volumes in a multi-dimensional choice situation must take account of prior chioces, even if they are structurally independent for the individual behavioral unit, if they are not statistically independent in the population. That is, if the variables influencing the choices are not statistically independent in the aggregate population, the two dimensional shares must be predicted either jointly or sequentially using a conditional distribution of the choice variables.

This discussion indicates the need to take account of the multi-dimensional choice context of a choice for which aggregate predictions are desired. If the choice is joint with other choices the prediction must be based on the joint choice probabilities. When the choice structure is recursive or independent,the structural choice functions may be combined and used as a joint choice function or they may be used sequentially with a conditional distribution of choice influencing variables which represents the effect of prior choices which define the group for which the further prediction is desired.

The result can be used to define the relationship between the aggregate sequential prediction structure of the UTMS type model where the aggregate shares are

$$S_i = \frac{V_i}{T_i} \tag{2.36a}$$

$$S_{d/i} = \frac{V_{id}}{V_i} \tag{2.36b}$$

$$S_{m/id} = \frac{V_{idm}}{V_{id}} \tag{2.36c}$$

$$S_{r/idm} = \frac{V_{idmr}}{V_{idm}} \tag{2.36d}$$

The product of these shares gives the total share of travel:

$$\frac{V_{idmr}}{T_i} = S_i \cdot S_{d/i} \cdot S_{m/id} \cdot S_{r/idm} \tag{2.37}$$

The underlying choice process which parallel this structure is:

$$P_t(f) \tag{2.38a}$$

$$P_t(d/f) \tag{2.38b}$$

$$P_t(m/f,d) \tag{2.38c}$$

$$P_t(r/f,d,m) \tag{2.38d}$$

The product of the marginal choice probability and the conditional choice probabilities give the joint choice probability:

$$P_t(f,d,m,r) = P_t(f)\cdot P_t(d/f)\cdot P_t(m/f,d)\cdot P_t(r/f,d,m) \qquad (2.39)$$

Using the results of the preceding discussion, the relationships between the aggregate shares and the disaggregate choice probabilities, assuming that the frequency choice is 0, or 1, are:

$$S_i^f = \int_X P_t(f)\ g(X)\ dX \qquad (2.40a)$$

$$S_{d/i}^f = \int_X P_t(d/f)\ g(X/f)\ dX \qquad (2.40b)$$

$$S_{m/id}^f = \int_X P_t(m/f,d)\ g(X/f,d)\ dX \qquad (2.40c)$$

$$S_{r/idm}^f = \int_X P_t(r/f,d,m)\ g(X/f,d,m)\ dX \qquad (2.40d)$$

When the frequency choice is not constrained the corresponding set of equations is:

$$S_i = \sum_f f\cdot S_i^f \qquad (2.41a)$$

$$S_{d/i} = \sum_f f\ S_i^f\ S_{d/i}^f/S_i \qquad (2.41b)$$

$$S_{m/id} = \sum_f f\ S_i^f\ S_{d/i}^f\ S_{m/id}^f/S_{d/i} \qquad (2.41c)$$

$$S_{r/idm} = \sum_f f\, S_i^f\, S_{d/i}^f\, S_{m/id}^f\, S_{r/idm}^f / S_{m/id} \tag{2.41d}$$

If the conditional choice probabilities for destination, mode and route are choice independent and statistically independent of the non-zero frequency we have:

$$S_i = \sum_f f\, S_i^f \tag{2.42a}$$

$$S_{d/i} = S_{d/i}^f \tag{2.42b}$$

$$S_{m/id} = S_{m/id}^f \tag{2.42c}$$

$$S_{r/idm} = S_{r/idm} \tag{2.42d}$$

In all of these cases the sequential aggregate shares must be determined from the conditional choice probabilities and the conditional distributions of choice influencing variables.

2.10 Summary

The need for aggregate travel predictions is contrasted with the disaggregate based theory of travel choice behavior. The problems associated with the formulation and estimation of aggregate relationships which are consistent with underlying disaggregate choice models is described.

The structure of an aggregated prediction model is defined in response to the need to provide aggregate travel predictions which are consistent with the relevant underlying individual travel choice behavior. This structural framework includes three components:

- the disaggregate choice model,
- the distribution of choice influencing characteristics, and
- the aggregation procedure.

A range of different distribution representations and aggregation procedures is described. Only one of these, complete enumeration of data and explicit summation of choice probabilities, is consistent with relevant theories of travel behavior. However, this approach places extreme demands on the prediction of independent variables. Criteria are proposed for the evaluation of alternative procedures which have less extreme variable prediction requirements.

The implications of the aggregated model structure for prediction in multi-dimensional choice situations is described. This description indicates the need to recognize statistical relationships between variables which influence choices in the prediction group as well as the structural relations between these choices for different individuals.

CHAPTER III

ANALYSIS OF AGGREGATION BIAS

3.1 Introduction

The aggregated prediction model, described in Chapter II, includes three main components: a disaggregate choice model, a representation of the distribution of choice influencing characteristics, and an aggregation procedure. The aggregation procedure transforms the disaggregate choice model and the distribution of choice influencing characteristics into predictions of aggregate travel behavior. The enumeration procedure is the theoretically correct aggregation procedure. All other procedures are approximate and introduce error into the aggregate predictions.

The purpose of this chapter is to study the prediction error of three approximate aggregation procedures. These are the naive, statistical differentials and classification procedures. This is accomplished by comparison of prediction by each of these methods with the true aggregate predictions based on complete knowledge of the distribution of choice influencing characteristics. The analysis identifies and describes the conditions under which each method will have zero bias, the direction of bias when it occurs and the conditions which result in maximum bias. The analysis includes comparison of share predictions and share sensitivity or slope predictions. The analysis of share prediction is concerned with the error in the prediction of shares for a single set of conditions. The analysis of share sensitivity or slope predictions is concerned with the error in the differences in the prediction of shares between two alternative sets of conditions such as with or without a proposed highway link, free transit versus fare paid transit, with or without a CBD parking charge, etc.

The emphasis placed on these three procedures, naive, statistical differentials and classification, is due to their conceptual simplicity, the ease with which they can be used in practice and the modest input data requirement necessary for their use. The naive prediction, in particular has input data and implementation requirements which are essentially the same as the requirements of presently used models. The statistical differentials and classification methods have only slightly greater data input and implementation requirements. For these reasons, any of these procedures would be highly desirable for use in prediction when they do not contain significant errors as compared to alternative procedures which are more difficult to understand, more difficult to use and require more extensive representation of the distribution of choice influencing characteristics.

The bias in prediction for the naive procedure is analyzed first for different conditions on the structure of the utility distributions, the mean differences in utility between pairs of alternatives and the number of alternatives in the choice set. The analysis of the bias of naive predictions includes bias in share prediction for two alternatives (Section 3.2), bias in share prediction for multiple alternatives

(Section 3.3), bias in share sensitivity prediction for binary and multiple choice situations (Section 3.4) and bias in binary choice prediction when the net utility between alternatives is normally distributed (Section 3.5).

The bias of the statistical differentials procedure (Section 3.6) and the classification procedure (Section 3.7) is examined as extensions to the analysis of bias in naive predictions.

The chapter concludes with summarization of the results obtained (Section 3.8)

3.2 Bias in Share Prediction by the Naive Method for Binary Choice Case

The approach of this anlaysis is to, first, establish the formal statement of bias in naive prediction; second, apply this statement to the analysis of bias in the case where net utility values (difference in utility between two alternatives) are symmetrically distributed in the population; and, finally, generalize the analysis to a wider range of distributions of net utility values.

The analysis of the binary choice case is approached by use of the net values of utility between the two alternatives rather than the utility values for each alternative. The structure of independent random utility models with additive disturbance is such that it is possible to formulate the binary choice model in terms of net utility:

$$V = V_i - V_j \tag{3.1}$$

where V_i, V_j ..... are the utilities of alternatives i and j, respectively, and

V ..... is the net utility difference between alternatives i and j.

The analysis of bias makes use of two important properties of commonly used choice functions. First, choice functions are monotonic in utilities. That is, the probability of choosing an alternative increases with increasing value of its own utility and decreases with increasing value of the utility of all other alternatives. Second, the choice function is symmetric with respect to alternatives. That is, the choice function is structurally identical for each alternative. In the binary choice case this establishes that:

$$f(V) = 1 - f(-V) \tag{3.2}$$

where $f(V)$ ..... is the disaggregate choice function.

The aggregate share for a binary choice set is:

$$S_i = \frac{1}{T} \sum_{t \in T} f^i(V_t) \tag{3.3}$$

where S_i ..... is the aggregate share choosing alternative i,

$f^i(V_t)$ ..... is the choice probability for individual t choosing alternative i when,

V_t ..... is the net utility of alternative i relative to the remaining alternative,

T ..... is both the set of individuals in the group and their number, and

$\sum_{t \in T}$ ..... indicates summation over all members of the group T.

When the distribution of the net utilities is known and can be expressed in terms of a density function, the aggregate share is also given by:

$$S_i = \int_{-\infty}^{\infty} f^i(V)\, g(V)\, dV \tag{3.4}$$

where $g(V)$ ..... is the density function of the distribution of net utility values.

This expression for aggregate share is used in the following discussion as it is more mathematically tractable than the summation.

The naive share is:

$$S_{iN} = f^i(\overline{V}) \tag{3.5}$$

where S_{iN} ..... is the aggregate share for alternative i predicted by the naive method, and

$f(\overline{V})$ ..... indicates that $f(V)$ is evaluated at $V = \overline{V}$.

The bias of the share predicted by the naive method is:

$$B_N^i = S_{iN} - S_i$$

$$= f^i(\overline{V}) - \int_{-\infty}^{\infty} f^i(V)\, g(V)\, dV \tag{3.6}$$

where B_N^i ..... is the bias in predicting the share for alternative i by the naive method.

The bias can also be expressed in terms of the bias in prediction for the individuals included in the aggregate group, defined by:

$$B^i_{Nt} = f^i(\bar{V}) - f^i(V_t) \tag{3.7}$$

This is the bias in the prediction for an individual t when he is assumed to have average net utility in place of individual net utility.

The total bias in prediction can be expressed in terms of individual bias as:

$$\begin{aligned} B^i_N &= \int_{-\infty}^{\infty} B^i_{Nt}\; g(V)\; dV \\ &= \int_{-\infty}^{\infty} \left[f^i(\bar{V}) - f^i(V)\right] g(V)\; dV \end{aligned} \tag{3.8}$$

The argument of this integral is the bias in prediction for a single member of the aggregate group with true net utility, V, represented as having average net utility, $\bar{V}$, weighted by the density of the net utility distribution for each individual net utility value.

Based on the monotonic property of the choice function the individual bias is positive when true utility is less than average utility and negative when true utility is greater than average net utility. Zero bias occurs when the positive bias from individuals with net utility less than the average is balanced by the negative bias from individuals with higher than average net utility.

The bias in share prediction by the naive procedure is first analyzed for the condition where the distribution of net utilities is symmetric around its mean, that is:

$$g(V_D) = g(-V_D) \tag{3.9}$$

where V_D ..... is defined by

$$V_D = V - \overline{V} \tag{3.10}$$

The symmetry condition is placed on the distribution of net utility values. The distribution of utility values for each of the alternatives may have any shape provided that the distribution of net utilities satisfies the symmetry condition (Equation 3.9).

The expression for bias (Equation 3.6) is reformulated in terms of net utility differences from the mean (Equation 3.10) to obtain:

$$B_N^i = f^i(\overline{V}) - \int_{-\infty}^{\infty} f^i(V_D+\overline{V})\ g(V_D)\ dV_D \tag{3.11}$$

which can be reformulated as follows:

$$\begin{aligned} B_N^i &= f^i(\overline{V}) - \int_{-\infty}^{0} f^i(V_D+\overline{V})\ g(V_D)dV_D - \int_{0}^{\infty} f^i(V_D+\overline{V})\ g(V_D)\ dV_D \\ &= f^i(\overline{V}) - \int_{0}^{\infty} f^i(-V_D+\overline{V})\ g(V_D)\ dV_D - \int_{0}^{\infty} f^i(V_D+\overline{V})\ g(V_D)\ dV_D \\ &= f^i(\overline{V}) - \int_{0}^{\infty} \left[1-f^i(V_D-\overline{V})\right]\ g(V_D)dV_D - \int_{0}^{\infty} f^i(V_D+\overline{V})g(V_D)\ dV_D \end{aligned}$$

$$= f^i(\overline{V}) - \int_0^{\infty} g(V_D)dV_D - \int_0^{\infty} \left[f^i(V_D+\overline{V})-f^i(V_D-\overline{V})\right] \; g(V_D)\; dV_D$$

$$= f^i(\overline{V}) - \frac{1}{2} - \int_0^{\infty} \left[f^i(V_D+\overline{V})-f^i(V_D-\overline{V})\right] g(V_D)\; dV_D \qquad (3.12)$$

This expression is equal to zero when average net utility is zero, is positive when average net utility is positive and negative when average net utility is negative.* Further the magnitude of positive bias at some positive value of mean net utility is identical to the magnitude of negative bias when the negative value of mean net utility has the same magnitude:

$$B_N^i(-\overline{V}) = f^i(-\overline{V}) - \frac{1}{2} - \int_{-\infty}^{\infty} \left\{f^i[V_D+(-\overline{V})] - f^i[V_D-(-\overline{V})]\right\} g(V_D)dV_D$$

---

*When $\overline{V} = 0$, $f(\overline{V}) = \frac{1}{2}$; so $f(\overline{V}) - \frac{1}{2} = 0$, and $f(V_D+\overline{V}) - f(V_D-\overline{V}) = f(V_D) - f(V_D) = 0$ so the integral is zero also. When $\overline{V} > 0$, $f(\overline{V}) > \frac{1}{2}$ so $f(\overline{V}) - \frac{1}{2} > 0$, $f(V_D+\overline{V}) - f(V_D-\overline{V}) > 0$ so the integral is greater than zero, furthermore $\text{Max}[f(V_D+\overline{V}) - f(V_D-\overline{V})] = [f(\overline{V})-f(-\overline{V})] = [2f(\overline{V}) - 1]$ and $\int_0^{\infty} [2f(\overline{V}) - 1]\; g(V_D)\; dV_D = f(\overline{V}) - \frac{1}{2}$ so $\int_0^{\infty} [f(V_D+\overline{V}) - f(V_D-\overline{V})]\; g(V_D)\; dV_D < f(\overline{V}) - \frac{1}{2}$ so the total value of the bias expression is positive. For $\overline{V} < 0$ the opposite result occurs.

$$= \left[1-f^i(\overline{V})\right] - \frac{1}{2} - \int_{-\infty}^{\infty} \left[f^i(V_D-\overline{V}) - f^i(V_D+\overline{V})\right] g(V_D)dV_D$$

$$= - f^i(\overline{V}) + \frac{1}{2} + \int_{\infty}^{\infty} \left[f^i(V_D+\overline{V}) - f^i(V_D-\overline{V})\right] g(V_D)\, dV_D$$

$$= - B_N^i\,(\overline{V}) \qquad (3.13)$$

where $B_N^i\,(-\overline{V})$ ..... is the bias in naive share prediction with net utility mean equal to $-\overline{V}$, and

$B_N^i\,(\overline{V})$ ..... is the same for net utility mean equal to $\overline{V}$.

An intuitive understanding of this result is obtained by examination of Figure 3.1, Bias in Individual Share Estimates When Utility Mean is Zero, and Figure 3.2, Bias in Individual Share Estimates When Utility Mean is Greater than Zero. Both figures are plots of individual bias (Equation 3.7) for different values of mean net utility.* The share bias in naive prediction is obtained by integrating the area under this function weighted by the distribution of net utilities (Equation 3.8). If the distribution of net utilities is symmetric

*The function plotted is the choice probability for an individual with utility equal to the mean net utility (which is constant) less the choice probability for the individual with his own utility (which maps out the individual choice function). The difference is an inversion of the individual choice function located so that the bias is zero for an individual with net utility equal to mean net utility.

FIGURE 3.1

BIAS IN INDIVIDUAL SHARE ESTIMATES WHEN UTILITY MEAN IS ZERO

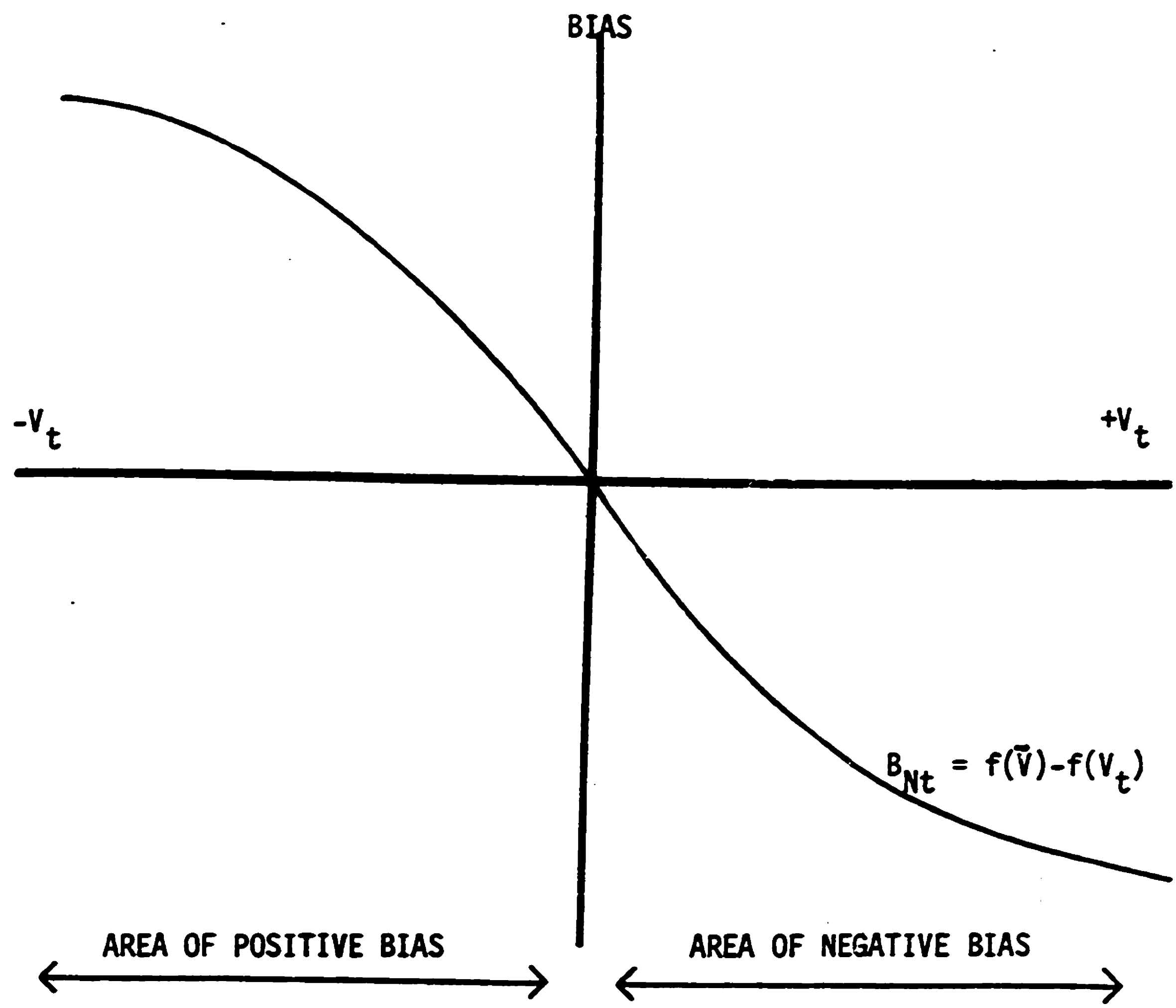
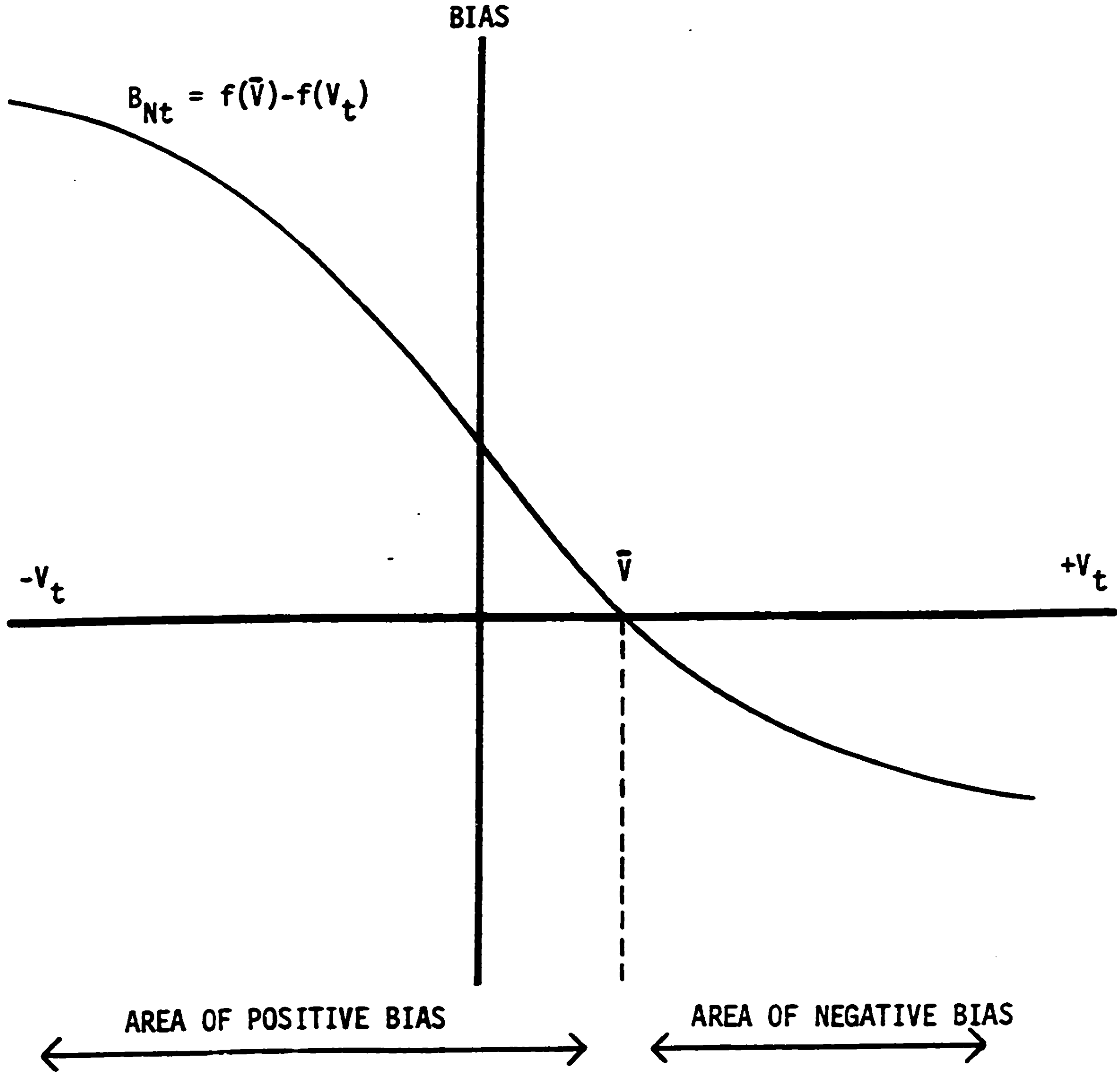


FIGURE 3.2

BIAS IN INDIVIDUAL SHARE ESTIMATES WHEN UTILITY MEAN IS GREATER THAN ZERO



the integral of net bias is zero when mean net utility is zero (Figure 3.1). In this case the weighting of equal magnitude positive and negative bias at equal distances from the zero bias point will be equal (due to symmetry of the distribution around the zero bias point) and therefore will exactly balance. When the average net utility is positive (Figure 3.2) the weighting of higher magnitudes of positive bias and smaller magnitudes of negative bias at equal distances from the new zero bias point are equal and the share bias is positive. Similarly, the share bias is negative when the mean net utility is negative.

The location of points of maximum bias are identified by taking the derivative of the bias function (Equation 3.11):

$$\frac{\partial B_N^i}{\partial \bar{V}} = \frac{\partial f^i(V)}{\partial \bar{V}} - \frac{\partial}{\partial \bar{V}} \left\{ \int_{-\infty}^{\infty} f^i(V_D+\bar{V})\ g(V_D)\ dV_D \right\} \tag{3.14}$$

where $\frac{\partial}{\partial \bar{V}}$ ..... indicates that the partial derivative is taken with respect to mean net utility.*

In Section 3.4 it is shown that this expression is equivalent to the expression for bias in share sensitivity and has two zero value points. When the distribution of net utilities is symmetric these points occur at equal magnitudes of positive and negative mean net utility. The

*Individual utilities expressed by $V_D+\bar{V}$ are all affected equally by changes in mean net utility, $\bar{V}$, as there are no changes in the individual deviations from mean net utility, V_D.

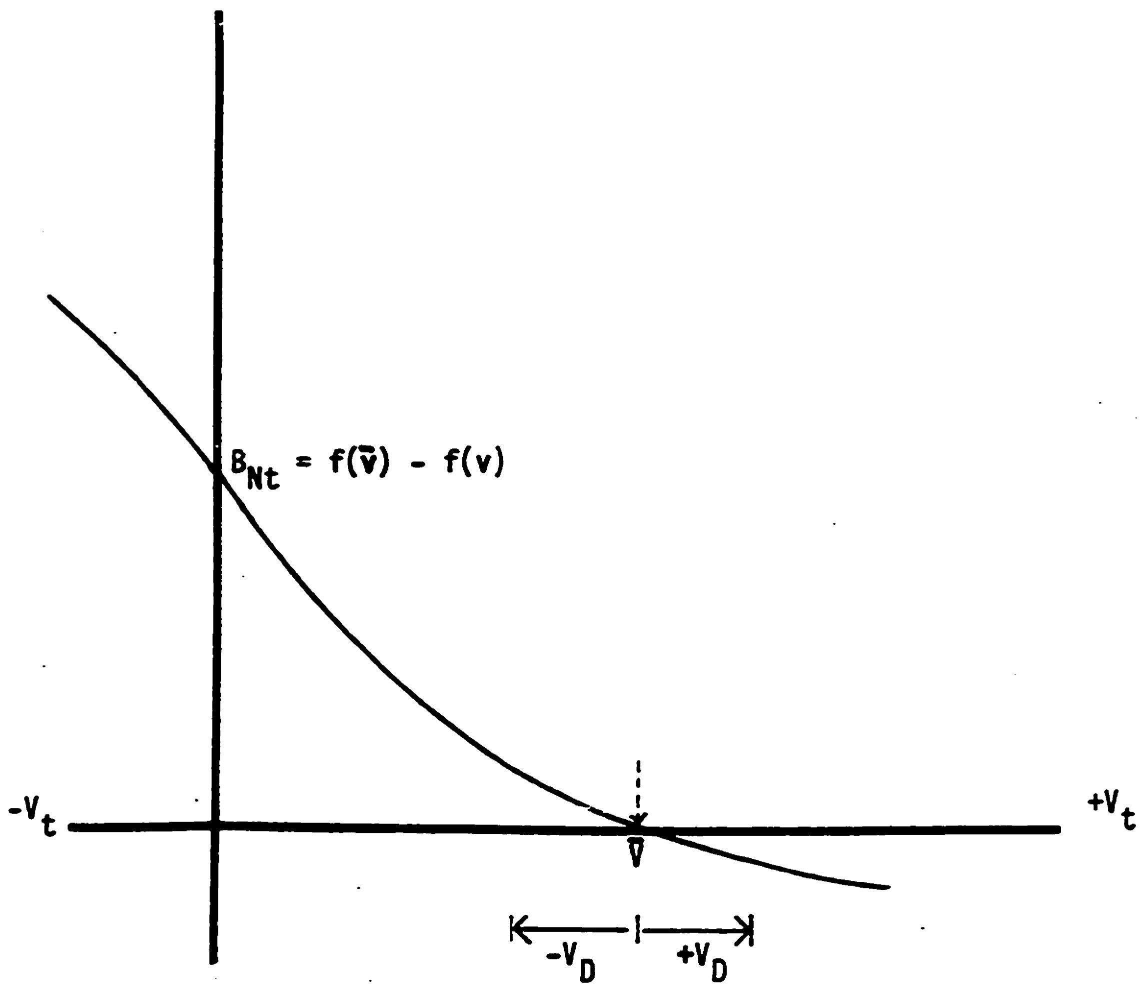
existence of maximum and minimum points indicates that the positive or negative share bias increases in magnitude to some maximum value and declines thereafter. As mean net utility becomes large (relative to the distribution of net utility deviations) the magnitude of bias becomes small (see Figure 3.3).

The entire character of the distribution of net utilities enters into the bias expression in the $g(V_D)$ term. The characteristics of the distribution can be expressed by its variance and shape. The effect of variance for a particular distributional shape (the normal distribution) is considered in Section 3.5. For the present discussion the only shape characteristic considered is the skewness of the distribution. This is done under the assumption that common distributions of net utilities will not differ substantially in other aspects of shape.

Two characteristics of skewness are important. First, for a positively skewed distribution the cumulative distribution function from minus infinity to its mean is greater than one half with the opposite effect for a negatively skewed distribution. Second, a positively skewed distribution has higher density at extreme positive deviation from the mean than at extreme negative deviation, with the opposite effect for a negatively skewed distribution. Since the individual bias function has diminishing slope with increasing deviation from the mean over most of its range (over the entire range when average net utility is zero as in Figure 3.1) the effect of mass concentrated away from the tail is most important in this situation. The effect of

FIGURE 3.3

LIMITING VALUE OF SHARE BIAS APPROACHES ZERO AS MEAN NET UTILITY BECOMES LARGE



skewness is to weight the bias away from the tail of the distribution more highly than the bias in the direction of the tail. Thus, the existence of positive skewness (right tail) produces greater positive bias or lower negative bias than would occur with a symmetric distribution with the same mean net utility value and negative skewness has the opposite effect. Furthermore, the magnitude of the affect of skewness is larger when the skewness is greater.

The skewness and mean value effects described earlier for the symmetric distribution interact so that positive mean and positive skewness (or negative mean and negative skewness) have compounding effects on the magnitude of bias while positive mean and negative skewness (or negative mean and positive skewness) have offsetting effects. Some conditions of offsetting effects create points of zero bias with negative means and positive skewness and vice versa. An algebraic increase in the mean net utility or the skewness from a zero bias point results in positive bias and an algebraic decrease results in negative bias:

The results of this section (the binary choice case) are:

1. For any distribution of net utility values around the mean, $g(V_D)$, there exists one and only one value of average net utility for which the naive method has zero bias.
2. The mean value of net utility for zero bias is negatively monotonically related to the skewness of the net utility distribution and is zero when skewness is zero.

3. An algebraic increase in the mean value of net utility or skewness of a distribution from the zero bias point creates a condition of positive bias. The opposite effect occurs when there is a decrease in mean or skewness.
4. As the value of average net utility increases (decreases) from the point of zero bias a point of maximum positive (negative) bias is reached after which the magnitude of bias declines.

3.3 Bias in Share Prediction by the Naive Method for Multiple Choice Case

The aggregate share for an alternative from a multiple choice set is:

$$S_i = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f^i(V_1,\ldots,V_R)\; g(V_1,\ldots V_R)\; dV_R \ldots dV_1 \qquad (3.15)$$

where S_i ..... is the aggregate share choosing alternative i,

$f^i(\ )$..... is the choice function for the i alternative

V_j ..... is the utility for alternative j, and

$g(\ )$ ..... is the multivariate distribution of utility values in the aggregate group.

The naive estimate of the aggregate share for an alternative in a multiple choice set is:

$$S_{iN} = f^i(\bar{V}_1,\ldots,\bar{V}_R) \qquad (3.16)$$

where $\bar{V}_j$ ..... is the average utility for alternative j.

The bias by the naive method is:

$$B_N^i = S_{iN} - S_i$$

$$= f^i(\bar{V}_1,\ldots\bar{V}_R) - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f^i(V_1,\ldots V_R) g(V_1,\ldots V_R) dV_R \ldots dV_1 \qquad (3.17)$$

The symmetric property of the choice function, in the multiple choice case is based on the use of the same functional form for all alternatives as follows:

$$f^i(V_{1,\ldots,}V_R) = f(V_i, V^i) \tag{3.18a}$$

$$f^j(V_{,\ldots,}V_R) = f(V_j, V^j) \tag{3.18b}$$

where $f^i(\ )$ ..... is the choice function for alternative i,

$f(\ )$ ..... is the common structural choice function,

V^i ..... is the vector of utility values excluding V_i.

That is, the choice function for alternative i is structurally identical to the choice function for alternative j for all i, j pairs. The common structural function, f( ), is monotonically increasingly with its first argument and monotonically decreasing with the remaining arguments. This property has been called simple scalability by Tversky (1972).

The analysis of the multiple choice case builds on the results for the binary choice case when the choice function, $f^i(V)$,* is a strict utility model** (Luce and Suppes, 1965) which has the form:

---

*This notation of the choice function will replace the complete form used in Equation 3.15 where possible to simplify notation. Similarly g(V) will replace $g(V_{1,\ldots}V_R)$.

**The commonly used logit function is a strict utility model.

$$f^i(V) = \frac{V_i}{\sum_j V_j} \tag{3.19}$$

One property of the strict utility models is that the relative probability of two alternatives is independent of other alternatives. When this condition exists, the choice function may be written as:

$$f^i(V) = \left[\frac{V_i}{V_i+V_k}\right]\left[\frac{V_i+V_k}{\sum V_j}\right] \tag{3.20}$$

and the expression for aggregate share can be written:

$$S_i = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{V_i}{V_i+V_k}\right)\int_{-\infty}^{\infty}\left(\frac{V_i+V_k}{\sum V_j}\right) g(V)\, dV^{ik} dV_k dV_i \tag{3.21}$$

where $\int_{-\infty}^{\infty} \ldots dV^{ik}$ implies integration over the utility distribution for all but the i and k alternatives.

If the distributions of alternative specific utility values are independent:

$$S_i = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{V_i}{V_i+V_k}\right) g(V_i)g(V_k)\int_{-\infty}^{\infty}\frac{V_i+V_k}{\sum V_j}\, g(V^{ik})dV^{ik}dV_k dV_i$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{V_i}{V_i+V_k}\right) g(V_i)g(V_j)Z_{ik}dV_k dV_i \tag{3.22}$$

where $$Z_{ik} = \int_{-\infty}^{\infty} \left(\frac{V_i + V_k}{\sum V_j}\right) g(V^{ik})\, dV^{ik} \quad . \tag{3.23}$$

Z^{ik} may be evaluated by consideration of the composite alternative, i and k, and scaling the utility values for every individual so that the utility of the composite alternative is constant across individuals. When the constant is chosen equal to one,

$$Z_{ik} = S_i + S_k \tag{3.24}$$

which is independent of the relative utility of alternatives i and k, so Equation 3.22 becomes:

$$S_i = \left\{ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{V_i}{V_i + V_k}\right) g(V_i) g(V_k) dV_k dV_i \right\} (S_i + S_k) \tag{3.25}$$

and similarly for the share choosing k:

$$S_k = \left\{ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{V_k}{V_i + V_k}\right) g(V_i) g(V_k) dV_k dV_i \right\} (S_i + S_k) \tag{3.26}$$

so the relative shares choosing i and k are independent of the share choosing all other alternatives. That is:

$$\frac{S_k}{S_k} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{V_k}{V_i+V_k}\right) g(V_i)\ g(V_k)\ dV_k\ dV_i}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{V_i}{V_i+V_k}\right) g(V_i)\ g(V_k)\ dV_k\ dV_i} \tag{3.27}$$

The relative shares choosing i and k predicted by the naive method are also independent of predictions for the other shares since:

$$\frac{S_{kN}}{S_{iN}} = \frac{\bar{V}_k / \sum \bar{V}_j}{\bar{V}_i / \sum \bar{V}_j}$$

$$= \frac{\bar{V}_k}{\bar{V}_i} \tag{3.28}$$

These share relationships (Equation 3.27 and 3.28) apply to all choice pairs.

In addition, the share choosing i and the predicted share choosing i by the naive method are fully defined by the relative shares between pairs of alternatives:

$$S_i = \frac{1}{\sum_j S_j / S_i} \tag{3.29}$$

$$S_{iN} = \frac{1}{\sum_j S_{jN}/S_{iN}} \tag{3.30}$$

The no bias condition for any share can be defined from these relationships. Two classes of no bias conditions exist. The first class includes situations where all of the relative shares between probabilities are unbiased. Under these conditions all shares are unbiased. That is, if:

$$\frac{S_{jN}}{S_{iN}} = \frac{S_j}{S_i} \quad \forall\ j \tag{3.31a}$$

then, from Equations 3.29 and 3.30,

$$S_{iN} = S_i \tag{3.31b}$$

and,

$$S_{jN} = \frac{S_{jN}}{S_{iN}} \cdot S_{iN}$$

$$= \frac{S_j}{S_i} \cdot S_i$$

$$= S_j \qquad \forall\ j \tag{3.31c}$$

Thus, when an alternative is unbiased because all of its relative shares are unbiased, all shares are unbiased. The situation corresponding to the identical mean-symmetric distribution in the binary choice case leads to:

$$\frac{S_j}{S_i} = 1 \tag{3.32a}$$

$$S_i = \frac{1}{\sum S_j/S_i} = \frac{1}{R} \tag{3.32b}$$

$$S_j = \frac{S_j}{S_i} \; S_i = \frac{1}{R} \qquad \forall\, j \tag{3.33c}$$

and,

$$\frac{S_{jN}}{S_{iN}} = 1 \tag{3.33d}$$

$$S_{iN} = \frac{1}{\sum S_{jN}/S_{iN}} = \frac{1}{R} \tag{3.33e}$$

$$S_{jN} = \frac{S_{jN}}{S_{iN}} \; S_{iN} = \frac{1}{R} \qquad \forall\, j \tag{3.33f}$$

which is a direct extension of the binary choice case (R=2). This class also includes cases where the relative shares between pairs of alternatives are unbiased due to the interacting effects of non-zero net mean values for pairs of utilities and non-symmetric distributions of net utility between pairs.

Furthermore, as an algebraic increase (decrease) in mean net utility or skewness of net utility distributions for alternative i relative to one or more other alternatives creates positive (negative) bias in the binary share for alternative i, the increase (decrease) will create a positive (negative) bias in the multiple choice share for alternative i as can be seen in Equations 3.29 and 3.30.

The other class of situations which leads to no bias in prediction of the share choosing alternative i occurs when the bias in different pair share ratios are offsetting. If:

$$\frac{S_{jN}}{S_{iN}} + \frac{S_{kN}}{S_{iN}} = \frac{S_j}{S_i} + \frac{S_k}{S_i} \tag{3.34a}$$

and

$$\frac{S_{lN}}{S_{iN}} = \frac{S_l}{S_i} \qquad \forall\ l \neq j,k \tag{3.34b}$$

then

$$S_{iN} = \frac{1}{\sum_j S_{jN}/S_{iN}} = \frac{1}{\sum_j S_j/S_i} = S_i \tag{3.34c}$$

but,

$$S_{jN} = \frac{S_{jN}}{S_{iN}} S_{iN} \neq \frac{S_j}{S_i} S_i = S_j \tag{3.34d}$$

That is, when the share predicted to choose alternative i by the naive method is unbiased due to offsetting biases in relative shares, those shares which are biased relative to share i will be biased in share predicted. In this case, as before, the effect of increasing (decreasing) the mean net utility or the positive skewness of the net utility distribution between alternative i and one or more other alternatives from a set of zero bias values creates a condition of positive (negative) bias.

Furthermore, for both classes of no bias in share i conditions offsetting changes in the net mean utility and the skewness of the net distribution of alternative i relative to other alternatives leads to other zero bias points with the same bias characteristics for other shares as before the offsetting changes.

The results of this section are:*

1. For a set of independent distributions of utility values around their means and given mean values of the utilities for other alternatives, there exists one and only one mean utility value for an alternative for which the naive share

*The results for the binary choice case may be derived as a special case of these results.

prediction is unbiased.

2. Zero bias may be obtained for any alternative share prediction with or without zero bias for other share predictions.
3. The mean value of utility for zero bias of an alternative is negatively monotonically related to the skewness of the distribution of utilities for that alternative. The mean value relationship is such that the zero bias share for an alternative is the inverse of the number of alternatives when the alternative utility distributions are identically distributed.
4. An algebraic increase (decrease) in the mean value or skewness of a utility distribution from a zero bias point for the corresponding share leads to a condition of positive (negative) bias.
5. As the value of mean utility for an alternative increases (decreases) from the zero bias point, a point of maximum positive (negative) bias is reached beyond which the magnitude of bias declines monotonically.

3.4 Bias in Share Sensitivity Prediction by the Naive Method

The discussion to this point has been concerned with bias in the prediction of shares of an aggregate group choosing each of a set of alternatives under a given set of conditions. It is also useful to analyze the bias in the differences in predictions between two different sets of conditions. The need for this comparative prediction occurs whenever there is a need to predict a change in travel behavior as a result of a change in variables.

The true difference in share in a binary choice case with a discrete change is:

$$D_i = \int_{-\infty}^{\infty} f^i(V_A)\ g(V_A)\ dV_A - \int_{-\infty}^{\infty} f^i(V_B)\ g(V_B)\ dV_D \tag{3.35}$$

where D_i ..... is the difference in share choosing alternative i for policy A (after) as compared to policy B (before).

The difference in share predicted by the naive method is:

$$D_{iN} = f^i(\overline{V}_A) - f^i(\overline{V}_B) \tag{3.36}$$

where D_{iN} ..... is the difference in share choosing i predicted by the naive method.

The bias in difference predicted by the naive method is:

$$BD_N^i = D_{iN} - D_i$$

$$= \left[f^i(\bar{V}_A) - f^i(\bar{V}_B)\right] - \left[\int_{-\infty}^{\infty} f^i(V_A)g(V_A)dV_A - \int_{-\infty}^{\infty} f^i(V_B)g(V_B)dV_B\right] \tag{3.37}$$

or, alternatively,

$$BD_N^i = \int_{-\infty}^{\infty} \left[f^i(\bar{V}_A) - f^i(V_A)\right] g(V_A)\, dV_A - \int_{-\infty}^{\infty} \left[f^i(\bar{V}_B) - f^i(V_B)\; g(V_B)dV_B\right] \tag{3.38}$$

The bias in difference prediction is zero if the bias in each of the point predictions is identical. For the case where the change in net utility is identically additive for all individuals:

$$V_{At} = V_{Bt} + \Delta V \qquad \forall\, t \tag{3.39}$$

and the changes are small:

$$V_{At} = V_{Bt} + dV \qquad \forall\, t \tag{3.40}$$

the sensitivity bias may be expressed by:

$$BS_N^i = \frac{\partial f^i(\bar{V})}{\partial \bar{V}} - \int_{-\infty}^{\infty} \frac{\partial f^i(V)}{\partial \bar{V}}\; g(V)\, dV \tag{3.41}$$

where BS_N^i ..... is the bias in sensitivity by the naive method*, and

$\dfrac{\partial f^i(\overline{V})}{\partial \overline{V}}$ ..... is the first derivative of the disaggregate choice function evaluated at $V = \overline{V}$.

or

$$BS_N^i = \int_{-\infty}^{\infty} \left[\frac{\partial f^i(\overline{V})}{\partial \overline{V}} - \frac{\partial f^i(V)}{\partial \overline{V}} \right] g(V)\, dV \tag{3.42}$$

When the distribution of V is symmetric around the mean, points of equal sensitivity bias occur in pairs. This is shown by expressing net utility in Equation 3.42 as the sum of average net utility and the difference from average net utility:

$$V = V_D + \overline{V} \tag{3.43}$$

to obtain:

$$BS_N^i = \int_{-\infty}^{\infty} \left[\frac{\partial f^i(\overline{V})}{\partial \overline{V}} - \frac{f^i(V_D + \overline{V})}{\partial \overline{V}} \right] g(V_D) \quad dV_D \tag{3.44}$$

---

*Note: Shift from bias in difference prediction to bias in sensitivity follows from consideration of small changes as indicated in Equation 3.40.

From the symmetry property described in Equation 3.2, we obtain:

$$\frac{\partial f(V)}{\partial \overline{V}} = - \frac{\partial f(-V)}{\partial \overline{V}} \tag{3.45}$$

so that Equation 3.44 becomes,

$$BS_N^i = \int_{-\infty}^{\infty} - \left[\frac{\partial f^i(-V)}{\partial \overline{V}} - \frac{\partial f^i(-V_D-\overline{V})}{\partial \overline{V}}\right] g(V_D)\, dV_D \tag{3.46}$$

A further transformation of variables:

$$V_{D'} = -V_D \tag{3.47}$$

leads to,

$$\begin{aligned} BS_N^i &= -\int_{-\infty}^{\infty} \left[\frac{\partial f^i(-\overline{V})}{\partial \overline{V}} - \frac{\partial f^i(V_{D'}-\overline{V})}{\partial \overline{V}}\right] g(V_{D'})\, -dV_{D'} \\ &= \int_{-\infty}^{\infty} \left[\frac{\partial f^i(-\overline{V})}{\partial \overline{V}} - \frac{\partial f^i(V_{D'}-\overline{V})}{\partial \overline{V}}\right] g(V_{D'})\, dV_{D'} \end{aligned} \tag{3.48}$$

A final transformation of variables,

$$V = V_{D'} + (-\overline{V}) \tag{3.49}$$

results in,

$$BS_N^i = \int_{-\infty}^{\infty} \frac{\partial f^i(-\bar{V})}{\partial \bar{V}} - \frac{\partial f^i(V)}{\partial \bar{V}} g\ (V)\ dV \qquad (3.50)$$

which is identical to Equation 3.42 except that the mean net utility value is the negative of the original value.

The occurrance of pairs of points with equal bias in share sensitivity and the condition for zero bias in share sensitivity are illustrated in Figure 3.4, Bias In Individual Sensitivity Estimates When Utility Mean is Zero and Figure 3.5, Bias In Individual Sensitivity Extimates When Utility Mean Is Different From Zero.* These figures, which correspond to Figures 3.1 and 3.2 for individual share bias, are plots of the individual bias in slope estimates:

$$BS_{Nt}^i = \frac{\partial f^i(\bar{V})}{\partial^i V} - \frac{\partial f^i(V_t)}{\partial V} \qquad (3.51)$$

where BS_{Nt}^i ..... is the bias in slope by naive prediction with respect to alternative i for individual t,

for different values of average net utility. The overall bias in sensitivity prediction is obtained by integrating the area under this

*The function plotted is derived from the sensitivity for an individual with utility equal to mean net utility compared to the sensitivity for an individual with any selected value of net utility. The shape of the function is the first derivative of the functions used in Figure 3.1 and 3.2.

FIGURE 3.4

BIAS IN INDIVIDUAL SENSITIVITY ESTIMATES
WHEN UTILITY MEAN IS ZERO

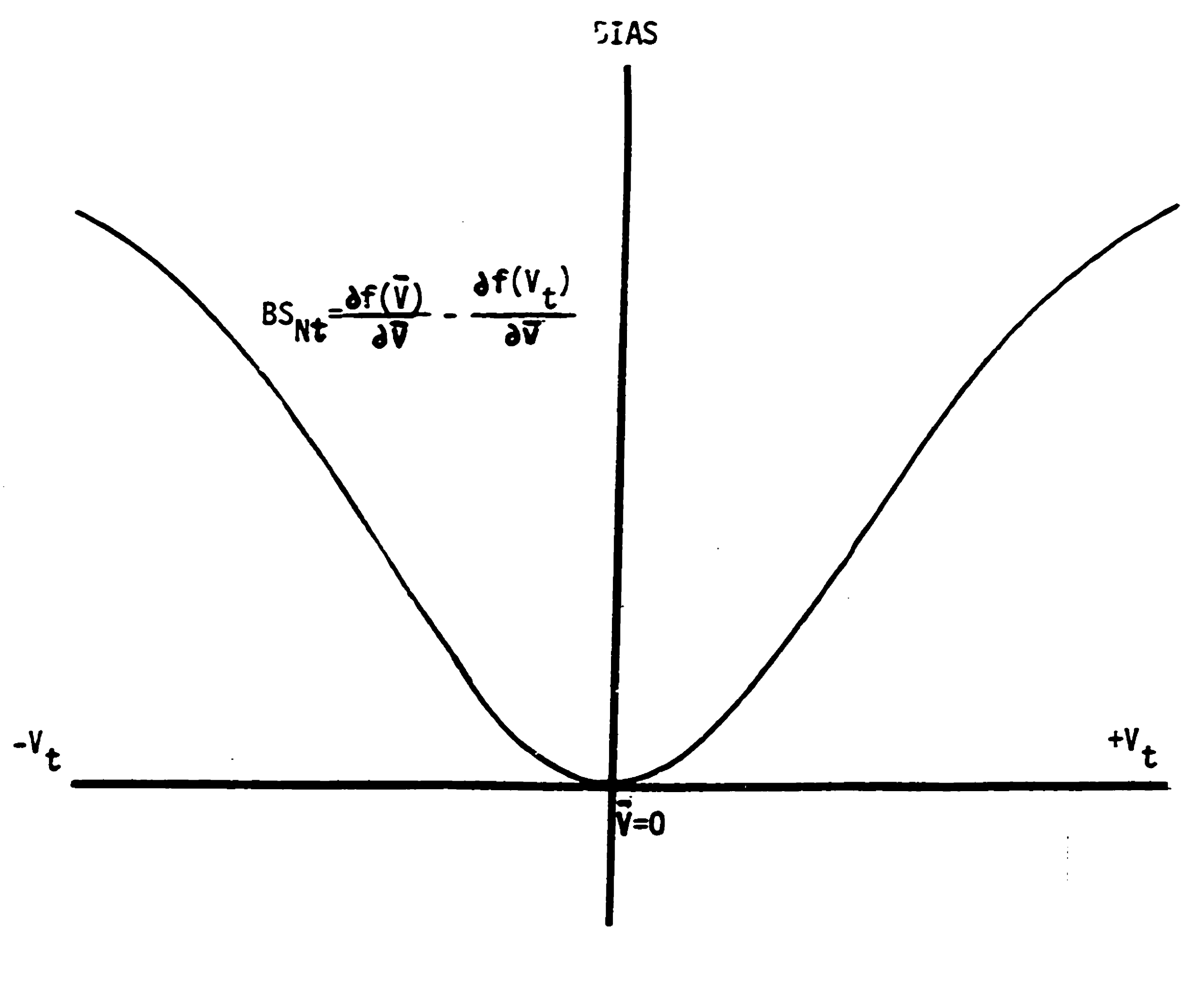
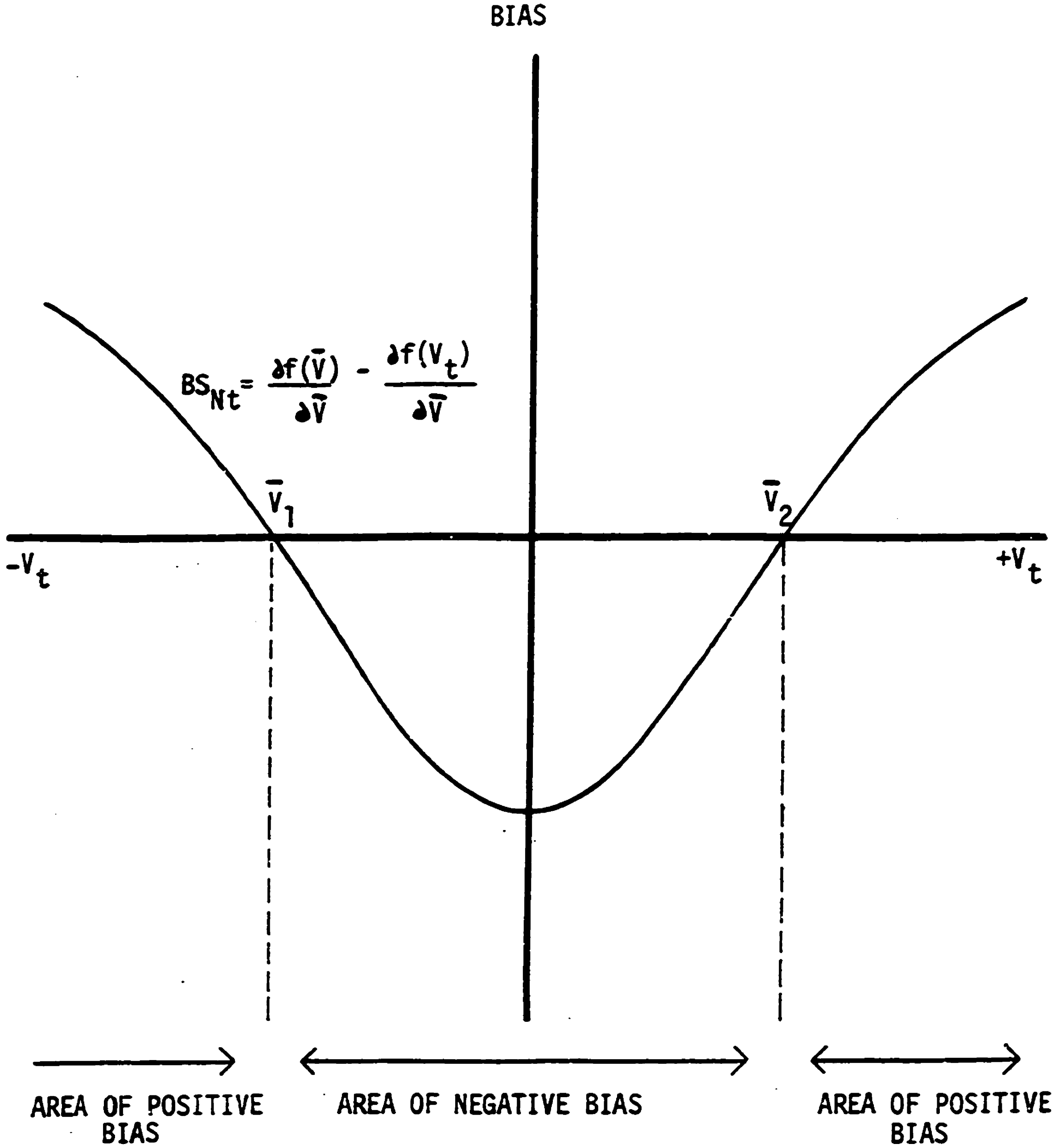


FIGURE 3.5

BIAS IN INDIVIDUAL SENSITIVITY ESTIMATES WHEN UTILITY MEAN IS DIFFERENT FROM ZERO



function weighted by the distribution of net utilities, $g(V_D)$. That is:

$$BS_N^i = \int_{-\infty}^{\infty} BS_{Nt}^i \, g(V_D) \, dV_D \tag{3.52}$$

When the mean net utility is equal to zero (Figure 3.4) the bias in share sensitivity for all individuals is positive. The effect of shifting the mean net utility away from zero is to translate the bias curve downward. Depending on the shape and spread of the distribution a zero slope bias point occurs followed by negative slope bias for further increases in the magnitude of average net utility. When the net utility distribution is symmetric, the zero bias points occur for positive and negative values of average net utility which are equal in magnitude. This is consistent with the comparison between Equations 3.42 and 3.50 made earlier.

The affect of skewness in the distribution is to increase the weight of the bias in the direction opposite the direction of skewness. To see this consider Figure 3.5 and assume that for a symmetric distribution the zero slope bias points are at $\overline{V}_1$ and $\overline{V}_2$. If the symmetric distribution is replaced by a positively skewed distribution the bias at $\overline{V}_1$ is positive and the bias at $\overline{V}_2$ is negative. The positive bias at $\overline{V}_1$ and the negative bias at $\overline{V}_2$ can be offset by a shift to a lower algebraic value of average net utility. This is similar to the offsetting effect of changes in average net utility value and skewness for share prediction and indicates the existence of the same type of monotonic relationship satisfying the zero bias condition except that there will be two points of zero slope bias for each distribution of net utilities. These points are on opposite sides of the point of zero share bias.

Before proceeding further, it is instructive to note the relationship of the condition for zero bias in sensitivity predictions and the conditions for maximum magnitude bias in share predictions. Subject to second order conditions, the points of maximum positive or negative bias in share prediction are identified by taking the derivative of bias with respect to net utility mean (Equation 3.14) and setting it equal to zero. Equation 3.14 is identical to Equation 3.42 and 3.50, which describe share sensitivity bias. Thus, the values of mean net utility for which share sensitivity bias is zero for any distribution are identically the same values for which share bias is at its maximum positive and negative values.

Share sensitivity bias is positive for mean net utility between the two values for which it is zero. Therefore, a maximum positive share sensitivity bias occurs in this range. It is apparent from Figures 3.4 and 3.5 that the maximum positive share sensitivity bias for symmetric distributions occurs for mean net utility equal to zero (the same value of mean net utility for which share bias is zero). When mean net utility is beyond the range of the zero share sensitivity bias points, the share sensitivity bias is negative.

The results obtained for the binary choice case for bias in sensitivity predictions can be extended to the multiple choice case by a series of arguments which parallel the arguments used to extend the results of the analysis of bias in share prediction. That is, the conditions for zero sensitivity bias are alternative specific, demonstrate (two) negative monotonic relationships between mean utility of an alternative and skewness of the distribution of the utility values for the alterntive, occur to either side of the point of zero share bias, and have maximum positive bias at or near the point of zero slope bias.

The results of this section are:

1. For any set of distributions of utility values, $g(V_{1D},\ldots,V_{RD})$ there exist two values of mean utility for an alternative for which the slope prediction by the naive procedure is unbiased.
2. The mean value of utility for an alternative for zero bias in share sensitivity by the naive procedure is negatively monotonically related to the skewness of the distribution of the utility of the alternative. The two zero bias slope points occur on either side (higher and lower mean utility values) of the zero bias share point.
3. An algebraic increase (decrease) in the mean value or skewness of a utility distribution from the zero slope bias point (greater than the zero share bias point) creates a condition of negative (positive) slope bias. The opposite effect occurs at the other zero bias point (less than the zero share bias point).

3.5 Bias in Share and Share Sensitivity Prediction by the Naive Method with Normally Distributed Net Utility Values

Further insight into the characteristics of the naive prediction method and some estimate of the magnitude of bias is obtained by consideration of the case where the utility values - net utility between two alternatives - are normally distributed in the aggregate group. Although this is a special case, evidence of robustness in the representation of the distribution (Chapter IV), indicates that the insights gained are applicable to a range of situations.

McFadden and Reid (1975) have shown that if the disaggregate choice probability is described by the probit function:

$$P_t(i) = \Phi(V_t)$$

$$= \int_{-\infty}^{V_t} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}y^2\right] dy \qquad (3.53)$$

and the net utility in the aggregate group is normally distributed with mean, $\overline{V}$, and variance σ_V^2, the expected share is:

$$S_i = \Phi\left(\frac{\overline{V}}{\sqrt{1+\sigma_V^2}}\right) \qquad (3.54)$$

The corresponding estimate of aggregate share using the naive method is obtained by ignoring the heterogeneity of utility values in the aggregate group or assuming $\sigma_V^2 = 0$:

$$S_{iN} = \Phi(\bar{V}) \tag{3.55}$$

These expressions can be compared to analyze more precisely the bias in share predictions and the bias in share sensitivity predictions. They also can be used to explore the effect of the degree of heterogeneity (variance of net utility) in the population on the value of the expected biases.

The formula for prediction bias is:

$$B_N^i = S_{iN} - S_i = \Phi(\bar{V}) - \Phi\left[\frac{\bar{V}}{\sqrt{1+\sigma_V^2}}\right]$$

$$= \int_{\frac{\bar{V}}{\sqrt{1+\sigma_V^2}}}^{\bar{V}} \left\{\frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}y^2\right]\right\} dy \tag{3.56}$$

Since the absolute value of the argument of the probit in the naive case is always greater in magnitude than for the distributed case, that is,

$$|\bar{V}| > \left|\frac{\bar{V}}{\sqrt{1+\sigma_V^2}}\right| \text{ if } \sigma_V^2 > 0 \tag{3.57}$$

the bias is negative when the share is less than one half and positive when it is greater than one half. The zero bias point is, as discussed before, at mean net utility of zero with equal probabilities for the two choices. In short, the probability estimate obtained by the naive method is always more extreme (further from equal shares) than the estimate which properly accounts for variance in the distribution of net utility values.

As previously discussed, the bias in use of the naive method results from the distribution of the utility values for members of the aggregate group. The naive method gives unbiased aggregate predictions when the groups members have identical utilities. To see this, consider the disaggregate choice function:

$$P_t(i) = f^i\ (V_{1t},\ldots,V_{RN}) \tag{3.58}$$

The homogeneity condition is:

$$V_{it} = \overline{V}_i \qquad \forall t,\ \forall i \tag{3.59}$$

so that,

$$\begin{aligned} S_i &= \frac{1}{T} \sum_t P_t(i) \\ &= \frac{1}{T} \sum_t f^i(V_{1t},\ldots,V_{Rt}) \end{aligned}$$

$$= \frac{1}{T} \sum_t f^i(\overline{V}_1, \ldots, \overline{V}_i, \ldots, \overline{V}_R)$$

$$= f^i (\overline{V}_1, \ldots, \overline{V}_i, \ldots, \overline{V}_R) \tag{3.60}$$

which is the same as the naive prediction.

The way in which increasing heterogeneity, represented by the within group variance of the utility values, affects the bias which results from the naive method is analyzed by taking the derivative of the bias expression (Equation 3.56) with respect to the variance of the net utility distribution:

$$\frac{\partial B_N^*}{\partial \sigma_V^2} = \frac{1}{\sqrt{2\pi}} \left[\exp\left[-\frac{1}{2} \overline{V}^2 / (1 + \sigma_V^2)\right] \right] \left[\frac{1}{2} \frac{\overline{V}}{(1+\sigma^2)^{3/2}}\right] \tag{3.61}$$

This expression is positive when the net utility mean is positive and negative when the net utility mean is negative. Thus, increasing variance of the net utility distributions increases the magnitude of the bias in naive prediction for any net utility mean. This effect is illustrated in Table 3.1, Bias in Naive Prediction for Normally Distributed

*To simplify notation the superscript i will be deleted in the balance of this section.

TABLE 3.1

BIAS IN NAIVE PREDICTION FOR NORMALLY DISTRIBUTED NET UTILITY AS A FUNCTION OF MEAN AND VARIANCE OF THE UTILITY DISTRIBUTION

	ABSOLUTE VALUE OF MEAN NET UTILITY						
VARIANCE OF NET UTILITY	0.0	0.5	1.0	1.5	2.0	2.5	3.0
	A. TRUE SHARE						
1.0	.50	.64	.76	.86	.92	.96	.98
2.0	.50	.61	.72	.81	.88	.93	.96
3.0	.50	.60	.69	.77	.84	.89	.93
∞	.50	.50	.50	.50	.50	.50	.50
	B. NAIVE SHARE PREDICTION						
	.50	.69	.84	.93	.98	.99	.999
	C. BIAS OF NAIVE SHARE PREDICTION (ABSOLUTE VALUE)						
1.0	.0	.05	.08	.07	.06	.03	.02
2.0	.0	.08	.12	.12	.10	.06	.04
3.0	.0	.09	.15	.16	.14	.10	.07
∞	.0	.19	.34	.43	.48	.49	.499

Net Utility as a Function of Mean and Variance* of the Utility Distribution, where the increase in bias with increasing variance of the net utility is given for a range of absolute values of mean net utility. The bias increases with variance and is asymptotic to the values given in the last line of Table 3.1.

The way in which share bias is affected by increasing magnitude of average net utility is studied by taking the derivative of the bias term with respect to average net utility to obtain,

$$\frac{\partial B_N}{\partial \bar{V}} = \frac{1}{\sqrt{2\pi}} \left\{ \exp\left(-\frac{1}{2}\bar{V}^2\right) - \frac{1}{\sqrt{1+\sigma_V^2}} \exp\left[-\frac{1}{2}\bar{V}^2 / (1+\sigma_V^2)\right] \right\} \tag{3.62}$$

which is zero (indicating a maximum bias point for a given variance) when,

$$|\bar{V}| = \left[\frac{1+\sigma_V^2}{\sigma_V^2} \ln(1+\sigma_V^2) \right]^{1/2} \tag{3.63}$$

provided σ^2 is not equal to zero. These values are shown in Table 3.2, Values of Average Net Utility for Maximum Share Bias as a Function of Net Utility Variance with Corresponding Shares and Bias.

---

*Note in analysis of numerical utility values that the relationships between probit and logit functions are such that the utility value of a probit function of 1.0 corresponds approximately to a utility value of a logit function of 0.6 (Cox, 1966).

TABLE 3.2

VALUE OF AVERAGE NET UTILITY FOR MAXIMUM SHARE BIAS AS A FUNCTION OF NET UTILITY VARIANCE WITH CORRESPONDING SHARES AND BIAS

NET UTILITY VARIANCE	NET UTILITY MEAN	NAIVE SHARE	TRUE SHARE	NAIVE BIAS IN SHARE
0.1	1.02	.85	.84	.01
1.0	1.18	.88	.80	.08
2.0	1.28	.90	.77	.12
3.0	1.36	.91	.75	.16

The points of maximum bias in share prediction by the naive method (Table 3.2) occur near the point of maximum curvature of the disaggregate choice function which, for the probit case, is at net utility equal to plus or minus one.

The relationships between true shares and bias of the naive method with mean net utility and net utility variance are depicted in Figure 3.6, Bias in Naive Prediction for Normally Distributed Net Utility as a Function of Mean and Variance of the Net Utility Distribution.

A similar approach is used to analyze the bias in slope predictions for the naive method. The true slope, obtained from the derivative of the aggregate probit function (Equation 3.54) is the value of the normal probability density function:

$$SE_i = \frac{\partial S_i}{\partial \bar{V}} = \frac{1}{\sqrt{2\pi(1+\sigma_V^2)}} \exp\left[-\frac{1}{2}\bar{V}^2 / (1+\sigma_V^2)\right] \qquad 3.64)$$

where SE ..... is the share sensitivity (slope) of alternative i with respect to a shift in the mean value of the net utility distribution.

The corresponding slope estimate by the naive procedure is:

$$SE_{iN} = \frac{\partial S_{iN}}{\partial \bar{V}} = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\bar{V}^2\right] \qquad (3.65)$$

where SE_{iN} ..... is the slope estimate of share i by the naive procedure.

FIGURE 3.6

BIAS IN NAIVE PREDICTION FOR NORMALLY DISTRIBUTED NET UTILITY AS A FUNCTION OF MEAN AND VARIANCE OF THE NET UTILITY DISTRIBUTION

PART A - PLOT AGAINST VARIANCE

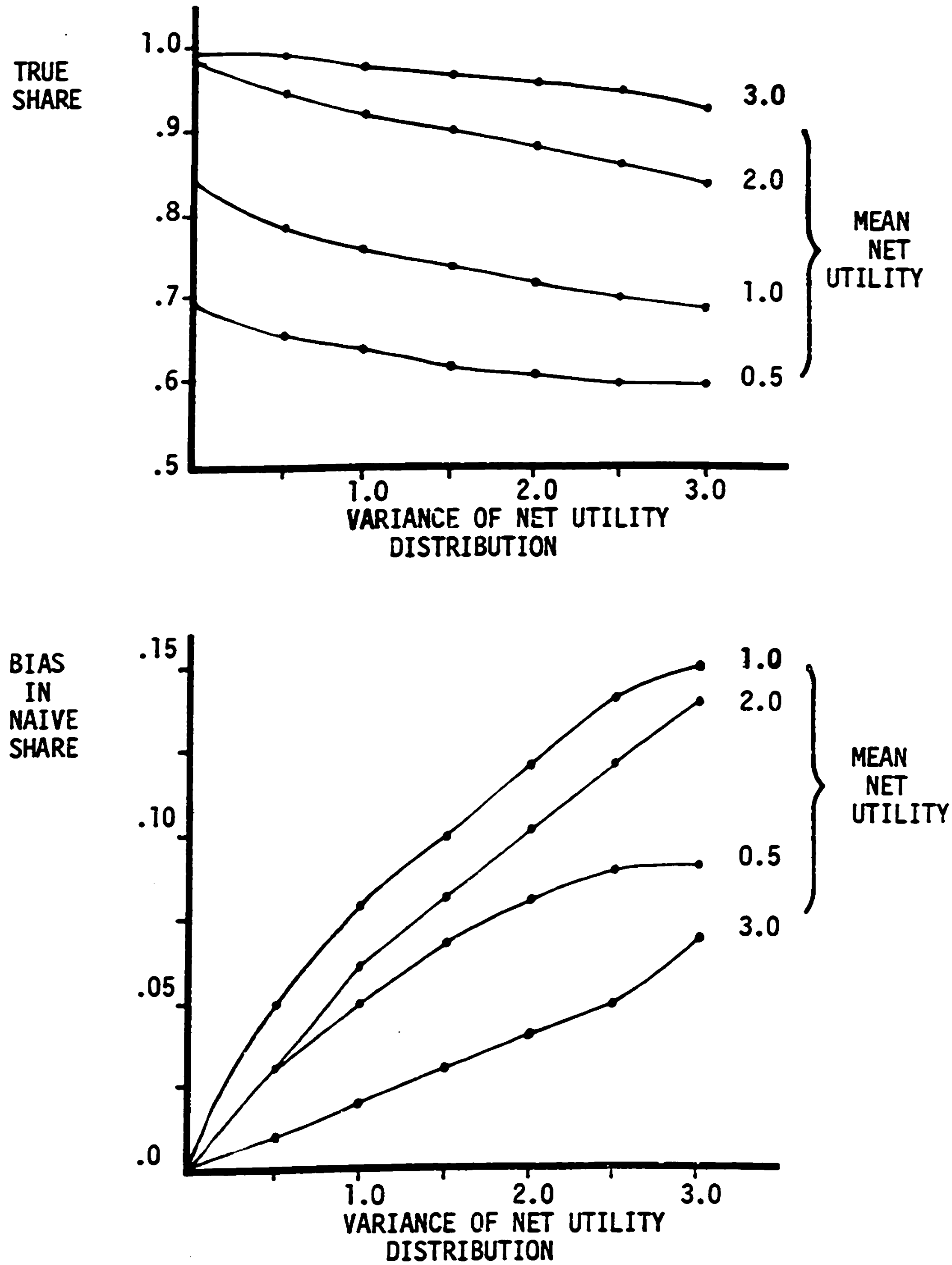
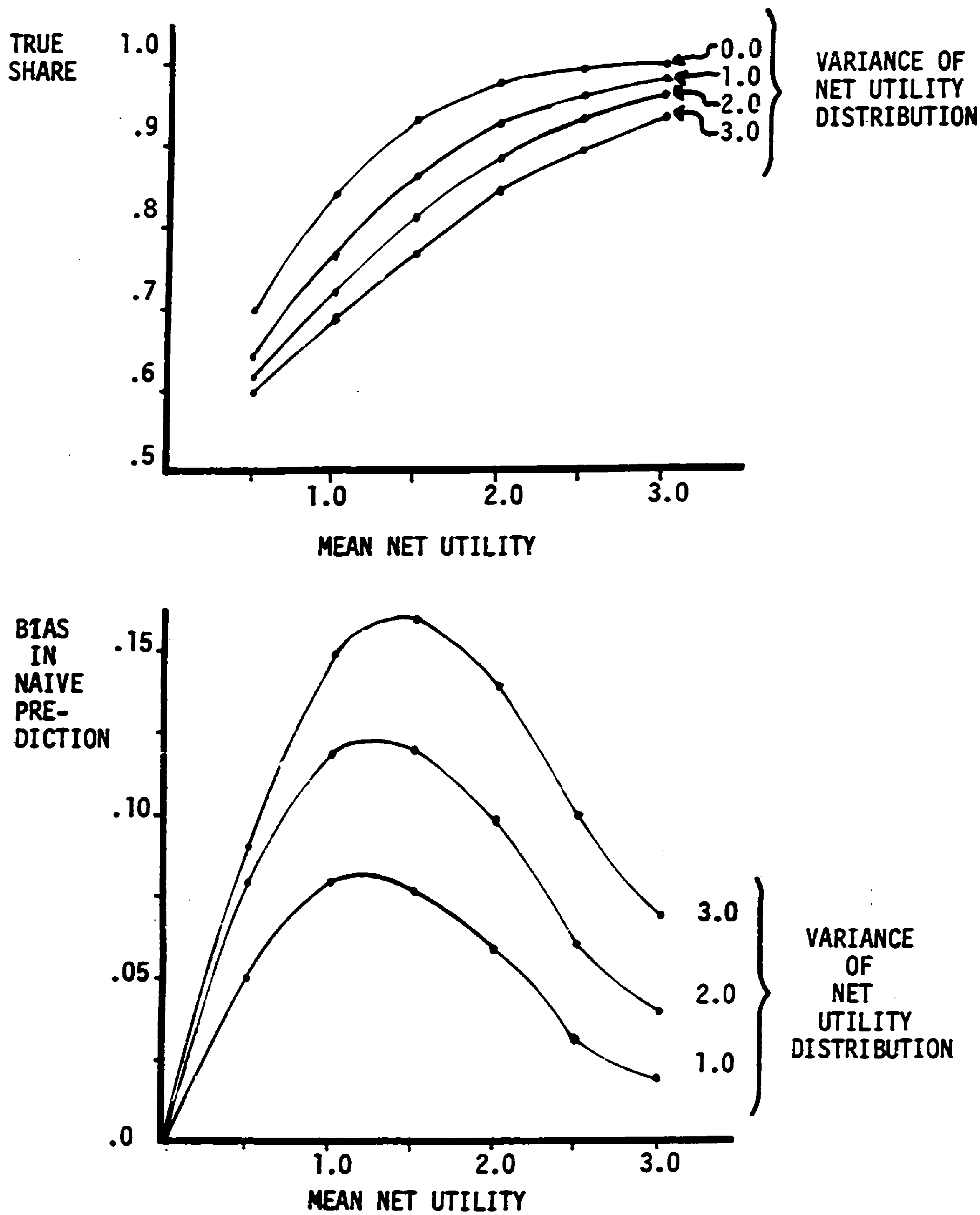


FIGURE 3.6

BIAS IN NAIVE PREDICTION FOR NORMALLY DISTRIBUTED NET UTILITY AS A FUNCTION OF MEAN AND VARIANCE OF THE NET UTILITY DISTRIBUTION

PART B - PLOT AGAINST MEAN NET UTILITY



The bias in estimated slope is given by,

$$BS_N = \frac{1}{\sqrt{2\pi}}\left[\exp\left[-\frac{1}{2}\bar{V}^2\right] - \frac{1}{\sqrt{1+\sigma_V^2}}\exp\left[-\frac{1}{2}\bar{V}^2/(1+\sigma_V^2)\right]\right] \quad (3.66)$$

which is identical to the equation for the derivative of share bias with respect to mean net utility and consequently has zero bias for values of net utility mean which satisfy equation 3.63 or as given in Table 3.2.

The locations of maximum and minimum slope bias are at values of average net utility such that the derivative of slope bias with respect to average net utility,

$$\frac{\partial BS_N}{\partial \bar{V}} = \frac{1}{\sqrt{2\pi}}\left[-\bar{V}\exp\left[-\frac{1}{2}\bar{V}^2\right] + \frac{\bar{V}}{1+\sigma_V^2{}^{3/2}}\exp\left[-\frac{1}{2}\bar{V}^2\ (1+\sigma_V^2)\right]\right] \quad (3.67)$$

is zero which occurs when,

$$|\bar{V}| = \left[3\left(\frac{1+\sigma_V^2}{\sigma_V^2}\right)\ln\,(1+\sigma_V^2)\right]^{1/2} \quad (3.68a)$$

or,

$$\bar{V} = 0 \quad (3.68b)$$

For the condition given by equation 3.68a, the slope bias is negative* and occurs for a mean net utility which is $\sqrt{3}$ times that for zero slope bias. For the condition given by Equation 3.68a, the slope bias is positive and has the value,

$$BS_N(\bar{V}=0) = \frac{1}{\sqrt{2\pi}}\left[1 - \frac{1}{(1+\sigma_V^2)^{1/2}}\right] \tag{3.69}$$

The effect of variance on slope bias is seen by:

$$\frac{\partial BS_N^i}{\partial \sigma_V^2} = \left\{\frac{1/2}{\sqrt{2\pi}}\left[\exp\left[-\frac{1}{2}\bar{V}^2 \quad (1+\sigma_V^2)\right]\right]\right\}\left\{\frac{1}{(1+\sigma_V^2)^{3/2}} - \frac{\bar{V}^2}{(1+\sigma_V^2)^{5/2}}\right\}$$

$$= \left\{\frac{1/2}{\sqrt{2\pi}}\frac{1}{(1+\sigma_V^2)^{5/2}} \exp\left[-\frac{1}{2}\bar{V}^2 \; (1+\sigma_V^2)\right]\right\}\left\{1+\sigma_V^2 - \bar{V}^2\right\} \tag{3.70}$$

which is positive, zero or negative as the value of $(1+\sigma_V^2 - \bar{V}^2)$ is positive, zero or negative. The result is that for magnitude of average net utility less than or equal to one the bias increases monotonically with variance. When the magnitude of average net utility is

---

*In Section 3.4, the slope bias is determined to be negative for mean net utility greater than the value of mean net utility for zero slope bias.

greater than one, the bias declines from zero to some maximum negative value for $\sigma^2 = \bar{V}^2 - 1$ and then increases (becomes less negative and may become positive) with further increases in variance. This effect is illustrated in Table 3.3, Bias in Share Sensitivity Estimates by Naive Method for Normally Distributed Net Utility Values. Part A of the Table contains the bias estimates from Equation 3.66 for a selected range of variance and mean net utility values. The slope bias changes from positive to negative for mean net utility values between 1.0 and 1.5 which is the range of maximum share bias described in Table 3.2. The maximum positive bias occurs for mean net utility equal to zero (Equation 3.69 and 3.68b) and maximum negative bias occurs at mean net utility values near two which is roughly $\sqrt{3}$ times the zero bias mean net utility values as expected (Equation 3.68a).

Part B of Table 3.3 illustrates the slope bias as a percent of true slope,

$$\frac{SE_{iN} - SE_i}{SE_i} = (1 + \sigma_V^2)^{1/2} \exp\left[-\frac{1}{2} \bar{V}^2 \frac{\sigma_V^2}{1+\sigma_V^2} \right] - 1.0 \qquad (3.71)$$

The important feature of these results is that, as a proportion of true slope, the negative slope bias increases throughout the range of mean net utility values examined. With further increases in mean net utility values, it can be shown that the negative slope bias by the naive method approaches one hundred percent.

TABLE 3.3

BIAS IN SHARE SENSITIVITY ESTIMATES BY NAIVE METHOD FOR NORMALLY DISTRIBUTED NET UTILITY VALUES

VARIANCE OF NET UTILITY DISTRIBUTION	ABSOLUTE VALUE OF NET UTILITY MEAN					
	0.0	0.5	1.0	1.5	2.0	3.0
	A. BIAS IN PERCENT SLOPE					
0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.0	11.7	8.7	2.2	-3.1	-5.0	-2.5
2.0	16.9	13.1	4.7	-2.9	-6.4	-4.7
3.0	19.9	15.9	6.6	-2.1	-6.7	-6.0
	B. BIAS IN SLOPE AS PERCENT OF TRUE SLOPE					
0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.0	41.4	32.9	10.1	-19.4	-48.0	-85.1
2.0	73.2	59.4	24.1	-18.2	-54.3	-91.4
3.0	100.0	82.1	37.5	-14.0	-55.4	-93.2

The results obtained in the analysis of bias in naive prediction with normally distributed net utility values in binary choice* are consistent with the results of the general discussion of bias in binary choice with any symmetric distribution (Section 3.2). The additional information obtained from this analysis for normally distributed net utility and a probit choice function can be summarized as follows:

1. Zero share bias occurs at mean net utility equal to zero.
2. Maximum share bias for a given level of variance occurs at mean net utility greater than one (the exact location depends on the variance but is between mean net utility of 1.0 and 1.5 for variance up to 3.0).
3. The magnitude of share bias for any level of mean net utility increases monotonically with variance.
4. Maximum positive slope bias occurs at mean net utility of zero and increases monotonically with variance.
5. Zero slope bias occurs at the same points as maximum share bias (i.e. location depends on variance).
6. Maximum negative slope bias occurs at mean net utility values which are $\sqrt{3}$ times the values for zero slope bias. However, the negative slope bias as a percent of true slope increases to one hundred percent with increasing mean net utility values.

---

*Although the analysis is based on the probit choice function the results are similar for the logit choice function.

7. The magnitude of slope bias is positive and increases monotonically with increasing variance for mean net utility values less than or equal to one. For larger mean net utility values slope bias is negative and increases to a maximum negative value when the variance is equal to the mean net utility squared minus one and then becomes less negative with further increases in variance.

3.6 Bias in Prediction by the Statistical Differentials Method

The previous sections describe bias in share and sensitivity predictions using the naive method. The insights gained can be extended to some related methods of aggregation. In this section, the bias in predictions using the statistical differentials method is analyzed.

The method of statistical differentials, described in Chapter II. linearizes the choice function using a Taylor series and predicts aggregate shares in terms of the choice function and the moments of the distribution of the alternative utility functions. The expression for the binary and multiple choice cases are, respectively:

$$S_{iS} = f^i(\bar{V}) + \frac{1}{2}\sigma_V^2 \left.\frac{\partial^2 f^i(V)}{\partial V^2}\right|_{\bar{V}} + \ldots{}^* \qquad (3.72)$$

and

$$S_{iS} = f^i(\bar{V}) + \frac{1}{2}\sum_j \sum_k \mathrm{Cov}(V_j, V_k) \left.\frac{\partial^2 f^i(V)}{\partial V_j dV_k}\right|_{\bar{V}} + \ldots{}^* \qquad (3.73)$$

where S_{iS} ..... is the aggregate share for alternative i predicted by the statistical differentials method.

---

*The additional terms consist of (1) the inverse of n!, (2) the nth derivative with respect to combinations of utility terms totaling n evaluated at the average of each utility distribution and (3) the nth order moments of the distribution of utility.

$\bar{V}$ ..... is the mean net utility in the binary choice case and the vector of mean utilities for each alternative in the multiple choice case,

σ_V^2 ..... is the variance of the net utility distribution in the binary choice case,

$Cov(V_j, V_k)$ ..... is the covariance in the distribution of utilities for alternatives j and k; when j=k it is the variance of the utility distribution, and

$\Big|_{\bar{V}}$ ..... indicates that the preceding expression is evaluated at the values of $\bar{V}$.

In both cases the additional terms, which include higher distribution moments, are usually ignored. The second term in the series is a function of the second order derivatives with respect to the various choice utilities and the variance-covariance matrix describing the distribution of these utilities in the aggregate population.

The bias in prediction by the statistical differentials procedure for the binary choice case is:

$$B_S^i = S_{iS} - S_i$$

$$= f^i(\bar{V}) + \frac{1}{2}\sigma_V^2 \left.\frac{\partial^2 f^i(V)}{\partial V^2}\right|_{\bar{V}} - \int_{-\infty}^{\infty} f^i(V)\, g(V)dV \qquad (3.74)$$

which can be expressed in terms of the bias for the naive prediction by,

$$B_S^i = B_N^i + \frac{1}{2}\sigma_V^2 \left.\frac{\partial^2 f^i(V)}{\partial V^2}\right|_{\bar{V}} \qquad (3.75)$$

The sign of the added term is determined by the second derivative of the choice function which, for symmetric s-shaped choice functions is zero when the mean net utility value is zero, positive when the mean net utility is negative and negative when then mean net utility is positive. The analysis in earlier sections of this chapter indicated that bias in the naive prediction is zero when the mean net utility is zero, positive when the mean net utility is positive and negative when the mean net utility is negative. Thus, the variance term is in the opposite direction of the bias in naive prediction for symmetric distributions. Whether the variance terms over or under corrects for bias in the naive prediction can be determined by comparing it with the results of naive prediction using the probit choice function with normal distribution of net utilities described in the preceding section. The additional term for the probit choice function is:

$$- \frac{1}{2} \frac{\bar{V}}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} \bar{V}^2 \right] \cdot \sigma_V^2 \tag{3.76}$$

Table 3.4, Comparative Bias of Statistical Differentials and Naive Predictions, sets forth (1) the estimated bias of the naive procedure with the probit choice function from Table 3.1, (2) the value of the adjustment term from the statistical differentials series (Equation 3.75) and (3) the estimated bias by the statistical differentials adjustment.

The additional term in the statistical differentials series over

TABLE 3.4

COMPARATIVE BIAS OF STATISTICAL DIFFERENTIALS AND NAIVE PREDICTIONS

VARIANCE OF NET UTILITY	MEAN NET UTILITY 0.0	0.5	1.0	1.5	2.0	2.5	3.0
	A. BIAS BY NAIVE PREDICTION						
1.0	.0	.05	.08	.07	.06	.03	.02
2.0	.0	.08	.12	.12	.11	.06	.04
3.0	.0	.09	.15	.16	.14	.10	.07
	B. STATISTICAL DIFFERENTIALS ADJUSTMENT						
1.0	.0	-.09	-.12	-.10	-.05	-.02	-.01
2.0	.0	-.18	-.24	-.19	-.11	-.04	-.01
3.0	.0	-.26	-.36	-.29	-.16	-.07	-.02
	C. BIAS BY STATISTICAL DIFFERENTIALS PREDICTION						
1.0	.0	-.04	-.04	-.03	.0	.01	.01
2.0	.0	-.10	-.12	-.07	.0	.02	.03
3.0	.0	-.17	-.21	-.13	-.02	.03	.05

corrects for the bias in naive prediction for absolute values of mean net utility less than two, corrects almost exactly for absolute mean net utility equal to two, and under corrects for absolute mean net greater than two. That is, there are zero bias points at mean net utility equal to two as well as at mean net utility equal to zero. In the range where the statistical differentials term over corrects for the bias in naive prediction, the over correction is generally small enough so that the revised prediction is biased by lesser magnitude than before the addition of the variance term. However, when the variance of the distribution is large and the mean net utility is between plus and minus one the magnitude of bias by the statistical differentials procedure may exceed that for the naive procedure for some mean net utility values.

The additional term of the statistical differentials series is independent of the skewness of the distribution of net utility values. When the distribution of the net utility values is skewed to the right (left) the area of positive and negative bias is shifted to the left (right). However, the range of positive and negative value for the adjustment term is not affected by skewness in the distribution of net utility values. Thus, there is a range of values for mean net utility, between zero and the zero bias point for the actual distribution, where the additional term will result in larger bias than that from naive prediction. Except for an additional small area adjacent to the zero bias point, the statistical differentials procedure has lower magnitude of bias for all other values of mean net utility for low or moderate values of variance. At higher variance levels increased bias may occur for mean net utilities in the range of plus or minus one as with the symmetric distribution.

The effect of using the statistical differentials series in the multiple choice case can be explored by considering the adjustment to bias in naive prediction as in the binary choice case (Equation 3.75):

$$B_S^i = B_N^i + \frac{1}{2} \sum_j \sum_k \operatorname{Cov}(V_j, V_k) \left. \frac{\partial^2 f^i(V)}{\partial V_j \partial V_k} \right|_{\bar{V}} \tag{3.77}$$

Expanding the summations and computing the derivatives for the logit choice function gives:

$$\begin{aligned} SDA = {} & \frac{1}{2} S_N^i \sum_j (\theta_1 - S_N^j)(1 - 2S_N^j) \operatorname{Var}(V_j) \\ & - S_N^i \sum_j S_N^j \sum_{k \neq j} (\theta_2 - 2S_N^j) \operatorname{Cov}(V_j, V_k) \end{aligned} \tag{3.78}$$

where SDA ..... is the statistical differentials adjustment,

θ_1 ..... is 1 if $j = i$ and zero otherwise, and

θ_2 ..... is 1 if $k = i$ and zero otherwise.

The first term sums up the effect of variance in all of the utility distributions and the second sums up the effect of covariance between each pair of utility distributions.

When the distributions of choice utilities are independent the second term is zero so that the total adjustment is:

$$\frac{1}{2} S_N^i \sum_j (\theta_1 - S_N^j)\ (1-2\cdot S_N^i)\mathrm{Var}(V_j) \tag{3.79}$$

The value of this term depends on all of the shares estimated by the naive method. An understanding of the effect of this term in adjusting the share prediction for alternative i when the individual utility variance are equal is obtained by considering two extreme cases. In one case all of the remaining shares are equal:

$$S_N^j = \frac{1 - S_N^i}{R - 1} \qquad \forall\ j \neq i \tag{3.80}$$

In this case the statistical differentials adjustment is:

$$\mathrm{SDA} = S_N^i\ (1-S_N^i)(1-R\cdot S_N^i) \Big/ (R-1)\ \mathrm{Var}(V_i) \tag{3.81}$$

The SDA is zero when S_N^i equals 1/R and, by 3.80, $S_N^j = 1/R$. Thus, in the equal shares cases which is unbiased by the naive prediction the statistical differentials prediction is also unbiased. When S_N^i is less (more) than 1/R, the naive prediction is biased downward (upward).* The SDA in Equation 3.81 is positive (negative) since:

$$R \cdot S_N^i < 1 \tag{3.82}$$

so the statistical differentials adjustment is in the proper direction.

In the other case all of the remaining shares but one are zero. In this case the statistical differentials adjustment is:

$$SDA = S_N^i\,(1 - S_N^i)\,(1 - 2 \cdot S_N^i)\mathrm{Var}(V_i) \tag{3.83}$$

which is always positive for f^i near 1/R if R is greater than two. That is, when the remaining shares are (extremely) non-uniformly distributed the statistical differentials term may introduce a positive adjustment to an unbiased naive prediction.

The existence of skewed net utility distributions further complicates the effect of the statistical differentials adjustment. One effect is to increase the range of situations for which the statistical differentials procedure has greater bias than the naive procedure.

The slope bias for the statistical differentials procedure can be related to the slope bias of the naive procedure for the binary choice

---

*It was shown in Section 3.3 that for identically distributed utility values with different means the share bias is negative for shares less than 1/R and positive for shares greater than 1/R.

case by:

$$BS_S^i = \frac{\partial}{\partial \bar{V}}\left[f^i(\bar{V}) + \frac{1}{2}\frac{\partial^2 f^i(\bar{V})}{\partial \bar{V}^2}\sigma_V^2\right] - \int_{-\infty}^{\infty}\frac{\partial f^i(V)}{\partial \bar{V}} g(V_D)dV_D$$

$$= \frac{\partial f^i(\bar{V})}{\partial \bar{V}} + \frac{1}{2}\frac{\partial^3 f^i(\bar{V})}{\partial \bar{V}^3}\sigma_V^2 - \int_{-\infty}^{\infty}\frac{\partial f^i(V)}{\partial \bar{V}} g(V_D)dV_D$$

$$= BS_N^i + \frac{1}{2}\frac{\partial^3 f^i(\bar{V})}{\partial \bar{V}^3}\sigma_V^2 \tag{3.84}$$

When the choice model has the logit form the statistical differentials slope adjustment (SDSA) is:

$$SDSA = \frac{1}{2} S_N^i \left[1-S_N^i\right]\left[1-6\ S_N^i + 6(S_N^i)^2\right]\sigma_V^2 \tag{3.85}$$

which is equal to zero when the absolute value of net utility mean is approximately equal to 1.3, is negative for lower absolute values and positive for higher absolute values. It was previously shown (Section 3.5) that zero slope bias occurs near the point of maximum curvature of the choice function which is the point at which the statistical differentials slope adjusting term (Equation 3.85) is equal to zero. That is, the slope adjusting term is zero at, or near the point of zero slope bias, negative in the area of positive slope bias and positive in the area of negative slope bias. When the choice model is probit the statistical differentials slope adjustment term is:

$$\text{SDSA} = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\overline{V}^2\right] (\overline{V}^2-1) \tag{3.86}$$

which again is equal to zero at the point of maximum curvature (where $\overline{V} = 0$). In Table 3.3, the zero slope bias is close to this point although for some values of net utility mean (between 1.0 and 1.25) the positive bias will be increased by the additional term.

The results of the analysis of bias of statistical differentials predictions relative to the bias of naive predictions are:

1. When the net utility distribution is symmetric share predictions by the statistical differentials procedure have lower absolute bias than predictions by the naive procedure except for high values of variance (see Table 3.4).
2. When the net utility distribution is skewed predictions by the statistical differentials procedure have lower absolute bias than for the naive procedure except for a range of net utility values between zero and near the zero bias point for the naive prediction and for higher values of variance.
3. Slope predictions are more accurate except for a narrow range in the vicinity of the zero slope prediction by the naive method.
4. In the multiple choice case there are a variety of conditions which result in higher prediction bias by the statistical differentials procedure than by the naive procedure. These include

the conditions of high variance and skewed distributions mentioned for the binary choice case and an additional case which occurs when the remaining shares are extremely different.

3.7 Bias in Prediction by the Classification Method

This section analyzes the bias in prediction using the method of classification. This method consists of (1) assigning members of the aggregate group to two or more classes, (2) using the naive method to predict aggregate choice shares for each class and (3) computing overall aggregate share as the weighted average of the class shares. The share prediction by the classification method is:

$$S_{iC} = \sum_{G} f^{i}(\overline{V}_{G}) \frac{T_G}{T} \tag{3.87}$$

where S_{iC} ..... is the share predicted by the classification procedure,

$\sum_{G}$ ..... indicates summation over all of the subgroups,

$\overline{V}_G$ ..... is the average net utility for individuals in subgroup G,

T_G ..... is the total number of individuals in subgroup G, and

T ..... is the size of the total population

The bias in share prediction is given by,

$$B_C^i = \sum_{G} f^{i}(\overline{V}_{G}) \frac{T_G}{T} - \int_{-\infty}^{\infty} f^{i}(V)\, g(V)\, dV \tag{3.88}$$

where B_C^i ..... is the bias by the classification method, which may be rewritten as:

$$B_C^i = \sum_G \int_{-\infty}^{\infty} \left[f^i(\overline{V}_G) - f^i(V)\right] g(V_G)\, dV \qquad (3.89)$$

where $g(V_G)$ ..... is modified distribution of net utility values for individuals in group G such that $\int_{-\infty}^{\infty} g(V_G)\, dV_G = \frac{T_G}{T}$

In this expression (Equation 3.89) each term in the summation is the contribution to average bias of the naive prediction for the subgroup G, and the weighted summation is taken over all of the subgroups. The bias will be zero if either the bias for each group is zero or the biases are compensating in such a way that the weighted average is equal to zero.

This condition is independent of the number of subgroups used and the method of assignment of individuals to subgroups. To minimize bias, the classification strategy should assign individuals to subgroups so that the individual subgroups will satisfy the zero bias condition, the bias in different subgroups will be offsetting or the within group variance of net utility distribution in each subgroup will be significantly smaller than the original overall variance of net utility in the aggregate population. In practice the amount of information required to pursue either of the first two objectives is so great that it would probably be possible to use the same information to describe the distribution of utilities in a way which would allow use of the enumeration or integration procedures.* The only practical strategy is to assign members to groups to reduce within group variance. The need to predict the mean utility values and size for each subgroup suggests that the number of subgroups should be small.

In practice, classification will be based on grouping of individuals according to one or more of their socio-economic characteristics or the attributes of the choices available to them. The selection of the variable or variables to be used for this purpose should be based on their relative contribution to the overall variance of the utility function. The utility distributions of each class may be overlapping, that is, the values of individual utilities may not be fully rank ordered between subgroups. For the purpose of this discussion, it is assumed that the members of the aggregate group are classified according to their utility values. The classification is based on equal fractiles of the distribution with individual net utility values rank ordered between groups.

The bias for the case of a symmetric distribution classed into two groups at the median net utility is examined first. In this case the bias is:

---

*The effect of offsetting bias will operate in many situations. The point here is that structuring the classes to achieve this objective will require considerable information.

$$B_C = \int_{-\infty}^{G(V)=\frac{1}{2}} \left[f(\overline{V}_{G1}) - f(V)\right] g(V)dV + \int_{G(V)=\frac{1}{2}}^{\infty} \left[f(\overline{V}_{G2}) - f(V)\right] g(V)dV \tag{3.90}$$

where $G(V) = \frac{1}{2}$ ..... is the point where the cumulative distribution function is equal to 1/w, and

$\overline{V}_{G1}$ ..... is the average value of net utility for group 1, etc.

The half distributions created are mirror images of each other. In general, the share prediction for each subgroup is biased. The biases of these distributions are of equal magnitude and opposite sign when the means of the subgroups are of equal magnitude and opposite sign and they have mirrored shapes. This condition occurs when average net utility for the combined group is zero and the net utility values are symmetrically distributed.

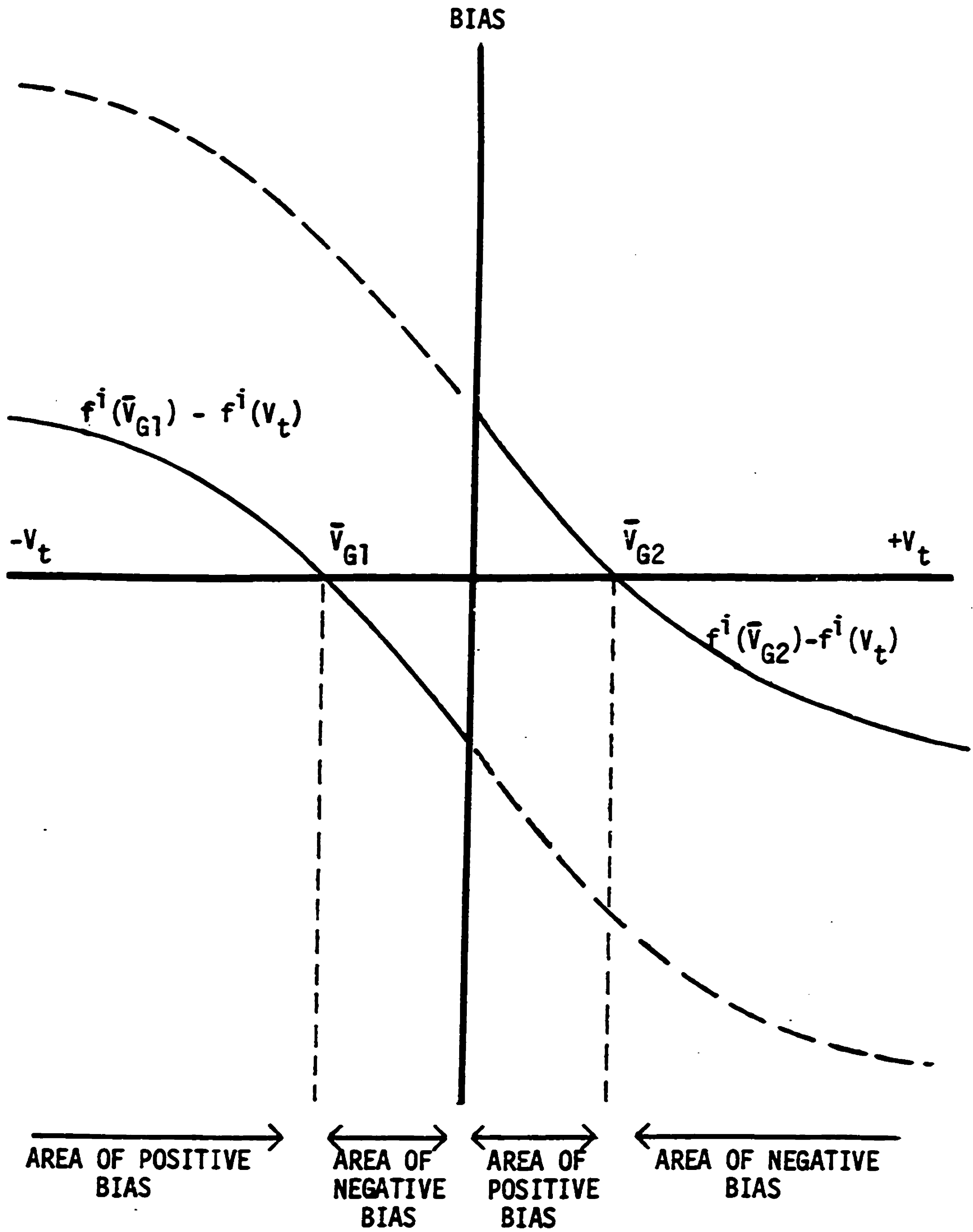
A graphical representation of Equation 3.90 illustrate the result obtained. Figure 3.7, Bias in Disaggregate Share Estimates with Classification Method When Utility Mean is Zero, shows the bias for individuals in the aggregate group. The bias is given by:

$$f^i(\overline{V}_{G1}) - f^i(V) \qquad \text{for} \qquad V < \overline{V} \tag{3.91a}$$

$$f^i(V_{G2}) - f^i(V) \qquad \text{for} \qquad V > \overline{V} \tag{3.91b}$$

FIGURE 3.7

BIAS IN DISAGGREGATE SHARE ESTIMATES WITH CLASSIFICATION METHOD WHEN UTILITY MEAN IS ZERO

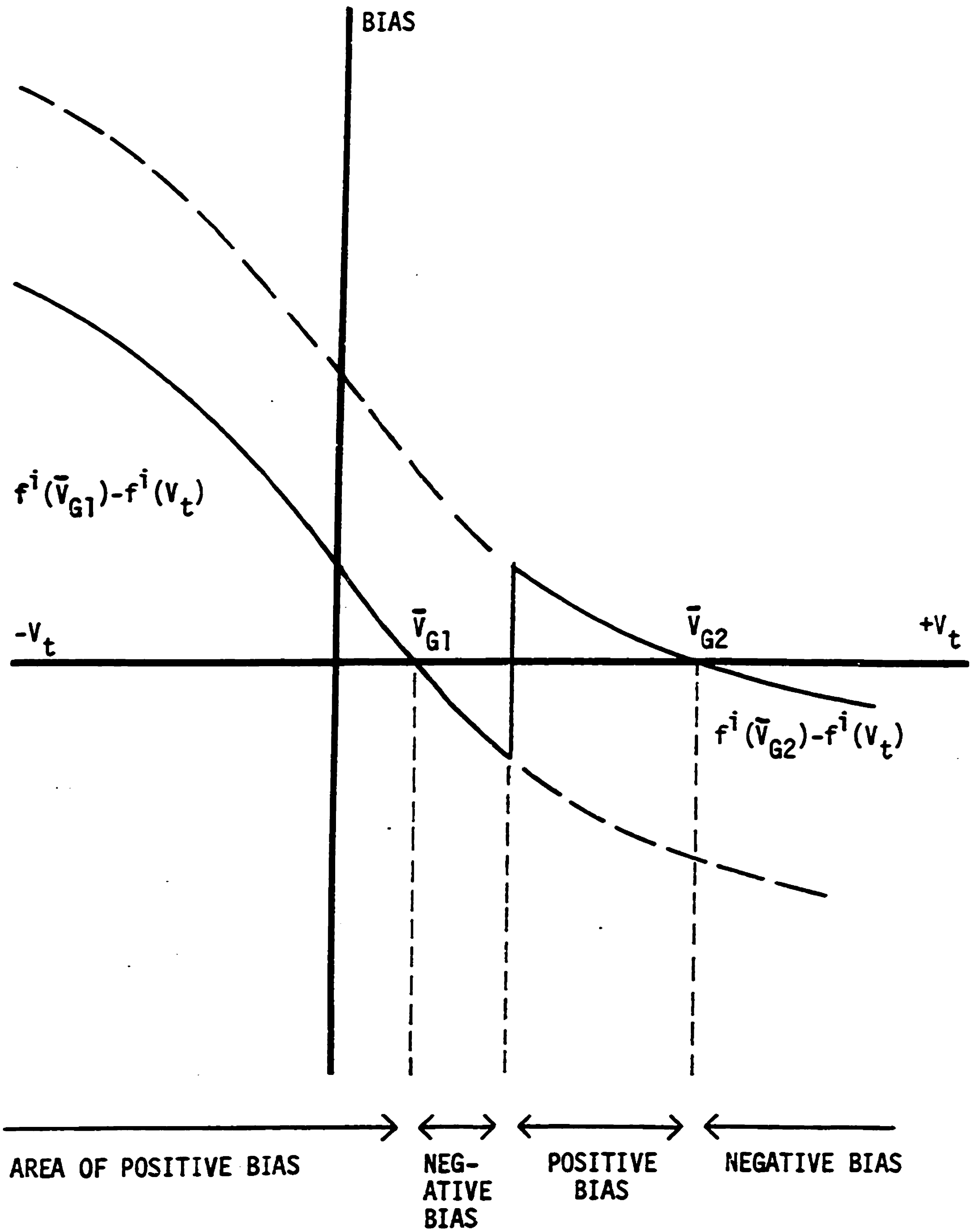


The average bias for group G1 is negative if the initial distribution was unimodal. This follows since the left hand portion of the distribution is skewed to the left and has negative mean. It was shown earlier that left skewed distributions with negative mean net utility are negatively biased. The opposite result holds for group G2. As described, above, these two effects are balancing in this case.

A modification of this discussion is used to examine the bias for symmetric distributions with mean values greater than zero represented by Figure 3.8, Bias in Disaggregate Share Estimates with Classification Method When Utility Mean is Greater than Zero. As average net utility increases, the mean net utilities of the two subgroups increase. From the analysis of bias in naive prediction, the change in bias for each of the subgroups and consequently for the total population can be inferred. As the mean net utility of group one increases, the negative bias of group one declines, reaches zero at some point as described in Section 3.2 and then becomes positive. As the mean net utility of group two increases, the positive bias of group increases until it passes the range of maximum bias and begines to decline. This combination of effects creates positive bias whenever the mean net utility is less than zero. The fact that maximum bias is reached for the subgroups at different mean net utilities and the smaller disaggregate bias by the classification method (Figures 3.7 and 3.8) as opposed to the naive procedure (Figures 3.1 and 3.2) suggests that the overall bias will be lower than for the naive procedure.

FIGURE 3.8

BIAS IN DISAGGREGATE SHARE ESTIMATES
WITH CLASSIFICATION METHOD WHEN UTILITY
MEAN IS GREATER THAN ZERO



The decline in the magnitude of bias with increases in the number of classes can be inferred by recognizing that as the number of classes approaches the number of members in the group the overall bias converges to zero. For symmetric distributions,* the bias is expected to decrease monotonically as the number of classes increases. The diminishing benefit in reduction of within group variance with increasing numbers of classes suggests that corresponding diminishing reduction in prediction bias occurs as the number of classes increases.

Since bias in slope prediction is the derivative of bias in share prediction, the expected overall reductions in share bias are expected to lead to reduced bias in slope prediction with increasing number of classes.

Due to the difficulty of expressing the classification bias in an easily manipulable form, no attempt is made here to analyze the bias by the classification procedure for either skewed distributions or multiple choices. The experience with analysis of bias structure with the statistical differentials procedure, however, suggests that some caution be used in making direct extensions of the results obtained for symmetric distributions in the binary choice case. That is, although

---

*In the next chapter, it is shown that when the distribution of net utilities is skewed, there are ranges of mean net utility for which an increase in the number of classes increases bias.

bias is expected to be less with the classification method than with the naive method, there may be situations, in particular those situations in which the naive procedure performs especially well, in which the bias magnitude is larger for more classes than less.

3.8 Summary

Aggregation bias in the prediction of shares and share sensitivity by the naive, statistical differentials and classification procedures is analyzed in this chapter. The analysis identifies the general structure of bias for each procedure with particular emphasis on the importance of the mean, shape and variance of the distribution of net utility values between pairs of alternatives. Changes in these characteristics affect both the absolute and relative bias of predictions by the three procedures studied.

Particular attention is directed at the conditions which result in zero or maximum bias in share and/or share sensitivity predictions. All of the procedures predict both share and share sensitivity with zero aggregation bias when the distributions of net utility are homogeneous (all members of the aggregate group for which the prediction is made are identical with respect to the choice being made). Increasing hetrogeneity (variance of the distribution of net utility) of the members of the prediction group results in a monotonic increase in the magnitude of positive or negative bias. The magnitude of the bias for different levels of variance depends on the mean and shape of the distribution.

When the distribution of net utility values is symmetric, zero bias in share prediction by all three procedures occurs when the shares for all alternatives are equal. The location and magnitude of the maximum positive and negative share bias occurs at points which are different for each procedure. The share bias of predictions by the naive procedure is greater than that for the classification procedure for all values of mean and variance of the net utility distribution. The share bias by the naive procedure is also greater than that for the statistical differentials procedure except when the variance is very large.

When the distribution of net utility is skewed the points of zero and maximum bias in share prediction occur at different values of mean net utility. When the distributions are skewed, a range of mean net utility values exists within which the naive procedure has least bias of the procedures examined.

The conditions for zero bias in share sensitivity prediction are identical to those for maximum bias in share prediction. In addition maximum bias in share sensitivity occurs under the same conditions (for symmetric distributions) or similar conditions (for skewed distributions) as zero bias in share prediction.

The results of this analysis serve as a base point for further studies of aggregation bias and provide a structure for interpretation of these studies. Bias in prediction by a larger number of procedures based on a simulation analysis is described in Chapter IV.

CHAPTER IV

SIMULATION ANALYSIS OF ALTERNATIVE AGGREGATION PROCEDURES

4.1 Introduction

A range of aggregation procedures were identified in Chapter II. The purpose of these aggregation procedures is to transform the disaggregate choice function and the distribution of choice variables into a set of aggregate predictions. A central element in the evaluation of aggregation procedures is the degree to which they introduce error into the prediction process over and above the error in the choice model - errors in the parameter estimates - and error in the representation of the distribution of choice variables. In the preceding chapter, three aggregation procedures were analyzed to obtain an understanding of the structure of the bias error which they introduced into predictions. Three important elements effecting the magnitude and direction of bias were identified. These are (1) the values of the mean net utilities between pairs of alternatives, (2) the variance of the distribution of net utilities between pairs of alternatives and (3) the skewness of the distribution of net utilities between pairs of alternatives.

The purpose of this chapter is to add breadth, in terms of the range of procedures studied, and richness of detail, in terms of actual estimates of bias for a limited range of situations, to the analysis of bias in aggregation. The results obtained verify and extend the analysis of the preceding chapter.

The aggregation procedures considered are:

1. Integration with known distribution. Since the distribution of utility values is precisely known, this procedure is equivalent to complete enumeration. It has no aggregation error and will be used as the base case against which other procedures will be compared.
2. Integration with assumed normal distribution. This procedure has been proposed by different researchers (McFadden and Reid, 1975; Westin, 1974) because of its relative ease of use (particularly the way in which the variance of individual variables may be collapsed to one variable for each alternative less one) and the moderate amount of predictive information required.
3. Integration with assumed uniform distribution. This procedure is suggested as an unrealistic assumption about the characteristics of the distribution to obtain some idea of the robustness of the distributional assumption.
4. Classification with two classes at median. This procedure represents the simplest type of classification. It gives a measure of the advantage of a simple classification as opposed to using the naive method.
5. Classification with three classes at tertiles. This procedure is used to examine the incremental reduction in bias which

results from increasing the number of classes.

6. Statistical differentials with mean and variance. This is the only statistical differential procedure proposed for regular use. It illustrates the effect on bias of the inclusion of variance information.

7. Naive procedure. This procedure represents the effect of using mean values only for prediction.

The simulation analysis is structured so that there is no source of error except for the bias due to aggregation independent of other sources of error.

The range of predictive situations considered parallel those of the preceding chapter except that the simulation analysis is restricted to the binary choice case. The key elements to be explored are those identified as important in the preceding chapter: (1) the mean value of the net utilities between alternatives, (2) the variance of the distribution of net utilities in the population and (3) the shape of the distribution of net utilities in the population.

The analysis investigates the bias in both share prediction and slope (share sensitivity) prediction. Particular attention is given to locating conditions for zero bias and maximum bias in both share and slope prediction. The effect of increasing variance is also studied.

4.2 Aggregation Procedures and Simulation Approach

The simulation procedure is based on a single variable binary choice logit* model,

$$P(i:A) = \frac{1}{1+e^{-V}} \tag{4.1a}$$

$$V = V_i - V_j = A_0 + A_1 X \tag{4.1b}$$

and a generated distribution for the values of the variable X. The distribution of X comes from a standardized binomial distribution. That is, a variable Y is generated with binomial distribution,**

$$\text{Prob}(Y=y) = \frac{N!}{y!(N-y)!} p^y (1-p)^{N-y} \tag{4.2}$$

The variable has expected value and variance:

---

*The general results observed will apply to other symmetric monotonic choice functions as well.

**The binomial distribution occurs in nature through a series of N independent trials which generate values of one with probability p and zero otherwise. The value of y is the sum of the N zero and one generated values and has the probability distribution given.

$$E(Y) = Np \tag{4.3a}$$

$$Var(Y) = Np(1-p) \tag{4.3b}$$

The real valued variable X with mean equal to zero and variance equal to one is obtained by transformation of the integer valued variable Y according to:

$$x = [y-E(y)] \;/\; [Var(y)]^{1/2} \tag{4.4}$$

The important shape characteristics of the distribution of the variable X (skewness and kurtosis) are:

$$\mu_3 = (1-2p)\ [Np(1-p)]^{-1/2} \tag{4.5a}$$

$$\mu_4 = 3 + [1-6p(1-p)]\ [Np(1-p)]^{-1} \tag{4.5b}$$

The distribution of the net utility values is obtained from the linear transformation (Equation 4.1b). The mean and variance of the utility distribution are:

$$\overline{V} = a_0 \tag{4.6a}$$

$$\sigma_v^2 = (a_1)^2 \tag{4.6b}$$

The variance of the utility distribution may be adjusted by making changes in a_1 and the mean value or location on the choice function may be varied by changing the parameter a_0. The shape of the distribution of net utility values is the same as that for the underlying standardized binomial distribution (Equations 4.5a and b).
The shape characteristics can be modified by selection of the parameter p.* In this manner, it is possible to explore a range of conditions of the mean net utility distribution, its variance and its shape.

The range of conditions to be simulated are:

1. Mean value of net utility between plus and minus four. This covers the range of aggregate shares between 3 and 97 percent.
2. The variance of the distribution of net utility in the population is for values of one, two and three.
3. The probability parameter of the standardized binomial distribution is set at values of 0.05, resulting in a distribution with skewness equal to 1.3 and kurtosis equal to 4.5, and 0.50, resulting in a symmetric distribution (skewness equal to 0.0) with kurtosis equal to 2.8.**

---

*The shape may also be modified by changes in the parameter N, however N determines the number of points in the distribution and is held constant in this study.

**The two distributions generated are described only as symmetric or skewed in the following discussion.

The results obtained are general enough to allow interpolation and/or extrapolation to a more extensive range of situations.

The information required for using each of these procedures is assumed to be perfectly known. The interacting effect of increased error in variables associated with more extensive data requirements must be explored to obtain an overall understanding of the effect on total error of alternative aggregation procedures. This analysis is designed to identify the bias error of the aggregation procedures only.

The simulation analysis includes the prediction of the estimated share and estimated share sensitivity using each of the aggregation methods described earlier. The predictions by each method are plotted as a function of the mean net utility values in the range between plus and minus four as illustrated by Figure 4.1, Estimates of Aggregate Share by All Methods, and Figure 4.2, Estimates of Aggregate Slope by All Methods.

The scale of aggregate share prediction ranges from zero to 100% which are the natural bounds of the aggregate shares. The scale of aggregate slope prediction ranges from zero to 25% which are the natural bounds of the aggregate slope (slope is always positive and achieves a maximum of 25% which is the maximum of the first derivative of the logit function for a point distribution of utilities with mean value of zero).

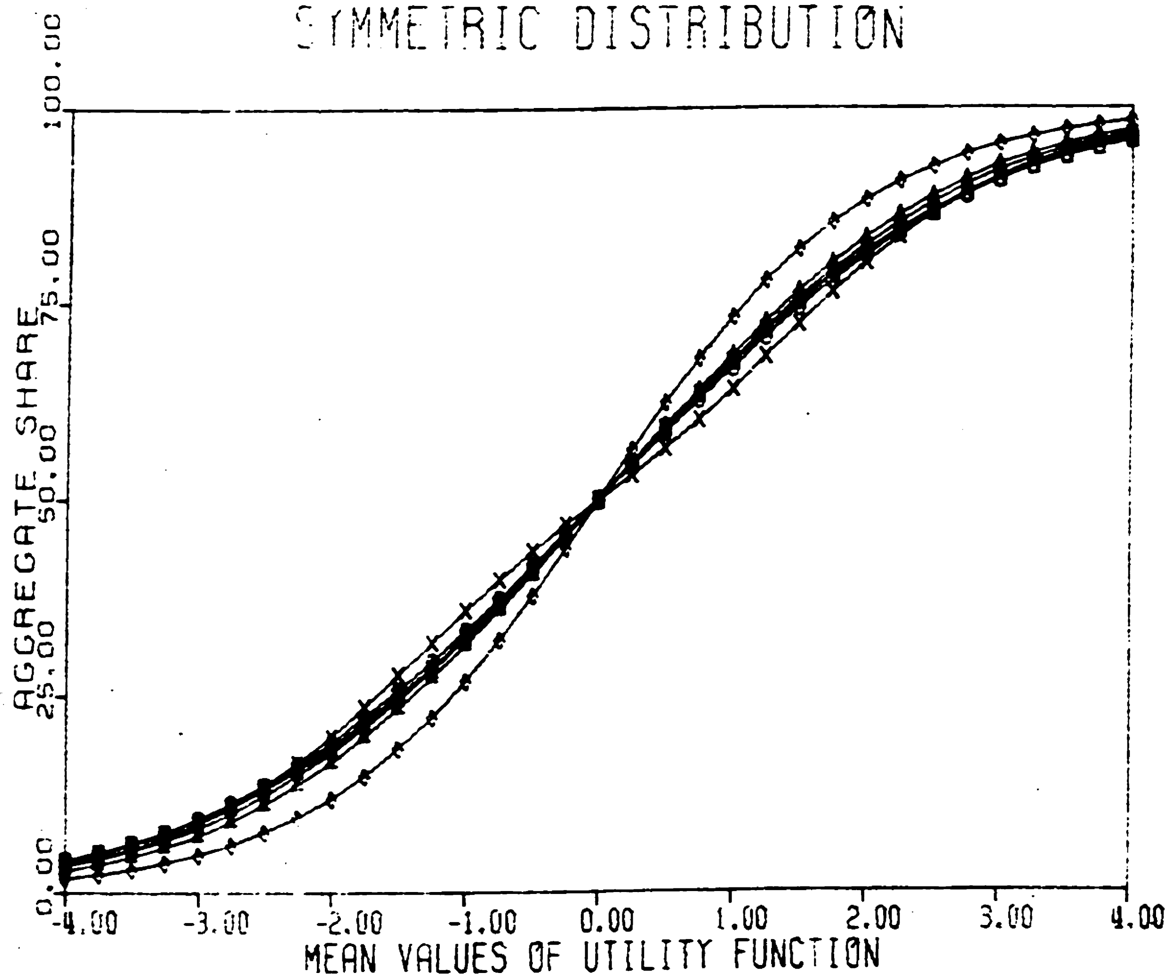
The error or bias in share or slope prediction is obtained by subtracting the correct prediction (by use of the numerical integration with known distribution procedure) from the prediction obtained by each

FIGURE 4.1

ESTIMATES OF AGGREGATE SHARE

BY ALL METHODS

SYMMETRIC DISTRIBUTION



AGGREGATION METHODS

——— NUM. INT. / KNOWN DISTRIBUTION

⊡⊡⊡⊡ NUM. INT. / NORMAL DISTRIBUTION

⊖⊖⊖⊖ NUM. INT. / UNIFORM DISTRIBUTION

▲▲▲▲ CLASSIFIC. 2 SETS / MEDIAN

++++ CLASSIFIC. 3 SETS AT THIRDS

XXXX STAT. DIFF. W/ MEAN & VARIANCE

◊◊◊◊ NAIVE METHOD

TRUE DISTRIBUTION FORM

BINOMIAL (N=10,P=0.50)

STANDARDIZED - VARIANCE IS 2.0

SKEWNESS INDEX IS 0.0

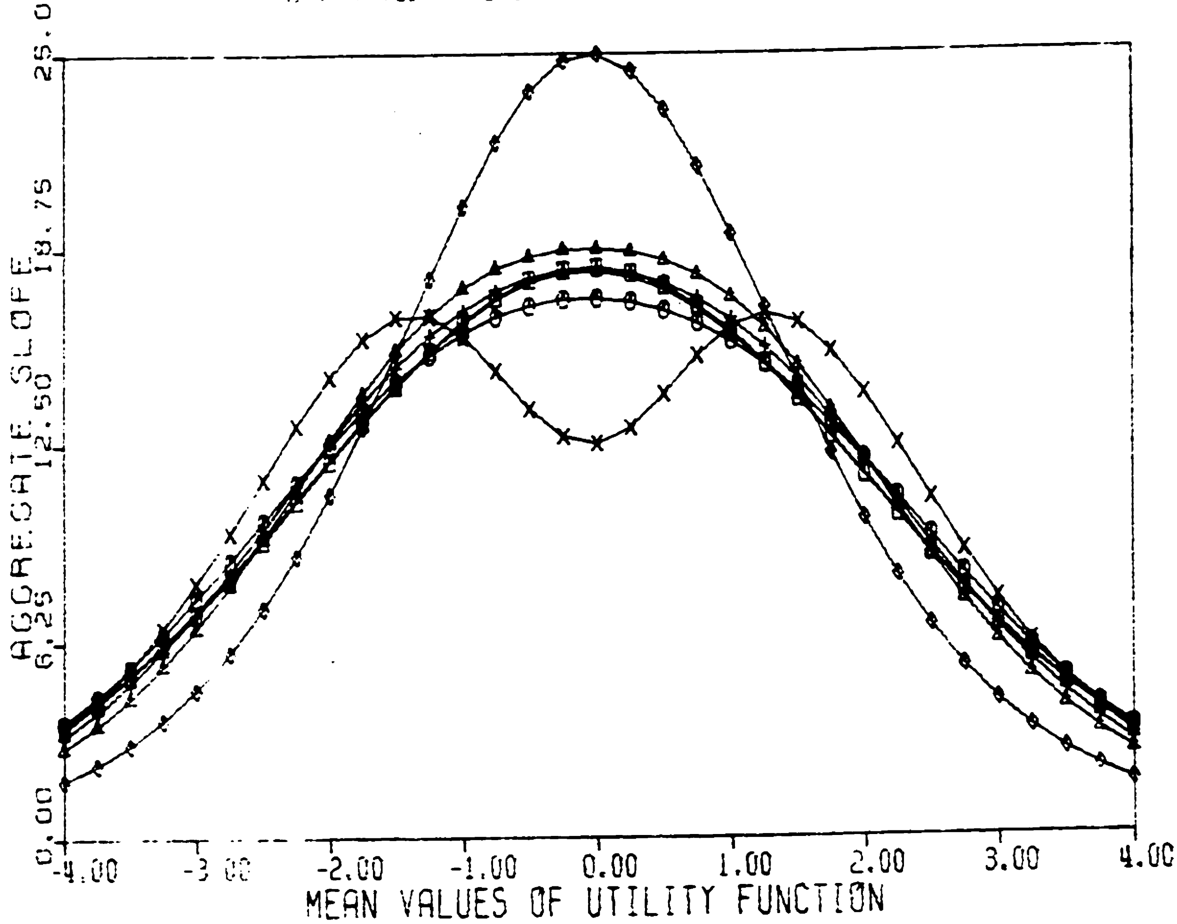
KURTOSIS INDEX IS 2.8

FIGURE 4.2

ESTIMATES OF AGGREGATE SLOPE

BY ALL METHODS

SYMMETRIC DISTRIBUTION



AGGREGATION METHODS

- ——— NUM. INT. / KNOWN DISTRIBUTION
- ⊡⊡⊡⊡ NUM. INT. / NORMAL DISTRIBUTION
- ⊖⊖⊖⊖ NUM. INT. / UNIFORM DISTRIBUTION
- ΔΔΔΔ CLASSIFIC. 2 SETS / MEDIAN
- ++++ CLASSIFIC. 3 SETS AT THIRDS
- XXXX STAT. DIFF. W/ MEAN & VARIANCE
- ◊◊◊◊ NAIVE METHOD

TRUE DISTRIBUTION FORM

BINOMIAL (N=10,P=0.50)

STANDARDIZED - VARIANCE IS 2.0

SKEWNESS INDEX IS 0.0

KURTOSIS INDEX IS 2.8

of the other methods. This is illustrated in Figure 4.3, Bias in Estimates of Aggregate Share by All Methods, and Figure 4.4, Bias in Estimates of Aggregate Slope by All Methods. The scale of the bias estimate in both cases ranges between plus and minus eight percent which includes the largest error in either share or slope prediction in the range of situations considered. Measurement of bias as a proportion of the magnitude of share or slope prediction is determined by comparison of the bias figures to the predicted shares and slopes in Figures 4.1 and 4.2. This measure of bias is included in the analysis in Section 4.7.

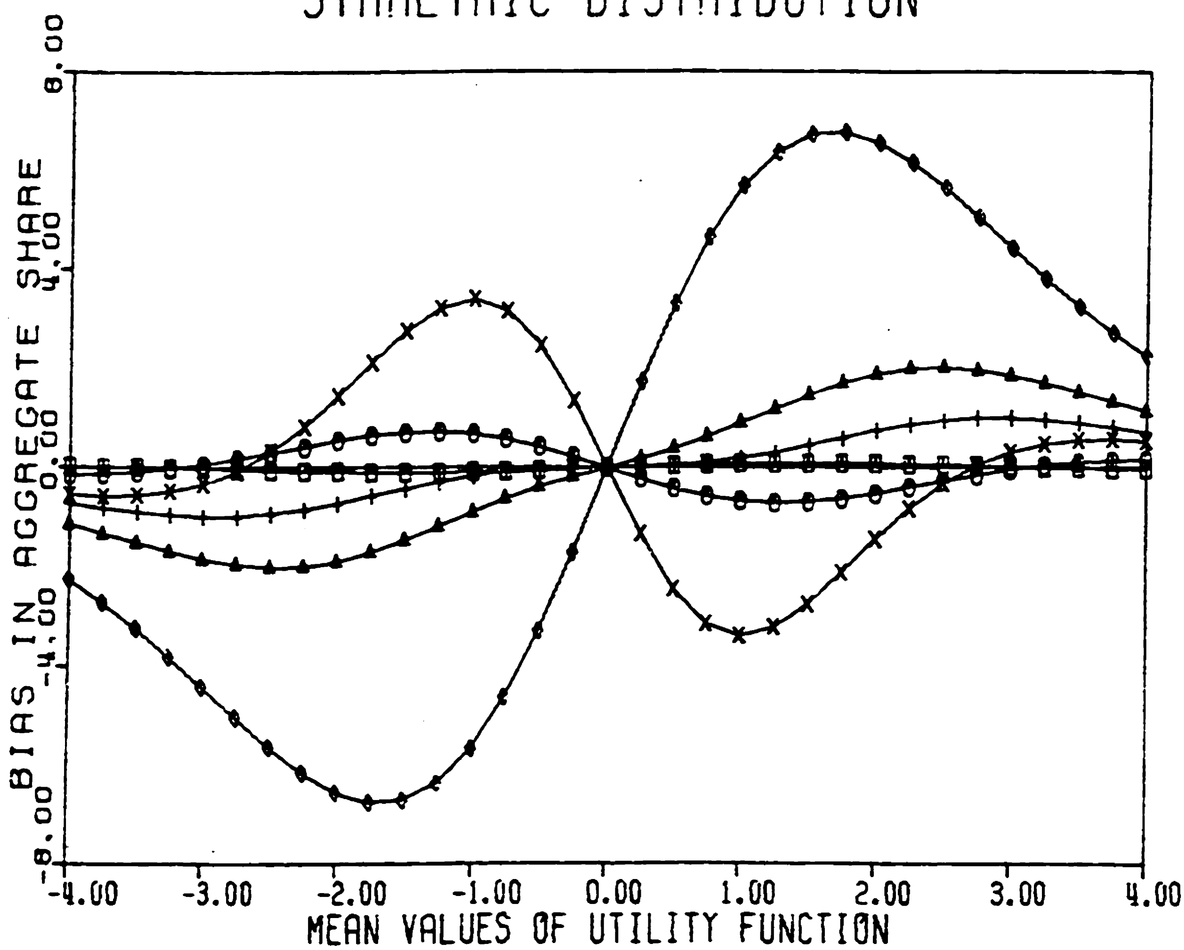
The methods represented in the graphs are identified in the lower left hand portion of the figure where each method and its corresponding symbol are identified. The shape and variance of the distribution of net utilities is described in the lower right hand portion of each figure. Figures 4.1 to 4.4 are based on a symmetric distribution with variance equal to two.

The next three sections describe the bias structure for different aggregation methods. The base case for comparison of methods is integration with the known distribution which, in this simulation study, is identical to the theoretically correct enumeration procedure. The analysis will be developed in terms of the bias of each method corresponding to Figures 4.3 and 4.4.

The simulation results are examined for three different groups of procedures. These are:

FIGURE 4.3

BIAS IN ESTIMATES OF AGGREGATE SHARE
BY ALL METHODS
SYMMETRIC DISTRIBUTION



AGGREGATION METHODS

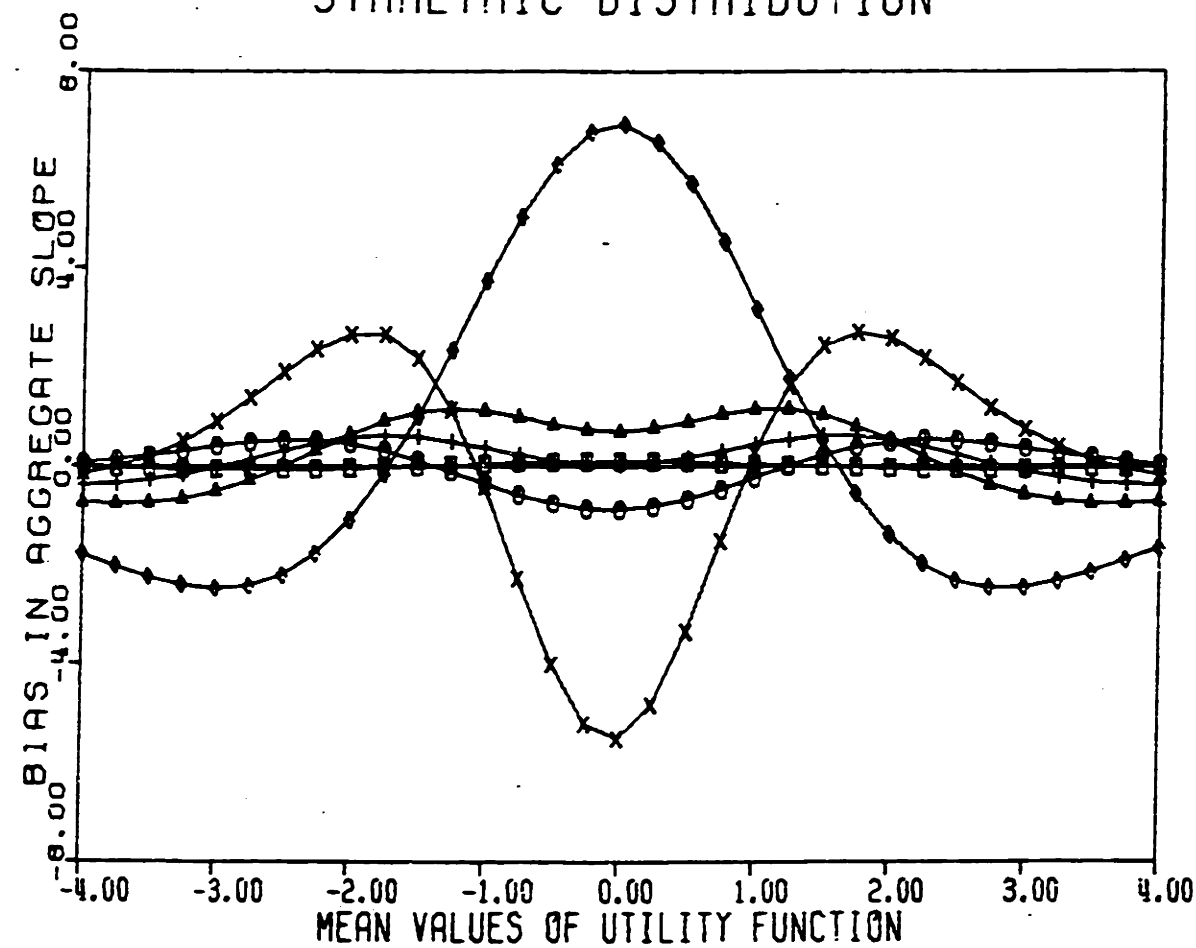
——— NUM. INT. / KNOWN DISTRIBUTION
⊟⊟⊟⊟ NUM. INT. / NORMAL DISTRIBUTION
⊖⊖⊖⊖ NUM. INT. / UNIFORM DISTRIBUTION
▲▲▲▲ CLASSIFIC. 2 SETS / MEDIAN
++++ CLASSIFIC. 3 SETS AT THIRDS
XXXX STAT. DIFF. W/ MEAN & VARIANCE
◆◆◆◆ NAIVE METHOD

TRUE DISTRIBUTION FORM

BINOMIAL (N=10,P=0.50)
STANDARDIZED - VARIANCE IS 2.0
SKEWNESS INDEX IS 0.0
KURTOSIS INDEX IS 2.8

FIGURE 4.4

BIAS IN ESTIMATES OF AGGREGATE SLOPE BY ALL METHODS SYMMETRIC DISTRIBUTION



AGGREGATION METHODS

- ———NUM. INT. / KNOWN DISTRIBUTION
- ▢▢▢▢ NUM. INT. / NORMAL DISTRIBUTION
- ○○○○ NUM. INT. / UNIFORM DISTRIBUTION
- △△△△ CLASSIFIC. 2 SETS / MEDIAN
- ++++ CLASSIFIC. 3 SETS AT THIRDS
- XXXX STAT. DIFF. W/ MEAN & VARIANCE
- ◇◇◇◇ NAIVE METHOD

TRUE DISTRIBUTION FORM

BINOMIAL (N=10,P=0.50)

STANDARDIZED - VARIANCE IS 2.0

SKEWNESS INDEX IS 0.0

KURTOSIS INDEX IS 2.8

1. <u>Methods of moments</u> including the naive method and the method of statistical differentials with mean and variance.
2. <u>Methods of classification</u> including methods using one class (the naive method), two classes and three classes.
3. <u>Methods of integration</u> including integration with assumed normal distribution, with assumed uniform distribution and assumed homogeneous distribution (naive method).

The inclusion of the naive method in each group provides a natural base for comparison of relative bias among methods.

The analysis for each group is made with a symmetric distribution and a skewed distribution both with variance equal to two. The effect of different variance levels in the prediction bias of all of the methods is described in Sections 4.6 and 4.7. The results obtained are specific to the simulated distribution; however, the general characteristics of the bias and orders of magnitude are indicative of aggregation bias for unimodal distributions as a class.

4.3 Aggregation Bias by Methods of Moments

This group includes the naive method and the method of statistical differentials with mean and variance. Figure 4.5 illustrates the bias in share estimates for these methods for a symmetric distribution with variance of two. The naive method has zero bias when the mean net utility is zero (one-half aggregate share). The bias is positive for higher mean net utility and negative for lower mean net utility. In either case the magnitude of bias increases to about six percent for mean net utility which corresponds to aggregate shares of about 20% and 80%. The magnitude of bias decreases with larger absolute values of mean net utility. The method of statistical differentials has a lower absolute bias for all possible values of mean net utility.* The bias by this method is zero at mean net utility equal to zero and also for mean net utility which corresponds to aggregate share of about 10% and 90%. For lower magnitudes of mean net utility the correction associated with the method of statistical differentials reverses the direction of bias of the naive method. The resulting bias is of lower magnitude than that for the naive procedure. The maximum bias in this range is about 3.5%. At higher values of absolute mean net utility the bias is less than one percent.

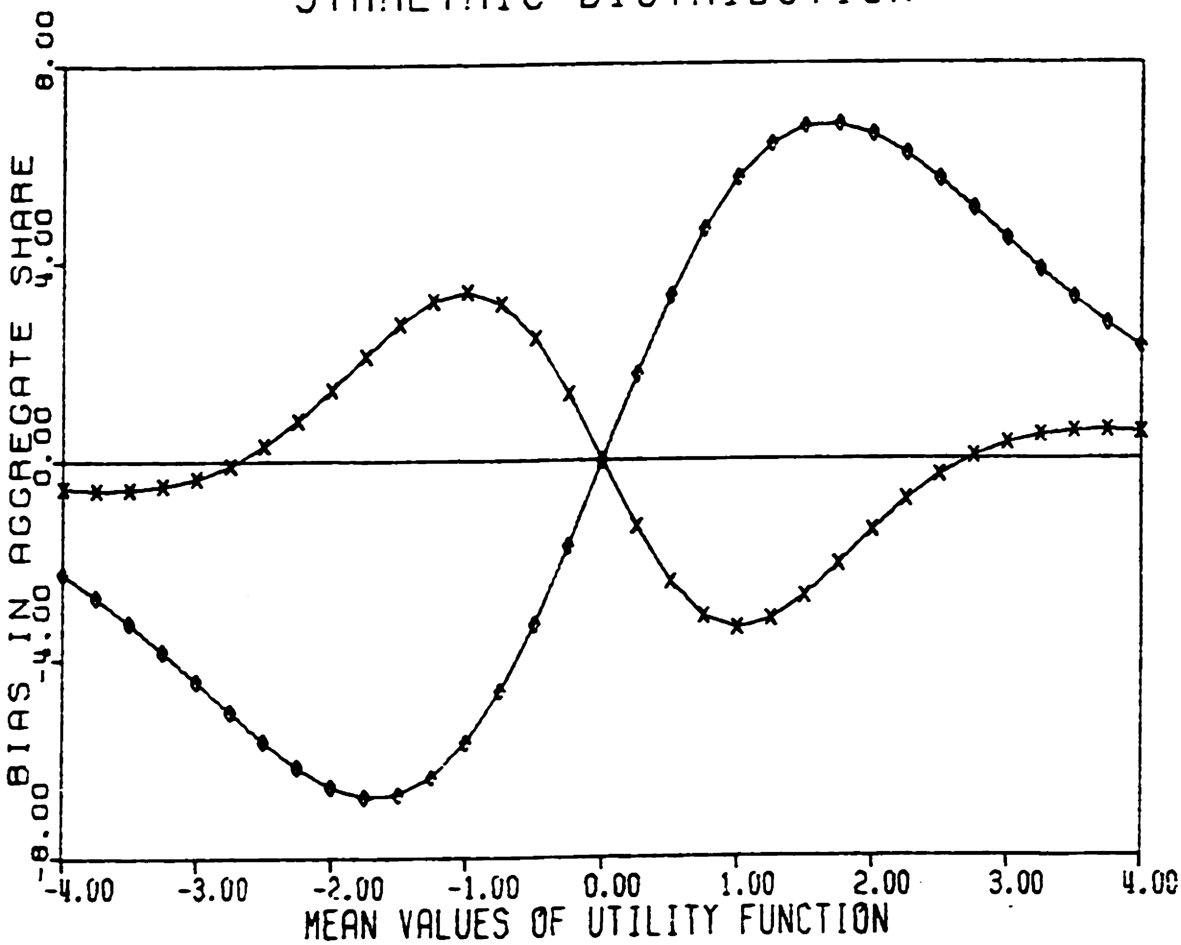
When the distribution of net utilities is skewed to the right

---

*This is true for variance equal to two. At higher levels of variance the statistical differentials procedure may result in greater bias than the naive procedure as shown in Chapter III.

FIGURE 4.5

BIAS IN ESTIMATES OF AGGREGATE SHARE
BY METHODS OF MOMENTS
SYMMETRIC DISTRIBUTION



AGGREGATION METHODS

——— NUM. INT. / KNOWN DISTRIBUTION

XXXX STAT. DIFF. W/ MEAN & VARIANCE

◊◊◊◊ NAIVE METHOD

TRUE DISTRIBUTION FORM

BINOMIAL (N=10, P=0.50)

STANDARDIZED - VARIANCE IS 2.0

SKEWNESS INDEX IS 0.0

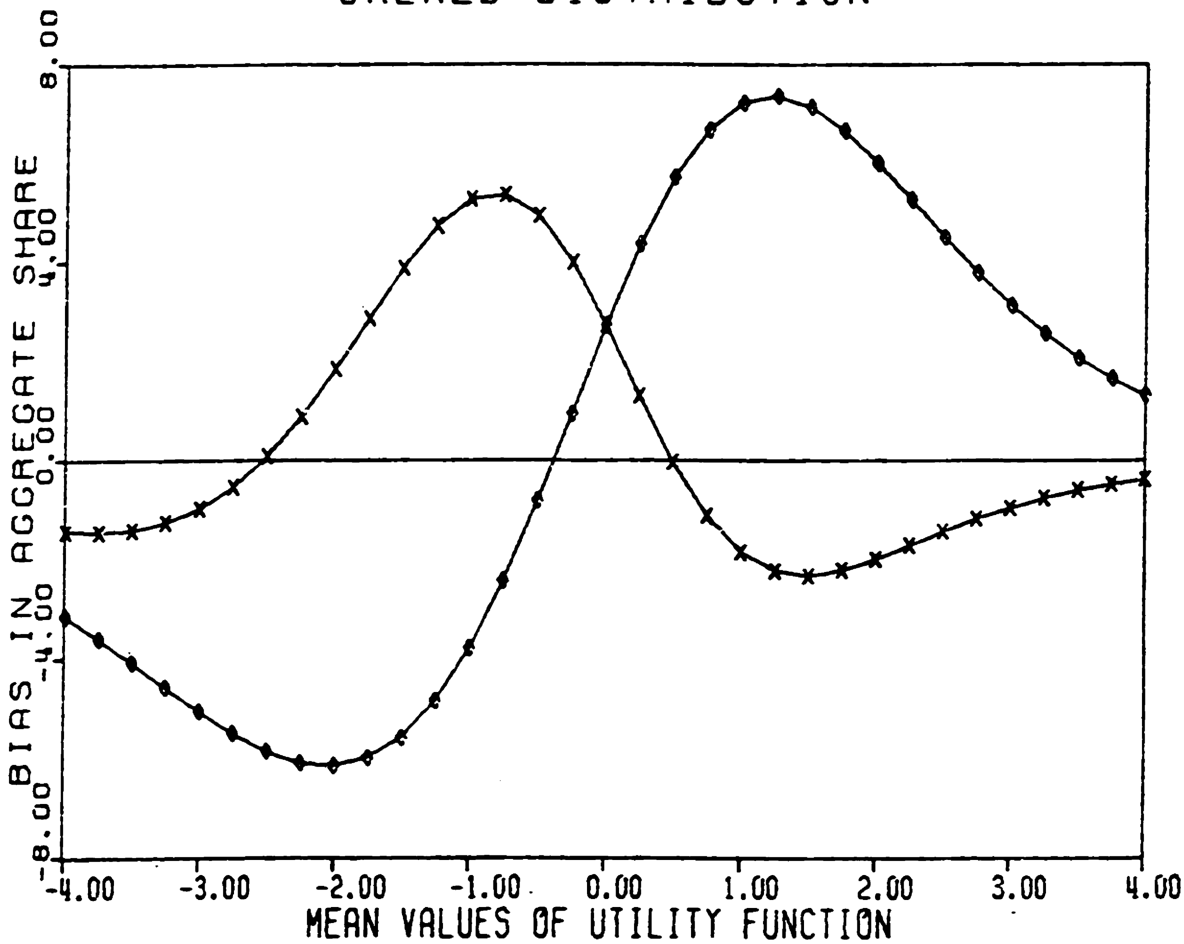
KURTOSIS INDEX IS 2.8

(positively skewed) all of the key points on the bias function of the naive method are shifted to the left (Figure 4.6). That is, the points corresponding to maximum negative bias, zero bias and maximum positive bias all occur at lower algebraic values of mean net utility. In addition, the maximum negative bias is reduced to about six percent and the maximum positive bias increases to about 7.5 percent. The opposite effect is observed for the method of statistical differentials. All three points of zero bias are shifted to the right. The maximum positive bias increases to about five percent and the maximum negative bias decreases to about two percent. One result of these shifts is that the naive method has lower magnitude of bias than the statistical differentials method for values of mean net utility which correspond to choice shares between twenty-five and fifty percent. When the distribution of net utilities is skewed to the left the affect on the bias structure is opposite to that when the distribution is skewed to the right.

Figures 4.7 and 4.8 illustrate the bias in slope estimates for methods of moments for symmetric and skewed distributions, respectively. The slope predictions for both the naive and statistical differentials procedures are zero at the same mean net utility values for which the share predictions (Figures 4.5 and 4.6) have maximum or minimum values. In addition, local maximum or minimum slope bias occurs at (for the symmetric distribution) or near (for the skewed distribution) the mean net utility values for which the share predictions are unbiased. The maximum biases in slope prediction are around plus seven percent and

FIGURE 4.6

BIAS IN ESTIMATES OF AGGREGATE SHARE BY METHODS OF MOMENTS SKEWED DISTRIBUTION



AGGREGATION METHODS

——— NUM. INT. / KNOWN DISTRIBUTION

xxxx STAT. DIFF. W/ MEAN & VARIANCE

◊◊◊◊ NAIVE METHOD

TRUE DISTRIBUTION FORM

BINOMIAL (N=10, P=0.05)

STANDARDIZED - VARIANCE IS 2.0

SKEWNESS INDEX IS 1.3

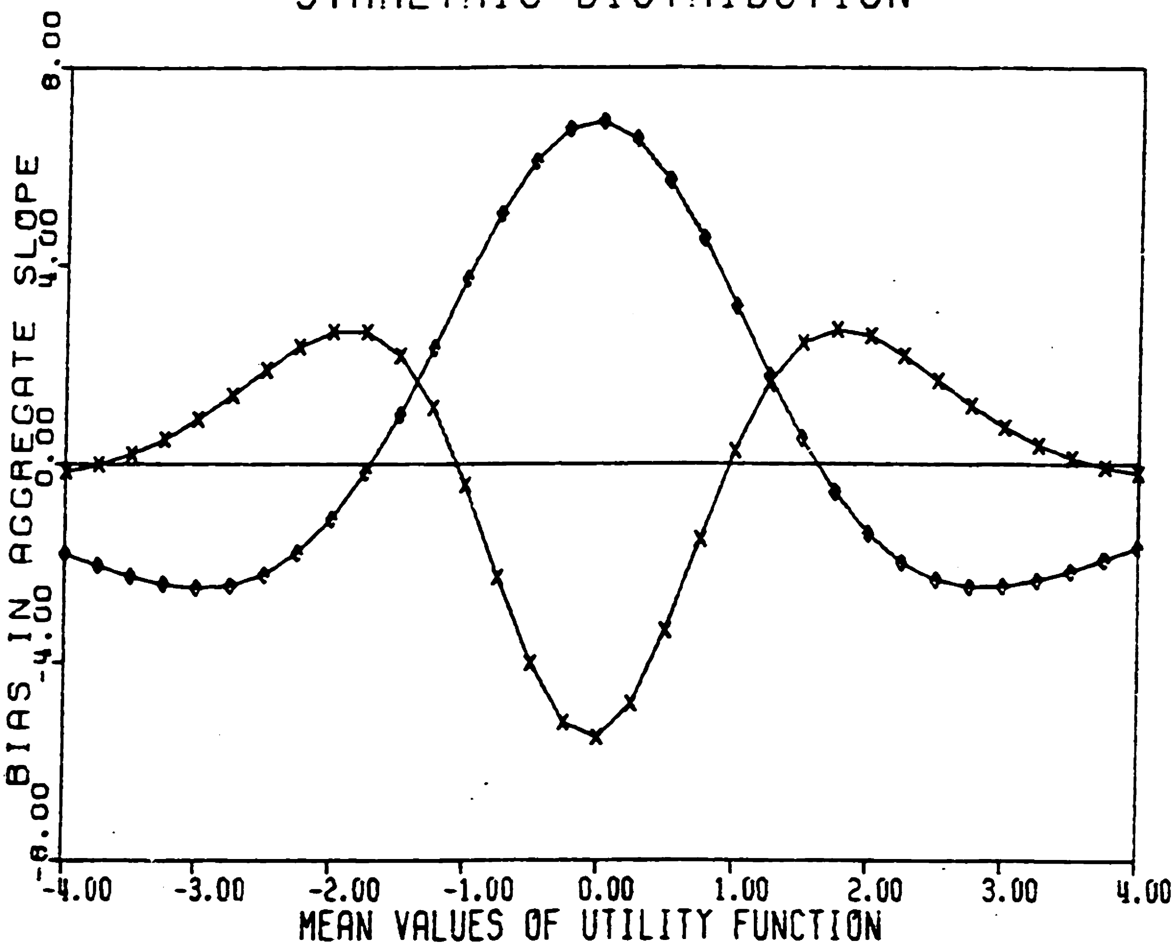
KURTOSIS INDEX IS 4.5

FIGURE 4.7

BIAS IN ESTIMATES OF AGGREGATE SLOPE

BY METHODS OF MOMENTS

SYMMETRIC DISTRIBUTION



AGGREGATION METHODS

———NUM. INT. / KNOWN DISTRIBUTION

XXXX STAT. DIFF. W/ MEAN & VARIANCE

◊◊◊◊ NAIVE METHOD

TRUE DISTRIBUTION FORM

BINOMIAL (N=10,P=0.50)

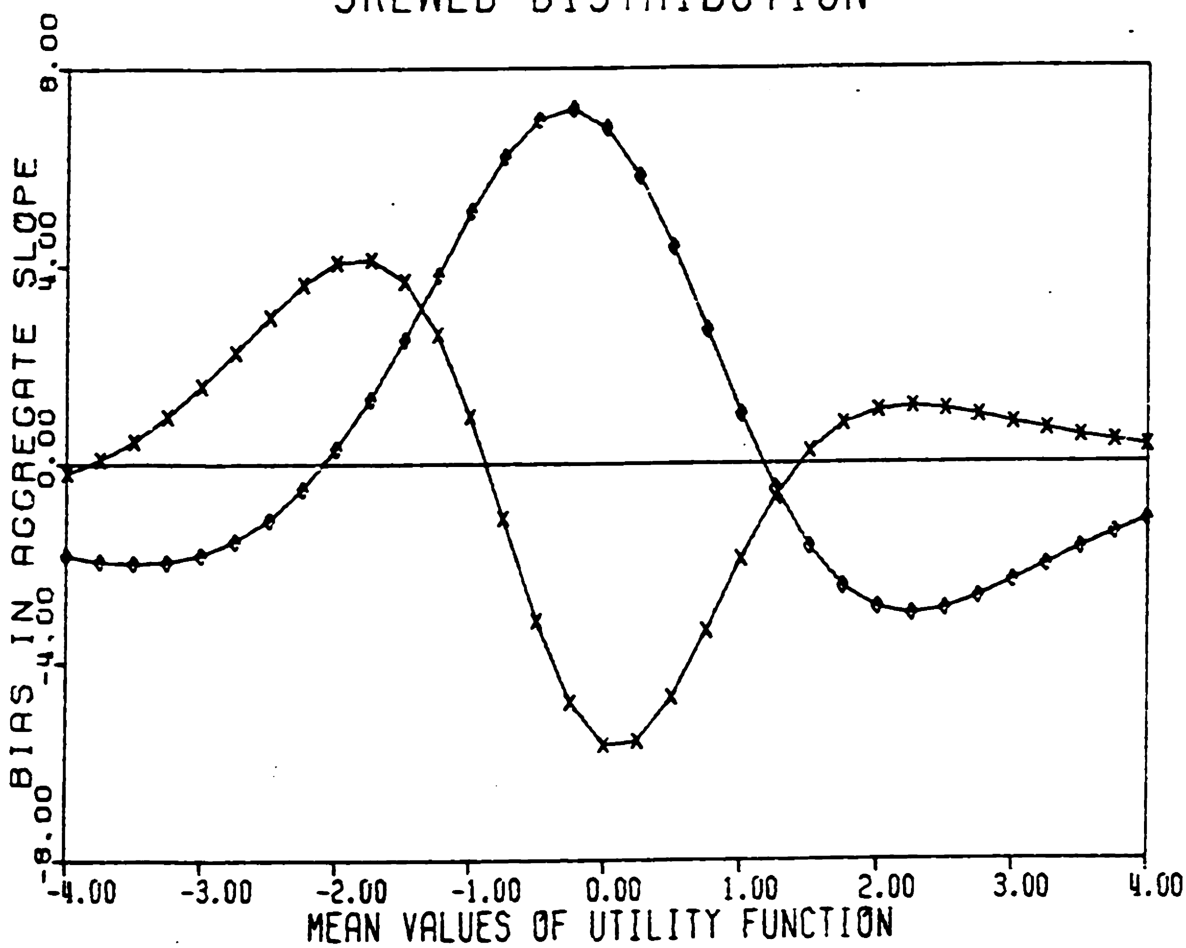
STANDARDIZED - VARIANCE IS 2.0

SKEWNESS INDEX IS 0.0

KURTOSIS INDEX IS 2.8

FIGURE 4.8

BIAS IN ESTIMATES OF AGGREGATE SLOPE BY METHODS OF MOMENTS SKEWED DISTRIBUTION



AGGREGATION METHODS

———— NUM. INT. / KNOWN DISTRIBUTION

XXXX STAT. DIFF. W/ MEAN & VARIANCE

◆◆◆◆ NAIVE METHOD

TRUE DISTRIBUTION FORM

BINOMIAL (N=10,P=0.05)

STANDARDIZED - VARIANCE IS 2.0

SKEWNESS INDEX IS 1.3

KURTOSIS INDEX IS 4.5

minus three percent for the naive procedure and plus three percent and minus six percent for the statistical differentials procedure.

These results are consistent with the mathematical analysis of bias described in Chapter III. The statistical differentials procedure has less bias in the share prediction than the naive procedure except for a narrow range of mean net utility values when the distribution of net utility values is skewed.* The statistical differentials procedure also has less bias in share sensitivity prediction than the naive procedure except for narrow ranges of mean net utility where the share sensitivity bias by the naive method is at or near zero. The correspondence between conditions for maximum share bias and zero slope bias and the near correspondence between the conditions for maximum slope bias and zero share bias identified in Chapter III are confirmed in the simulation results.

*This statement applies for variance equal to two. At higher levels of variance the bias by the statistical differentials procedure may be greater than that by the naive procedure (see sections 4.6 and 4.7).

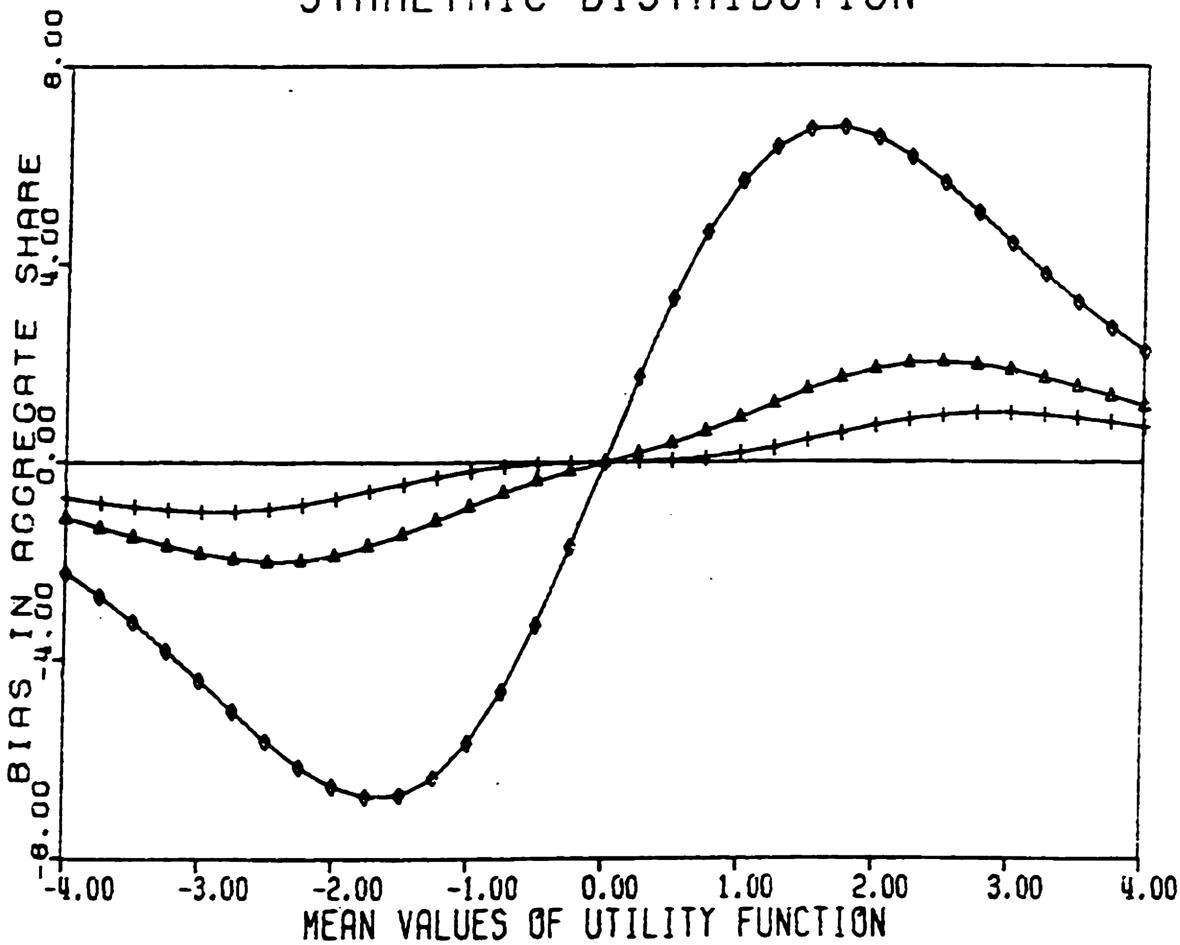
4.4 Aggregation Bias by Methods of Classification

This group includes the naive method, classification in two groups and classification in three groups. The bias of the naive method was discussed in the preceding section. This section is concerned with the bias of classification procedures. The share bias for symmetric distributions with variance equal to two is shown in Figure 4.9. The bias of classification in two sets is less than for the naive method and the bias of classification in three sets is less than for classification in two sets. As expected from the analysis of the preceding chapter the points of maximum bias occur at greater absolute values of the mean net utility as the number of classes increases. The maximum bias is reduced from seven percent for the naive procedure to approximately two percent and one percent for classification in two or three groups, respectively.

When the distribution of mean net utilities is skewed to the right (Figure 4.10) the location of the zero and maximum bias points for all three procedures are shifted to lower mean net utility values. The displacement of the zero bias point is monotonically related to the number of classes. The results of the shifts is to create ranges of values for mean net utility in which the magnitude of bias for classification in three sets is greater than for classification in two sets and the magnitude of bias for classification in two sets is greater than for the naive procedure. The magnitude of increased bias by procedures with

FIGURE 4.9

BIAS IN ESTIMATES OF AGGREGATE SHARE
BY METHODS OF CLASSIFICATION
SYMMETRIC DISTRIBUTION



AGGREGATION METHODS

——NUM. INT. / KNOWN DISTRIBUTION

ΔΔΔΔ CLASSIFIC. 2 SETS / MEDIAN

++++ CLASSIFIC. 3 SETS AT THIRDS

◊◊◊◊ NAIVE METHOD

TRUE DISTRIBUTION FORM

BINOMIAL (N=10,P=0.50)

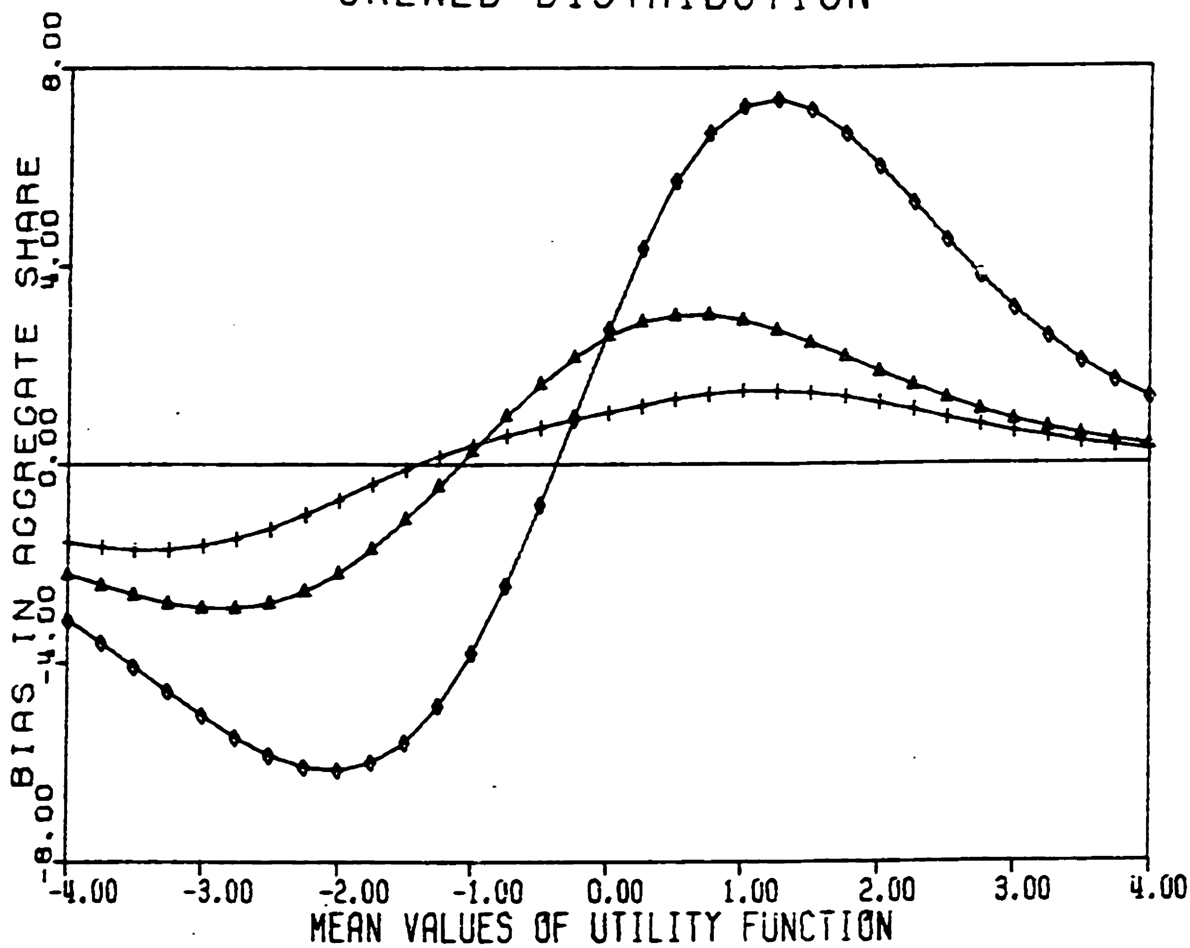
STANDARDIZED - VARIANCE IS 2.0

SKEWNESS INDEX IS 0.0

KURTOSIS INDEX IS 2.8

FIGURE 4.10

BIAS IN ESTIMATES OF AGGREGATE SHARE
BY METHODS OF CLASSIFICATION
SKEWED DISTRIBUTION



AGGREGATION METHODS

——— NUM. INT. / KNOWN DISTRIBUTION

▲▲▲▲ CLASSIFIC. 2 SETS / MEDIAN

++++ CLASSIFIC. 3 SETS AT THIRDS

◊◊◊◊ NAIVE METHOD

TRUE DISTRIBUTION FORM

BINOMIAL (N=10,P=0.05)

STANDARDIZED - VARIANCE IS 2.0

SKEWNESS INDEX IS 1.3

KURTOSIS INDEX IS 4.5

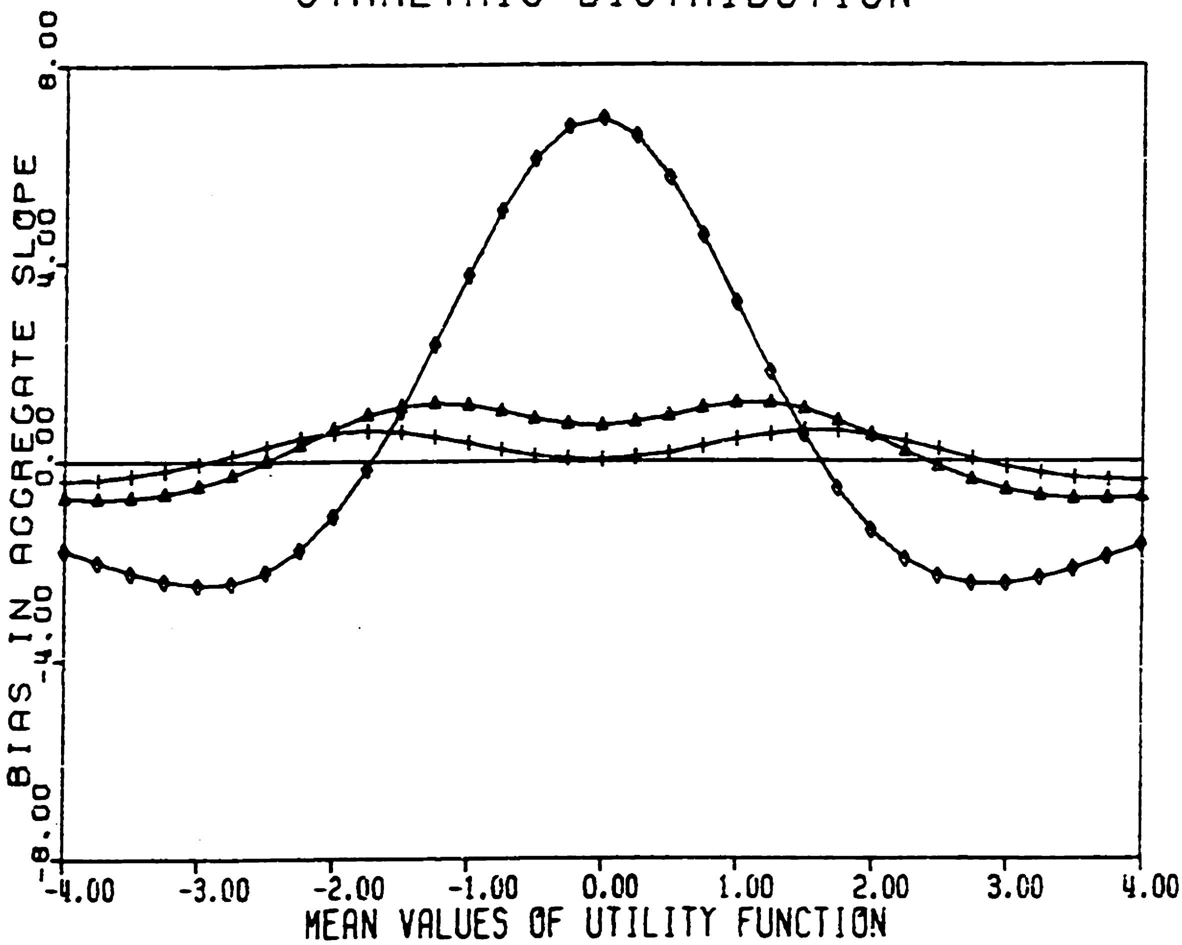
more classes is small when it occurs. The maximum biases by classification in two groups and three groups with skewed distribution increase to about three and one and one-half percent, respectively.

The slope bias by methods of classification is shown in Figures 4.11 and 4.12 for symmetric and skewed distributions, respectively. The maximum bias for both methods is under one and one-half percent with the symmetric distribution and less than three percent and one and one-half percent for the skewed distribution with two classes and three classes respectively. In some limited range of mean net utility values, the naive procedure has lower bias than either of the classification procedures when the distribution is skewed.

Overall, the classification procedures have considerably less bias than the naive procedure for prediction of both shares and slopes (share sensitivity). There are, however, narrow ranges of mean net utility values in which share predictions (for skewed distributions only) and slope predictions (for skewed and symmetric distributions) by classification procedures are slightly more biased than by the naive procedure.

FIGURE 4.11

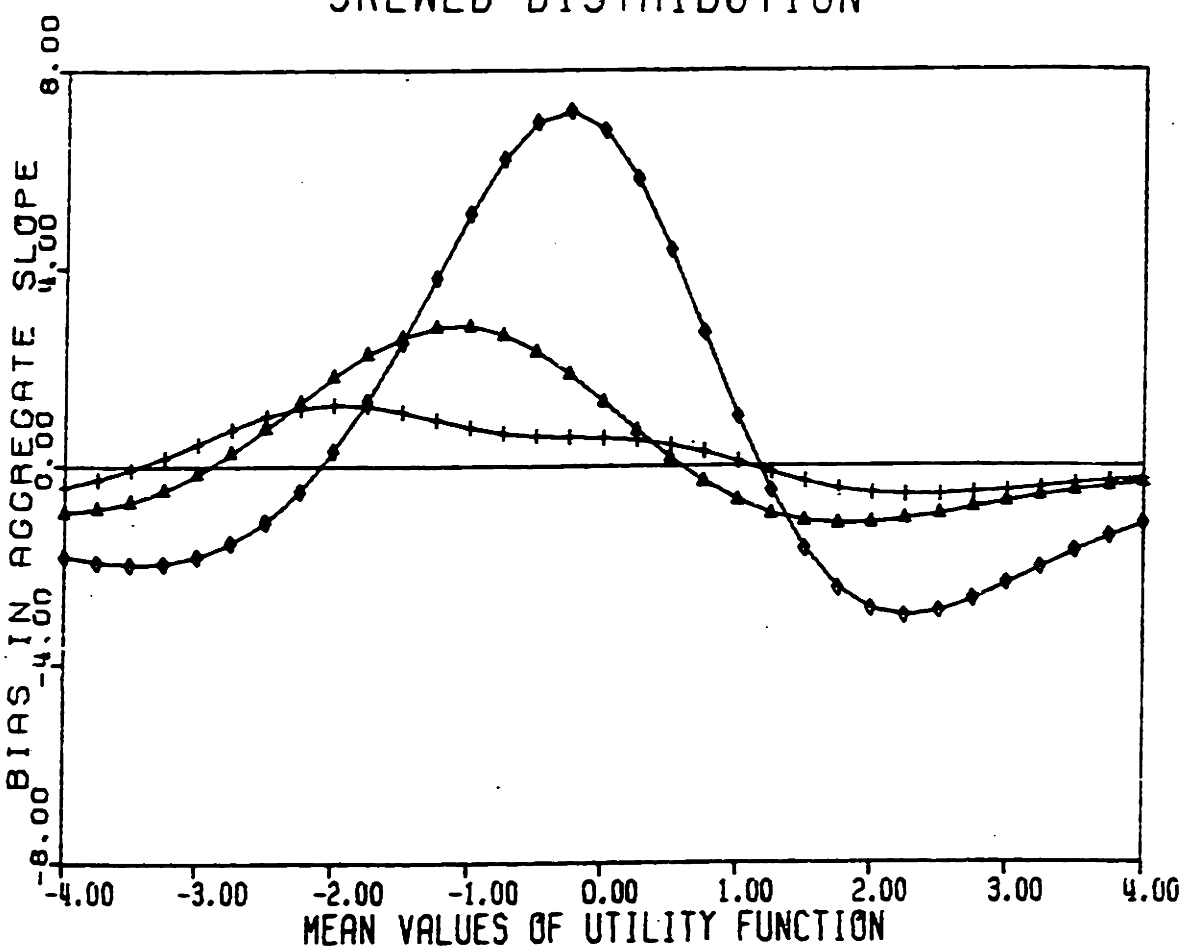
BIAS IN ESTIMATES OF AGGREGATE SLOPE
BY METHODS OF CLASSIFICATION
SYMMETRIC DISTRIBUTION



AGGREGATION METHODS	TRUE DISTRIBUTION FORM
——— NUM. INT. / KNOWN DISTRIBUTION	BINOMIAL (N=10,P=0.50)
ΔΔΔΔ CLASSIFIC. 2 SETS / MEDIAN	STANDARDIZED - VARIANCE IS 2.0
++++ CLASSIFIC. 3 SETS AT THIRDS	SKEWNESS INDEX IS 0.0
◊◊◊◊ NAIVE METHOD	KURTOSIS INDEX IS 2.8

FIGURE 4.12

BIAS IN ESTIMATES OF AGGREGATE SLOPE BY METHODS OF CLASSIFICATION SKEWED DISTRIBUTION



AGGREGATION METHODS

—— NUM. INT. / KNOWN DISTRIBUTION

ΔΔΔΔ CLASSIFIC. 2 SETS / MEDIAN

++++ CLASSIFIC. 3 SETS AT THIRDS

◊◊◊◊ NAIVE METHOD

TRUE DISTRIBUTION FORM

BINOMIAL (N=10,P=0.05)

STANDARDIZED - VARIANCE IS 2.0

SKEWNESS INDEX IS 1.3

KURTOSIS INDEX IS 4.5

4.5 Aggregation Bias by Methods of Integration

This group includes the naive method (integration over a homogeneous-point-distribution), integration with assumed normal distribution and integration with assumed uniform distribution.* The bias in share prediction by each of these methods for a symmetric distribution with variance of two is illustrated in Figure 4.13. All three methods have zero bias when the mean net utility is equal to zero. The magnitude of bias by integration with assumed normal distribution is nil over the entire range** and by integration with assumed uniform distribution is less than one percent over the entire range.

When the distribution of net utilities is skewed the bias by integration with assumed normal and uniform distributions have similar bias patterns (Figure 4.14). The maximum bias is under three percent for either method and the difference in bias between the two methods is always less than one percent. As we have seen before, there is a range of mean net utility values within which the naive method has lower bias than either of the distributed integrations when the distribution is skewed.

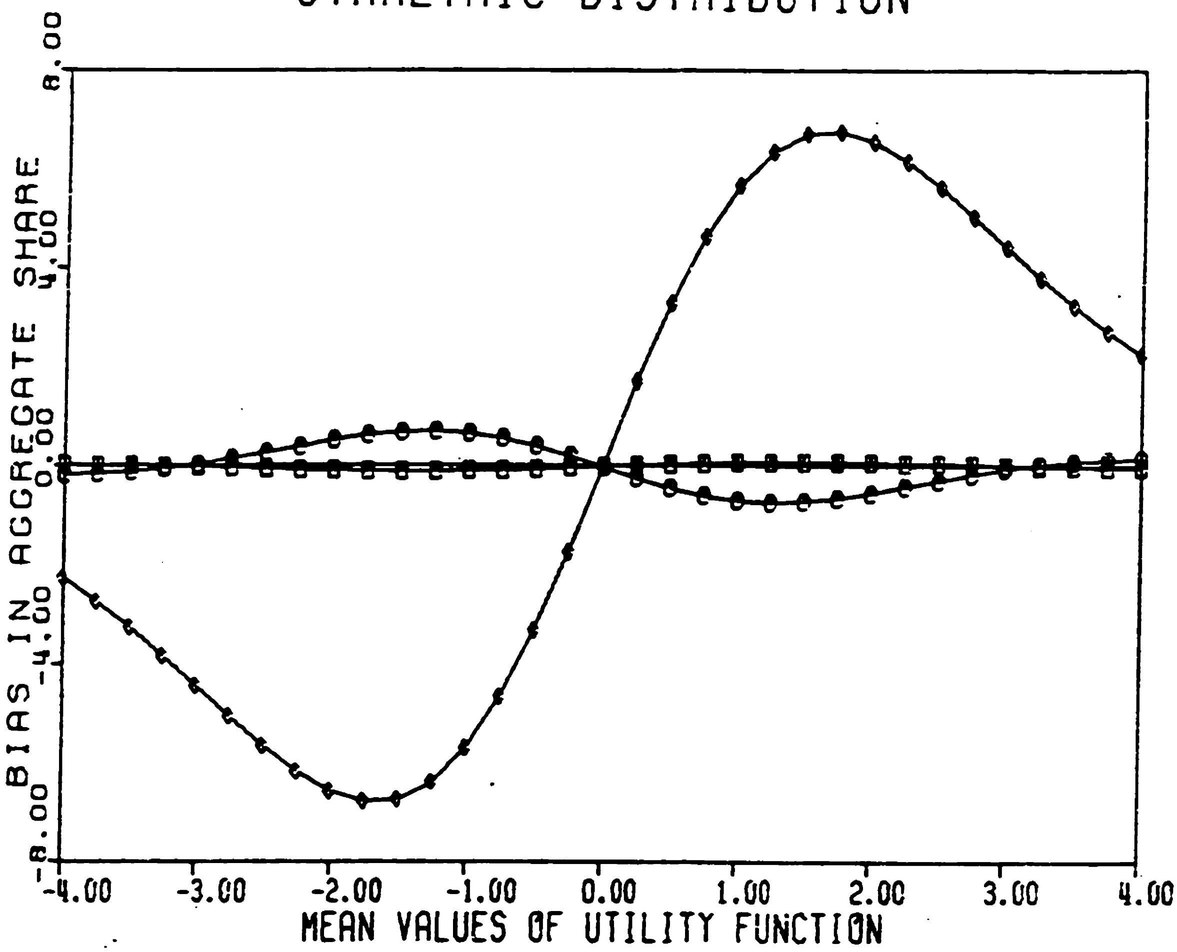
---

*The limits of the uniform distribution are determined so that its mean and variance are a_0 and a_1^2, respectively.

**The binomial distribution with large N is an excellent representation of the normal distribution and vice versa. Apparently N=10 is "large" when the binomial probability is 0.5.

FIGURE 4.13

BIAS IN ESTIMATES OF AGGREGATE SHARE BY METHODS OF INTEGRATION SYMMETRIC DISTRIBUTION



AGGREGATION METHODS

——— NUM. INT. / KNOWN DISTRIBUTION

□□□□ NUM. INT. / NORMAL DISTRIBUTION

○○○○ NUM. INT. / UNIFORM DISTRIBUTION

◇◇◇◇ NAIVE METHOD

TRUE DISTRIBUTION FORM

BINOMIAL (N=10,P=0.50)

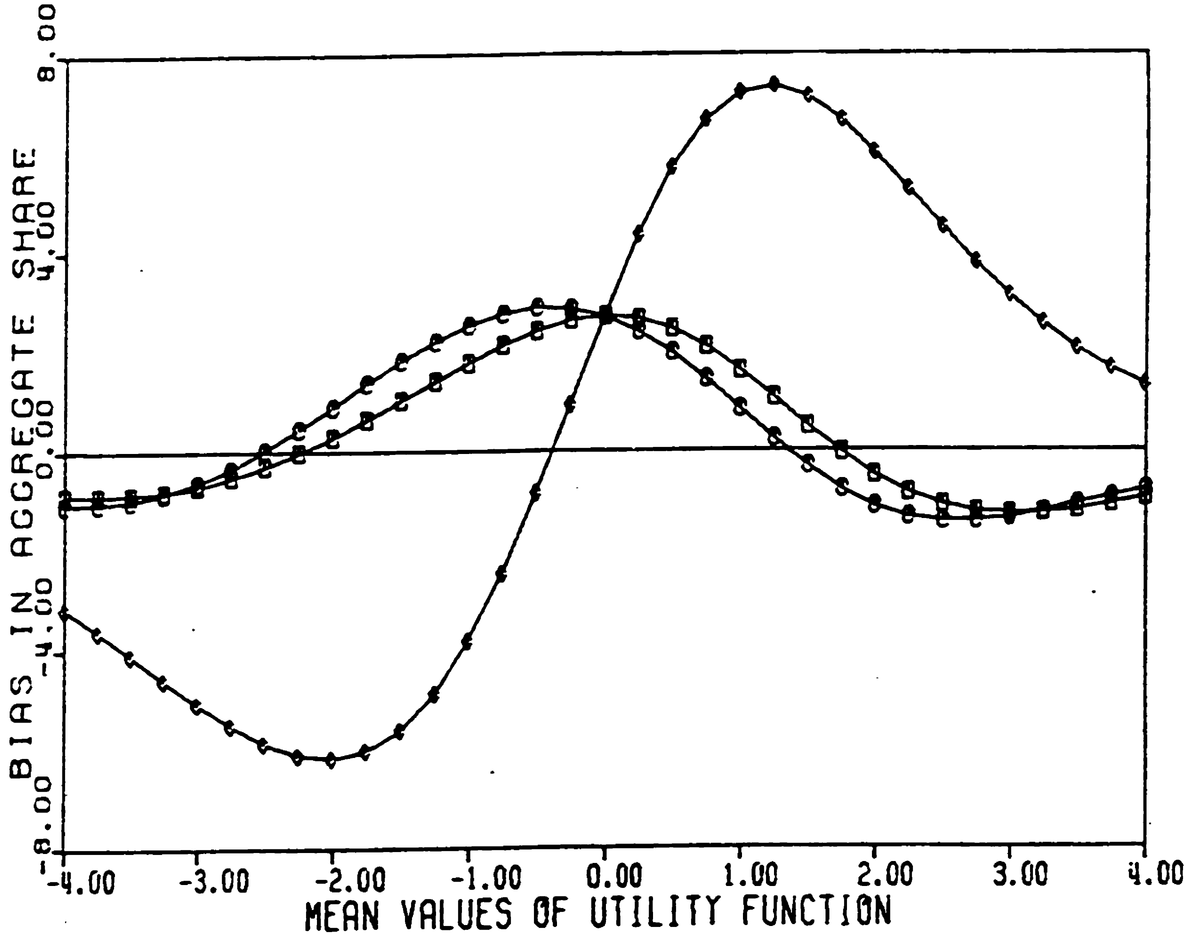
STANDARDIZED - VARIANCE IS 2.0

SKEWNESS INDEX IS 0.0

KURTOSIS INDEX IS 2.8

FIGURE 4.14

BIAS IN ESTIMATES OF AGGREGATE SHARE BY METHODS OF INTEGRATION SKEWED DISTRIBUTION



AGGREGATION METHODS

——— NUM. INT. / KNOWN DISTRIBUTION

NUM. INT. / NORMAL DISTRIBUTION

NUM. INT. / UNIFORM DISTRIBUTION

NAIVE METHOD

TRUE DISTRIBUTION FORM

BINOMIAL (N=10,P=0.05)

STANDARDIZED - VARIANCE IS 2.0

SKEWNESS INDEX IS 1.3

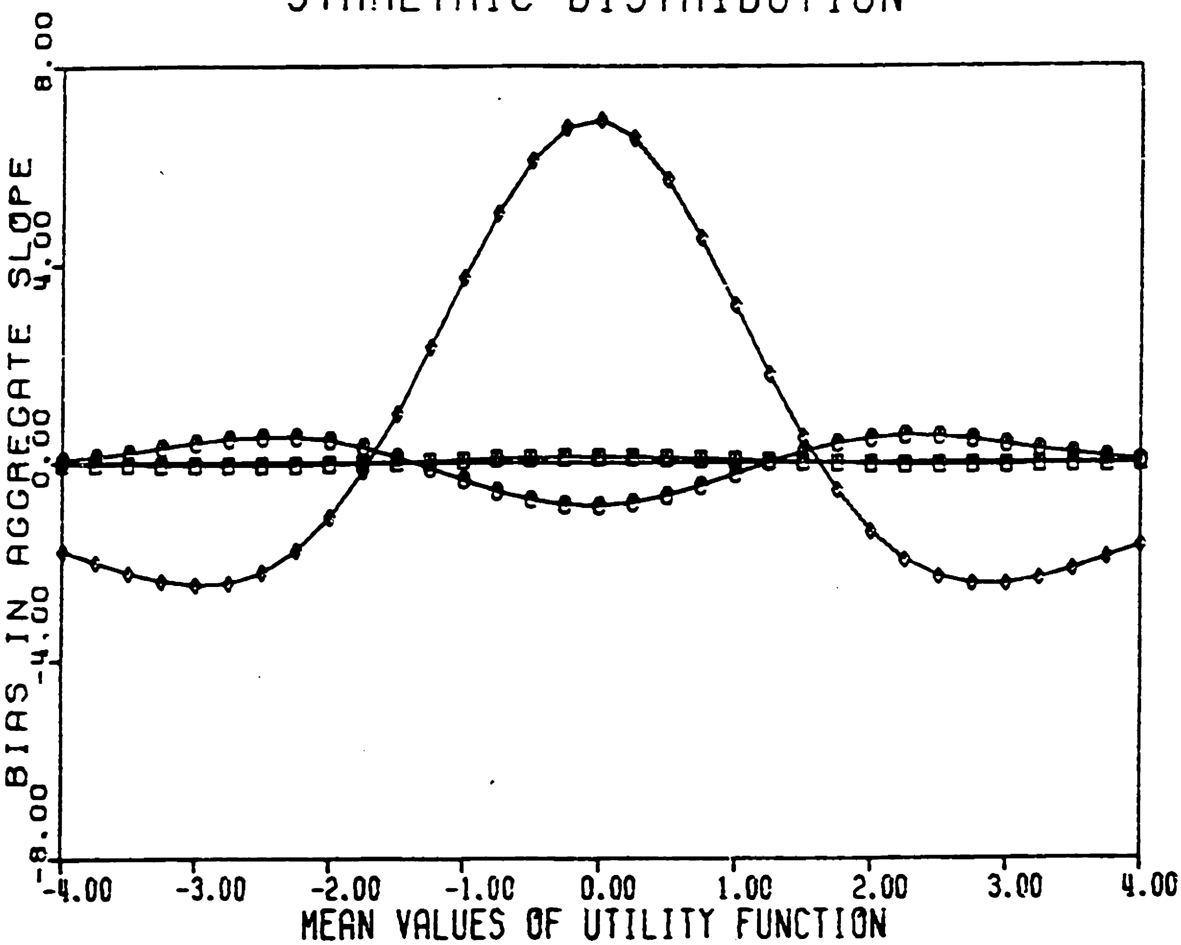
KURTOSIS INDEX IS 4.5

The slope bias for symmetric distributions with variance equal to two is illustrated in Figure 4.15. The slope bias for numerical integration with assumed normal distribution is nil and for integration with assumed uniform distribution is less than one percent over the entire range of mean net utility values. When the net utility distribution is skewed both methods have similar bias (less than one percent difference) over the entire range (Figure 4.16). The maximum bias of either method is less than one percent over the entire range of mean net utility values. The naive method has lower slope bias than either of the distributed integrations in very narrow ranges of mean net utility values for both symmetric and skewed distributions.

The generally low, and similar, bias of the methods of distributed integration as compared to the naive method indicate both the benefits of representing the distribution and the robustness of the distribution representation. Both methods have much lower share and slope bias than the naive method except in very narrow ranges near the points where the naive method gives unbiased share or slope estimates.

FIGURE 4.15

BIAS IN ESTIMATES OF AGGREGATE SLOPE BY METHODS OF INTEGRATION SYMMETRIC DISTRIBUTION



AGGREGATION METHODS

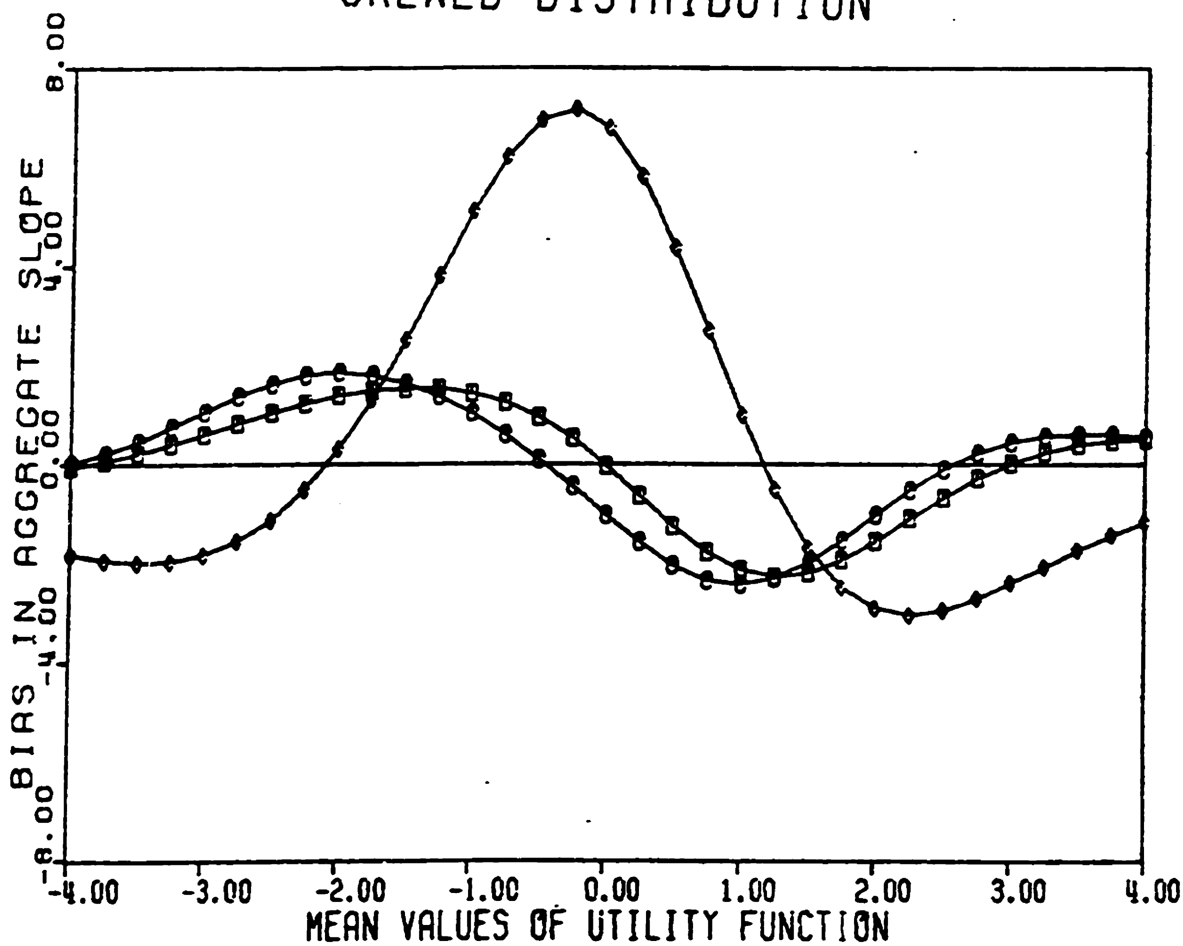
- NUM. INT. / KNOWN DISTRIBUTION
- NUM. INT. / NORMAL DISTRIBUTION
- NUM. INT. / UNIFORM DISTRIBUTION
- NAIVE METHOD

TRUE DISTRIBUTION FORM

BINOMIAL (N=10,P=0.50)
STANDARDIZED - VARIANCE IS 2.0
SKEWNESS INDEX IS 0.0
KURTOSIS INDEX IS 2.8

FIGURE 4.16

BIAS IN ESTIMATES OF AGGREGATE SLOPE BY METHODS OF INTEGRATION SKEWED DISTRIBUTION



AGGREGATION METHODS

——— NUM. INT. / KNOWN DISTRIBUTION

▫▫▫▫ NUM. INT. / NORMAL DISTRIBUTION

○○○○ NUM. INT. / UNIFORM DISTRIBUTION

◇◇◇◇ NAIVE METHOD

TRUE DISTRIBUTION FORM

BINOMIAL (N=10,P=0.05)

STANDARDIZED - VARIANCE IS 2.0

SKEWNESS INDEX IS 1.3

KURTOSIS INDEX IS 4.5

4.6 Absolute Bias Related to the Variance and Skewness of the Net Utility Distribution

In the three preceding sections, the structure of bias in share and slope prediction for different aggregation procedures has been described. In this section the bias in prediction of the different aggregation methods is compared. This is done by summarizing the bias structure of each method for both share and slope prediction in terms of:

- Maximum absolute bias, and
- Average absolute bias.

These error measures are computed for symmetric and skewed distributions with variance equal to one, two and three.

The maximum bias in share prediction for both distributions and three levels of variance is presented in Table 4.1, Maximum Bias in Share Prediction. The maximum bias in share prediction for each method increases with variance for both the symmetric and skewed distributions. The maximum bias for each method is larger for the skewed distribution than for the symmetric distribution at each level of variance. The ranking of methods by maximum bias with the symmetric distribution is: 1) integration with normal distribution, 2) integration with uniform distribution, 3) classification with three classes, 4) classification with two classes, 5) statistical differentials, and 6) naive, except at low levels of variance when the statistical differentials procedure has lower maximum bias than classification with two classes. When the distribution is skewed the rankings of the different methods according

TABLE 4.1

MAXIMUM BIAS IN SHARE PREDICTION
(percent share)

AGGREGATION	DISTRIBUTION SHAPE	SYMMETRIC			SKEWED		
METHOD	DISTRIBUTION VARIANCE	1.0	2.0	3.0	1.0	2.0	3.0
Integration - Normal		0.05	0.10	0.14	1.41	2.72	3.80
Integration - Uniform		0.28	0.72	1.14	1.51	2.89	3.89
Classification - Two Classes		1.28	2.02	2.58	1.78	3.01	4.00
Classification - Three Classes		0.64	1.00	1.30	1.09	1.71	2.19
Statistical Differentials		1.07	3.40	6.37	2.17	5.36	8.91
Naive		3.94	6.76	8.98	4.07	7.36	10.19

to their maximum bias in share prediction is changed by the shifting of classification with three classes from third best method to best method.

The average absolute bias in share prediction for each method is presented in Table 4.2, Average Bias in Share Prediction. As with maximum bias, the average bias is larger at higher levels of variance and for the skewed rather than symmetric distribution of net utility values. The rank ordering of the different methods in terms of average share bias is similar to that for maximum bias in share prediction except that there are more reversals in rank among the methods for one or more variance levels especially with the skewed distribution.

The maximum bias in slope prediction is given in Table 4.3, Maximum Bias in Slope Prediction. The rank ordering of methods in terms of maximum bias in slope prediction is similar to that for maximum bias in share prediction except that at the variance level of three the statistical differentials procedure has greater maximum bias than the naive procedure for both the symmetric and skewed distributions. Average bias in slope prediction is presented in Table 4.4, Average Bias in Slope Prediction. Except for minor variations, the relative performance of the different methods in terms of average bias in slope prediction is similar to that expressed in terms of maximum bias in share prediction.

Considering the four indices of maximum and average bias for share and slope prediction together, the naive procedure performs most poorly

TABLE 4.2

AVERAGE BIAS IN SHARE PREDICTION
(percent share)

AGGREGATION METHOD	DISTRIBUTION SHAPE	SYMMETRIC			SKEWED		
	DISTRIBUTION VARIANCE	1.0	2.0	3.0	1.0	2.0	3.0
Integration - Normal		0.02	0.05	0.08	0.59	1.21	1.72
Integration - Uniform		0.12	0.35	0.59	0.63	1.34	1.95
Classification - Two Classes		0.86	1.32	1.54	1.11	1.89	2.47
Classification - Three Classes		0.42	0.58	0.63	0.66	1.01	1.20
Statistical Differentials		0.43	1.43	2.76	0.78	1.95	3.30
Naive		2.51	4.55	6.24	2.47	4.44	6.11

TABLE 4.3

MAXIMUM BIAS IN SLOPE PREDICTION
(percent slope)

AGGREGATION METHOD	DISTRIBUTION SHAPE	SYMMETRIC			SKEWED		
	DISTRIBUTION VARIANCE	1.0	2.0	3.0	1.0	2.0	3.0
Integration - Normal		0.05	0.10	0.13	1.24	2.28	3.11
Integration - Uniform		0.42	0.89	1.22	1.34	2.43	3.24
Classification - Two Classes		0.98	1.17	1.59	1.90	2.84	3.40
Classification - Three Classes		0.38	0.62	0.91	0.92	1.27	1.64
Statistical Differentials		1.85	5.53	10.01	2.24	5.66	9.76
Naive		4.38	6.93	8.68	4.39	7.19	9.34

TABLE 4.4

AVERAGE BIAS IN SLOPE PREDICTION
(percent slope)

AGGREGATION METHOD	DISTRIBUTION SHAPE / DISTRIBUTION VARIANCE	SYMMETRIC			SKEWED		
		1.0	2.0	3.0	1.0	2.0	3.0
Integration - Normal		0.02	0.05	0.07	0.52	0.97	1.34
Integration - Uniform		0.15	0.39	0.61	0.57	1.09	1.49
Classification - Two Classes		0.52	0.72	0.82	0.72	1.15	1.43
Classification - Three Classes		0.25	0.32	0.42	0.40	0.56	0.68
Statistical Differentials		0.57	1.80	3.35	0.81	2.02	3.41
Naive		1.69	2.79	3.55	1.66	2.78	3.65

of all the procedures except for the statistical differentials procedure in terms of maximum bias in slope prediction when the distributions have high levels of variance. The next worst procedure is the statistical differentials procedure. Although this procedure reduces the maximum and average bias in share prediction by significant amounts, it has a small effect on the maximum and average bias in slope prediction. Both of these procedures are particularly sensitive to the variance of the distributions of net utility values. The other four procedures considered have significantly lower values of bias as measured by all four indices and do not increase in bias as rapidly as the naive or statistical differentials procedures with increasing variance. In general, both of the integration procedures have lowest bias when distributions are symmetric followed by the two classification procedures. However, when the distributions are skewed, classification with three classes usually has lower bias than either of the integration procedures.

4.7 Absolute Bias per Unit of Prediction Related to the Variance and Skewness of the Net Utility Distribution

The preceding discussion is entirely based on the analysis of absolute bias in prediction. Little direct attention is given to comparison of the magnitude of bias with the magnitude of the prediction. In this section the bias in share and slope prediction is measured as a proportion of the magnitude of the corresponding prediction. The bias structure for the different prediction procedures is summarized in terms of the average absolute per unit bias. The maximum absolute per unit bias is not included as the true maximum ratio for all of the prediction methods occurs for mean net utility values lower than the range of values included in the simulation.*

The average absolute per unit bias in share prediction is presented in Table 4.5, Average Absolute Bias in Share Prediction per Unit of Prediction. The bias measure reported in this manner has, as expected, significantly larger numerical values than the average or maximum bias described earlier. The per unit bias for each procedure is larger for the skewed distribution than the symmetric distribution at each level of variance and increases with variance for both distributions. The overall ranking is (1) numerical integration with normal distribution,

---

*For the naive and classification methods the bias ratio increases as mean net utility decreases to minus infinity.

TABLE 4.5

AVERAGE ABSOLUTE BIAS IN SHARE PREDICTION
PER UNIT OF PREDICTION
(PER CENT OF SHARE PREDICTION)

AGGREGATION METHOD	DISTRIBUTION SHAPE	SYMMETRIC			SKEWED		
	DISTRIBUTION VARIANCE	1.0	2.0	3.0	1.0	2.0	3.0
Integration - Normal		0.1	0.2	0.3	2.9	4.5	5.2
Integration - Uniform		0.5	1.2	1.8	3.2	5.2	6.1
Classification - Two Classes		5.4	8.5	9.8	9.0	14.3	16.7
Classification - Three Classes		2.8	3.9	4.1	5.4	7.1	6.8
Statistical Differentials		1.6	4.7	8.0	4.2	8.2	11.3
Naive		16.6	32.6	47.4	19.1	36.5	51.2

(2) numerical integration with uniform distribution, (3) classification with three classes, (4) statistical differentials, (5) classification with two classes and naive. At low levels of variance the statistical differentials procedure has less bias than classification with three classes. The naive procedure has extremely high values of bias compared to any of the other procedures. The statistical differentials procedure is most sensitive to increases in the variance of the net utility distributions.

The corresponding bias measure for slope prediction is given in Table 4.6, Average Absolute Bias in Slope Prediction per Unit of Slope Prediction. The pattern of error measures in slope prediction are similar to those for share prediction with higher magnitudes. The major exception to this is that the statistical differentials procedure moves from fourth to fifth ranked position.

The results of this analysis are generally consistent with those of the preceding sections except for the increased magnitudes of errors which result from comparison with the magnitude of the corresponding prediction.

TABLE 4.6

AVERAGE ABSOLUTE BIAS IN SLOPE PREDICTION
PER UNIT OF SLOPE PREDICTION
(PER CENT OF SLOPE PREDICTION)

AGGREGATION	DISTRIBUTION SHAPE	SYMMETRIC			SKEWED		
METHOD	DISTRIBUTION VARIANCE	1.0	2.0	3.0	1.0	2.0	3.0
Integration - Normal		0.2	0.5	0.7	4.5	8.4	12.1
Integration - Uniform		1.2	3.5	5.8	5.0	9.8	14.0
Classification - Two Classes		6.5	8.6	9.0	8.0	11.3	13.2
Classification - Three Classes		3.1	3.8	4.3	4.5	5.9	7.3
Statistical Differentials		4.2	14.2	33.1	6.6	16.7	34.5
Naive		21.7	37.5	48.5	20.5	34.6	45.3

4.8 Summary and Conclusions

The simulations of bias of different methods of aggregation described in this chapter reinforces the results described in Chapter III based on the mathematical analysis of bias in aggregation. That is, the importance of the shape and variance of the distribution and its mean value on the bias by different methods of aggregation is confirmed. In addition, the simulation analysis provides some numerical values for the expected bias in aggregation by different methods applied in different situations. In particular, a general rank ordering of methods is established with recognition that the rank ordering does not apply in all situations.

The numerical integration and the classification procedures generally perform better than the naive and statistical differentials procedures. Within each group, however, the procedure which has the minimum bias either for a particular situation or the minimum value of average or maximum bias over a range of situations depends on the variance and shape of the distribution of net utility values.

CHAPTER V

ANALYSIS OF ERRORS IN PREDICTION

5.1 Introduction

Predictions of future travel behavior and of the performance of alternative transportation systems is needed by transportation planners and decision makers in order to make judgements about the desireability of alternative transportation plans. The usefulness of predictions, and consequently of prediction methods, depends on their accuracy. The importance of errors in prediction is derived from the cost to the decision maker and/or society of making an erroneous decision due to inaccuracy in prediction. This chapter attempts to identify the sources of error in the prediction process, to describe the way in which errors in various stages of the prediction process contribute to the overall error in prediction and to identify a set of measures which may be used to evaluate the performance of alternative predictive procedures. In addition, it (1) examines the structure of prediction errors and (2) identifies the portion of prediction errors which are contributed by those elements in the predictive process which are to be evaluated. In particular, the discussion is directed toward the analysis of errors introduced by the aggregation procedure.

The chapter is organized in five sections. Section 5.2 describes the structure of the model formulation and prediction process and identifies the sources of error in this process. Section 5.3 describes the interaction and propagation of errors from different sources to identify their effect on the overall prediction errors. Section 5.4 describes a strategy for the use of error measures in identifying the relative contribution of the errors contributed by the three major components of the prediction model: the disaggregate choice model, the distribution of choice influencing variables and the aggregation procedures. Section 5.5 describes a series of error measures which may be used to evaluate the performance of different prediction methods.

5.2 Sources of Error in the Model Formulation and Prediction Process

Prediction of future travel behavior is 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 which are independent of changes in socio-economic activity system and transportation service characteristics, to relatively complex relationships regarding 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 and estimation of model parameters, to prediction of future travel demand.

The application of the model formulation and prediction process varies considerably depending on the hypotheses on which the models are based. At one extreme, the no change hypothesis compresses the entire structure to the simple element of setting the prediction equal to the measured level of travel. The trend model modifies this by adding a simple growth factor which is applied to the measured level of travel to obtain the predicted level of travel. Correlative models require the specification of some relationships between travel and some socio-economic characteristics and, on occasion, some transportation system variables. This approach then requires the prediction of the included socio-economic variables and the selection of future service variables.

The future level of travel is predicted by assuming that its relationship to these variables is as previously estimated.

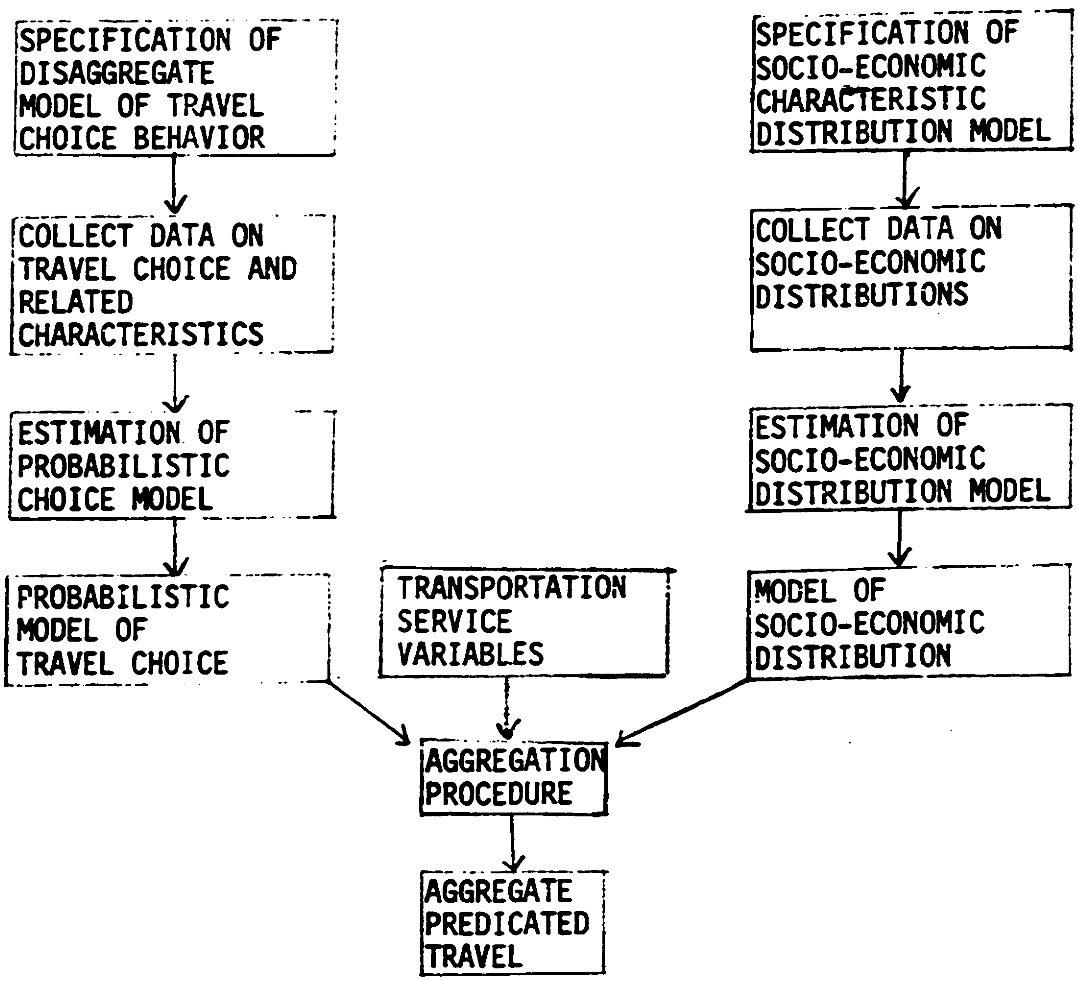
The case of interest here is the prediction of the share of travelers choosing each of a precise set of alternatives based on a probabilistic model of individual choices. The shares predicted may be described in terms of all of the characteristics of travel choices (frequency, destination, mode, time of day, route, etc.) and thus take the form of a direct demand model or may be in terms of a single element of travel choice and thus take the form of a component in a share model (Manheim, 1970). In either instance the model formulation and prediction process to be used, illustrated in Figure 5.1, Model Formulation and Prediction with Disaggregate Choice Models, consists of:

1. A disaggregate choice model of travel choice behavior based on observations of the characteristics of and facing individual travelers and their actual choices estimated by maximum likelihood methods.
2. An independent variable model designed to predict information on the distribution of socio-economic and level of service characteristics. It is based on data describing individual travelers and the alternatives they face collected in groups. The estimation procedure is designed to obtain information, at different levels of detail, on the distribution of those characteristics which are included in the choice model.
3. The aggregation procedure uses the information on the distri-

FIGURE 5.1

MODEL FORMULATION AND PREDICTION

WITH DISAGGREGATE CHOICE MODELS



bution of choice influencing variables in conjunction with the probabilistic choice model to obtain predictions of aggregate shares.

The utility function for an alternative to a particular individual is represented as linear in parameters with additive disturbance (LPAD) as described by Manski (1973), that is,

$$V_{it} = X_{it}'b + \epsilon_{it} \tag{5.1}$$

Errors in the choice model come from misspecifications of the utility functions and/or from errors in the measurement of independent variables. Manski classifies these errors into the following categories:

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

Manski shows that each of these sources of error may be represented as a real valued random variable which is consistent with the LPAD structure. He also demonstrates that the errors may not satisfy the independent, identically distributed assumption which underlies the development of common choice models such as logit or probit. The choice model parameter estimates contain random error due to the presence of any of the four sources of error described.

If the assumption that the disturbance term is independent and identically distributed holds, the parameter estimates using the maximum likelihood method are consistent and the asymptotic variance-covariance matrix is minus one times the inverse of the matrix of second derivatives of the log likelihood function (Theil, 1971).

If the independent identically distributed condition fails due to correlation between the disturbance term and included variables or between pairs of disturbance terms, the parameter estimates are biased.

Errors associated with attempts to apply the model calibrated on one data set (collected in one area during one time period) for prediction in a different time or place (transferability errors) are also included in the category of specification errors. Errors of transferability are generally 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 situations). The increased potential for transferability of disaggregate models has been widely assumed and is often cited as an important characteristic of disaggregate models. However, the hypothesis of transferability has not been extensively tested and is subject to verification.* Ultimately the question of transferability depends on the ability to properly and fully specify the relevant utility functions, the quality of the data used and the reliability of the parameter estimates. In any case, the use of disaggregate models avoids the aggregation specific nature of models calibrated on grouped data.

The present study is primarily concerned with the issue of prediction of travel demand based on disaggregate choice models. The choice influencing variables are assumed to be predicted by an independent model system. These predictions include errors introduced in the various stages of the model formulation and prediction process. These errors can be represented by a random additive disturbance term and a bias term. That is:

$$X_t^o = X_t + \delta_t + \gamma_t \tag{5.2}$$

where X_t^o ..... is a vector of observed or predicted variables,

X_t ..... is a vector of true variables,

δ_t ..... is a vector of random disturbance terms, and

γ_t ..... is a vector of bias terms.

*A recent study by Atherton (1975) determined 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, Mass.. Furthermore, the predictive performance of both models was similar.

Errors in aggregation are different from errors in parameter estimates or variable predictions as they are deterministic and thus introduce no random errors (though they propagate random errors from the choice model and the variable prediction model to the aggregate prediction). The errors introduced by imperfect aggregation procedures are structural errors which cause bias in the predictions obtained.

The errors coming from these sources are propagated through the model formulation and prediction process and, under some circumstances, interact in such a way as to increase or reduce the errors in aggregate prediction. The next section analyzes the propagation and interaction of the errors generated in the different components of the overall model.

5.3 Interaction and Propagation of Errors

Errors introduced at any stage in the model formulation and prediction process are propagated to succeeding stages. The method and effect of this propagation depends on the type of error and the way in which it enters into the model structure. In addition errors from different sources in the same element or in separate elements interact to increase or diminish the affect on the total error of prediction. The interactions are due to both the structure of the model, the relationship between model elements and the statistical properties of the data sets used for model estimation and for prediction.

The discussion of the interaction and propagation of errors covers:

- Random errors in choice model parameters and choice variables,
- Bias errors in choice model parameters and choice variables,
- Bias errors in aggregation.

The discussion focuses on the binary choice case for purposes of simplicity of notation and exposition. The choice model in the binary case is represented by:

$$P_t(i) = f^i(X_t'b) \tag{5.3}$$

where $P_t(i)$ ..... is the probability of individual t choosing alternative i,

$f^i(\ )$ ..... represents the form of the choice model, and

$X_t'b$ ..... represents the linear additive net utility of alternative i over the other alternative in the choice set.

The aggregate prediction model is:

$$S_i = \frac{1}{T} \sum_t P_t(i) \tag{5.4}$$

where S_i ..... is the aggregate share of the group T choosing alternative i,

T ..... is the number of members of group T, and

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

This model represents the enumeration aggregation procedure which was previously identified as the theoretically correct aggregation procedure (Chapter II). Errors of variability of prediction expressed in terms of the expected variance are considered first. Errors of bias will be discussed separately.

The variance of the aggregate share prediction is related to the variance of individual choice probabilities and the covariance of choice probabilities for every pair of individuals in the group:

$$\mathrm{Var}(S_i) = \frac{1}{T^2}\left[\sum_t \mathrm{Var}(P_t) + \sum_t \sum_{t' \neq t} \mathrm{Cov}(P_t, P_{t'})\right]^* \tag{5.5}$$

The variance of the individual choice probabilities is given

*The expression $P_t(i)$ is simplified to P_t.

approximately (Kendall and Stuart, 1969; Tukey, 1957) by:

$$\mathrm{Var}(P_t) = \left[\frac{dP_t}{dX_t'b}\right]^2 \mathrm{Var}(X_{t'}b) \tag{5.6}$$

Similarly, the covariance between pairs of probability estimates is given approximately by:

$$\mathrm{Cov}(P_t,P_{t'}) = \left(\frac{dP_t}{dX_t'b}\right)\left(\frac{dP_{t'}}{dX_{t'}'b}\right) \mathrm{Cov}(X_t'b,X_t'b) \tag{5.7}$$

That is, the variance of the aggregate share estimate depends on the variance of individual utility functions, the covariance of the utility functions for pairs of individuals and, through their effect on the choice derivatives, the values of the utility estimates.

The variance of the utility function may be disaggregated into the variance contributed by errors in parameters and variance contributed by the errors in choice variables:

$$\mathrm{Var}(X_t'b) = X_t' \,\mathrm{Var}(b)X_t + b\,\mathrm{Var}(\delta_t)b' + 1'\mathrm{Var}(b)\mathrm{Var}(\delta_t)1^{*}$$

$$= X_t' AX_t + b\,Y_t b' + 1'AY_t 1 \tag{5.8}$$

*This formulation assumes that there is no interacting variance between the model parameters and the choice variables.

where A ..... is the variance-covariance matrix for parameters,

Y_t ..... is the variance-covariance matrix for error in variables for individual t, and

1 ..... is a sum vector with unit value for each of its elements.

Similarly, the covariance of utility values between pairs of individuals can be disaggregated into that contributed by errors in the parameter and errors in choice variables.

$$Cov(X_t'b, X_{t'}'b) = X_t'Cov(b,b)X_{t'} + bCov(\delta_t, \delta_{t'})b' + 1'Cov(b,b)Cov(\delta_t, \delta_{t'})1$$

$$= X_t'AX_{t'} + bY_{tt'}b' + 1AY_{tt'}1 \qquad (5.9)$$

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

$Cov(b,b) = Var(b) = A$

In both cases, the third term, which represents the interaction between the variance in parameters and the variance in variables, is small relative to the other terms and will not be considered further. The variance in individual choice probabilities and covariance of choice probabilities between pairs of individuals is:

$$Var(P_t) = \left[\frac{dP_t}{dX_t'b}\right]^2 \left(X_t' A X_t + b Y_t b'\right) \qquad (5.10)$$

and

$$\mathrm{Cov}(P_t, P_{t'}) = \left(\frac{dP_t}{dX_t' b}\right)\left(\frac{dP_{t'}}{dX_{t'}' b}\right)\left[X_t' A X_{t'} + bY_{tt'} b'\right] \tag{5.11}$$

Finally, the variance of the aggregate share in terms of random errors in parameters and choice variables is:

$$\mathrm{Var}(S_i) = \frac{1}{T^2}\left\{ \sum_t \left(\frac{dP_t}{dX_t' b}\right)^2 \left(X_t' AX_t + bY_t b'\right) + \sum_t \sum_{t' \neq t} \left(\frac{dP_t}{dX_t' b}\right)\left(\frac{dP_{t'}}{dX_{t'}' b}\right)\left(X_t' AX_{t'} + bY_{tt'} b'\right) \right\} \tag{5.12}$$

For a relatively homogeneous population (similar derivatives, equal variance among individuals and equal covariance among pairs of individuals this simplifies to:

$$\mathrm{Var}(S_i) = \frac{1}{T}\left(\frac{dP_t}{dX_t' b}\right)^2 \left(X_t' AX_t + bY_t b'\right) + \left(\frac{T-1}{T}\right)\left(\frac{dP_t}{dX_t' b}\right)^2 \left(X_t' AX_{t'} + bY_{tt'} b'\right)$$

$$= \left(\frac{dP_t}{dX_t'b}\right)^2 \left\{ \frac{1}{T}\left[X_t'AX_t + bY_tb'\right] \right.$$

$$\left. + \left(\frac{T-1}{T}\right)\left[X_t'AX_{t'} + bY_{tt'}b'\right] \right\} \qquad (5.13)$$

The importance of the covariance in errors of choice probability estimates results from the substantially greater number of terms in the summation of covariances, T^2-T, than in the summation of variances, T. In the extreme cases where there is no covariance in errors in probability estimates between individuals or where covariance for pairs of individuals is equal to their variance* the variance in share estimates varies by a factor of T.

The variance in the former case is given by:

$$\mathrm{Var}(S_i) = \frac{1}{T^2}\left[T\left(\frac{dP_t}{dX_t'b}\right)^2 \left(X_t'AX_t + bY_tb'\right)\right]^*$$

$$= \frac{1}{T}\left(\frac{dP_t}{dX_t'b}\right)^2 \left(X_t'AX_t + bY_tb'\right) \qquad (5.14)$$

and in the latter case by,

---

*For a relatively homogeneous population in terms of variable values and measurement and parameter errors the variances for individuals are approximately equal.

$$\mathrm{Var}(S_i) = \frac{1}{T^2}\left[T\left(\frac{dP_t}{dX_t'b}\right)^2\left(X_t'AX_t+bY_tb'\right)\right.$$

$$\left. + (T^2-T)\left(\frac{dP_t}{dX_t'b}\right)^2\left(X_t'AX_t + bY_tb'\right)\right]$$

$$= \left(\frac{dP_t}{dX_t'b}\right)^2\left(X_t'AX_t + bYb'\right) \tag{5.15}$$

which is equal to the variance in choice probability for one individual. Where actual situations fall in this range depends on the values of variables for pairs of individuals and the covariance in errors in variable prediction for pairs of individuals. To see this the utility variance and covariance are expressed in algebraic notation,

$$\mathrm{Var}(X_t'b) = \sum_j \sum_k x_{tj}a_{jk}x_{tk} + \sum_j \sum_k b_j\, y_{jkt}b_k \tag{5.16a}$$

$$\mathrm{Cov}(X_t'b, X_{t'}'b) = \sum_j \sum_k x_{tj}a_{jk}x_{t'k} + \sum_j \sum_k b_j y_{jktt'}b_k \tag{5.16b}$$

where x_{tj} ..... is the j^{th} variable for individual t,

a_{jk} ..... is the element in the j^{th} row, i^{th} column of the $_A$th matrix,

y_{jkt} ..... is the j, k element of the Y_t matrix, and

$y_{jktt'}$..... is the j, k element of the $Y_{tt'}$ matrix.

The variance-covariance elements of the parameter estimates, a_{jk}, and the parameters, b_j or b_k, are identical in both equations for all individuals. The expected utility covariance value is zero or near zero if either all of the terms are zero or if the terms tend to offset each other. Since the members of a group for aggregate prediction are generally selected by some criteria which implicitly classifies by common value of variables (i.e. geographic groups) many of these $X_{tj} \cdot X_{t'k}$ products will be similar to the $X_{tj} \cdot X_{tk}$ products in the variance expression.

Since the same process is normally used to generate explanatory variables for all individuals the Y_t matrices are expected to be similar among individuals. The correspondence in errors in variables generated for pairs of individuals may or may not be high depending on the type of variable (i.e. level of service variables obtained from engineering estimates using common networks are highly correlated) and the method of prediction or estimation.

The relative importance of the variance and covariance in probability estimates on the variance in share estimates depends on both the number of individuals in the group and the within group homogeneity as the average covariance is closely related to the commonality of variables for group members. In any case, the variance in share estimate is always lower than the average variance in individual predictions.

To add perspective to the discussion of errors in predictions, it is useful to consider the error in the estimate of choice shares based on sample observation. The variance of the observed share depends on the true share and the number of observations collected.

$$Var(S_i^o) = \frac{1}{T_o} S_i^A (1-S_i^A) \tag{5.17}$$

where S_i^o ..... is the expected observed share choosing alternative i,

T_o ..... is the number of observation to be taken,

S_i^A ..... is the actual share choosing alternative i.

The error in prediction obtained by comparing predicted shares to observed shares includes both errors in prediction and errors in observations.

A similar approach is used to describe the propagation of bias error in parameters or in variables to the prediction of aggregate share. Bias in aggregate share is related to bias in individual choice probabilities by,

$$\text{Bias }(\hat{S}_{iT}) = \frac{1}{T} \sum_t \text{Bias}(P_t) \tag{5.18}$$

Bias in individual choice probability is approximately related to bias in parameters or variables by,

$$\text{Bias}(P_t) = \frac{dP_t}{dX_t'b}\,\text{Bias}(X_t'b)$$

$$= \frac{dP_t}{dX_t'b}\left[\sum_k X_k\text{Bias}(b_k) + \sum_k b_k\text{Bias}(X_k)\right] \qquad (5.19)$$

The bias in share prediction is expressed directly as,

$$\text{Bias}(S_{iT}) = \frac{1}{T}\sum_t \frac{dP_t}{dX_t'b}\left[\sum_k X_k\text{Bias}(b_k) + \sum_k b_k\text{Bias}(X_k)\right] \qquad (5.20)$$

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.

Equation 5.20 can be rewritten to highlight the effect of constant bias across individuals as,

$$\text{Bias}(S_{iT}) = \frac{1}{T}\left[\sum_k \text{Bias}(b_k)\sum_t \frac{dP_t}{dX_t'b}X_k + \sum_k \text{Bias}(X_k)\sum_t \frac{dP_t}{dX_t'b}b_k\right] \qquad (5.21)$$

Common bias in a parameter creates bias in the aggregate share by the factor,

$$\frac{1}{T}\sum_{t}\frac{dP_t}{dX_t'b}\ X_k \tag{5.22}$$

and the common bias in a variable creates bias in the aggregate share by the factor,

$$\frac{1}{T}\sum_{t}\frac{dP_t}{dX_t'b}\ b_k \tag{5.23}$$

which factors are the slope of the aggregate share with respect to a single parameter or variable respectively.

When the members of the group are essentially homogeneous, equation 5.19 simplifies to,

$$\text{Bias}(S_{iT}) = \left[\frac{dP_t}{dX_t'b}\right]\left[\sum_k X_k\text{Bias}(b_k) + \sum_k b_k\text{Bias}(X_k)\right] \tag{5.24}$$

which is identical to the bias in prediction of the individual choice probabilities (Equation 5.19).

There are two effects on error of using different aggregation procedures. First, the propagation of error from variables and parameters is affected by the aggregation procedure. Second, a structural bias is introduced into the share prediction by the aggregation procedure.

When the naive procedure is used the propagation of random and bias error is observable by considering the naive procedure to be a single prediction based on the same model structure but with variables defined by,

$$\overline{X} = \frac{1}{T} \sum_{t} X_t \tag{5.25a}$$

The variance in the estimates of the mean value parameter depend on the degree of correlation of errors in variables between individuals. That is,

$$\begin{aligned} Var(\overline{X}) &= \frac{1}{T} Var(X_t) + \frac{T-1}{T} Cov(X_t, X_{t'}) \\ &= \frac{1}{T} Y_t + \left(\frac{T-1}{T}\right) Y_{tt'}{}^{*} \end{aligned} \tag{5.25b}$$

the variance of share prediction by the naive procedure is:

$$\begin{aligned} Var(S_{iN}) &= \left(\frac{dP_t}{d\overline{X}'b}\right)^2 \left[\overline{X}A\overline{X} + b\, Var(\overline{X}) b'\right] \\ &= \left(\frac{dP_t}{d\overline{X}'b}\right)^2 \left[\overline{X}'A\overline{X} + b \left[\frac{1}{T} Y_t + \frac{T-1}{T} Y_{tt'}\right] b'\right] \end{aligned} \tag{5.26}$$

---

*This formulation assumes that the measurement variance is equal for individuals and that the measurement covariance is equal for pairs of individuals.

The variance in share prediction by the naive procedure (Equation 5.26) is similar to the variance in share prediction by the enumeration procedure (Equation 5.13). The variance of share prediction by the two procedures is identical when the aggregate group is perfectly homogeneous and is slightly greater by the enumeration procedure otherwise. The propagation of bias in parameters and variables by the naive procedures is also similar to the propagation of bias by the complete enumeration procedure.

The structural bias introduced by the naive and other approximate aggregation procedures is described in Chapters III and IV and is not discussed further in this chapter.

5.4 Disaggregation of Errors in Prediction

As described in preceding sections, error in prediction has its source in each of the components of the aggregate prediction structure. Two related analyses of error are relevant in the study of prediction methods. One is to identify the total error of a prediction method based on a specific choice model, representation of socio-economic variables and aggregation procedure. Identification of total error is appropriate to the choice among different methods, to the decision if any of the available methodologies is acceptable, and to obtain a measure of the accuracy of predictions of a particular procedure or alternatively of the risk or expected loss associated with use of the predictions as a basis for decision making.

The identification of the contribution to total error of different elements in the prediction structure is relevant to examination of the available methodologies with the objective of locating those portions of a methodology which should be improved or replaced. This approach is also useful in evaluating the relative importance of errors introduced by component models used in conjunction with other components. In this case the identification of sources of error is used to examine the performance of a range of aggregation procedures in conjunction with a selected set of choice models and distribution representations.

This section describes a strategy for evaluating the performance of aggregation procedures by comparing the total prediction error (using measures to be described in the following section) using a single choice model and a corresponding set of choice variables. This approach has the advantage of highlighting the additional contribution to error of aggregation procedures in the framework of the total process. Due to the complexity of the propagation and interaction of errors (Section 5.3) this approach is based on empirical studies. Since the results of these studies are influenced by the specific data structure used, care must be taken in generalizing the results.

The overall strategy is to identify the error in prediction which results from use of approximate aggregation procedures rather than the theoretically correct complete enumeration procedure. This approach is applied to two different prediction situations. The first is an attempt to replicate the observed travel behavior in which the choice model was calibrated. In this case, the choice model is considered to be correctly specified such that the only errors generated by the choice model are random errors in the estimated parameters. In addition, the data used for "prediction" is identical with that used for estimation. This is equivalent to making "expost" predictions which are suitable to analysis of performance of the prediction model (Klein, 1968). A variety of different aggregation procedures is used in conjunction with the choice model and the given data base. The prediction error resulting from use of the complete enumeration procedure provides a basis for comparison. This prediction is expected to be best due to its consistency with the theory of aggregate travel behavior (Chapter II). The performance of alternative aggregation procedures is evaluated by comparison with both the observed behavior (total error in prediction) and by comparison with the complete enumeration of prediction (aggregation error in prediction).*

The second prediction situation corresponds to a change in the travel service offered. The approach is identical to that for the prediction of observed behavior except that comparisons are made with the complete enumeration prediction only. In this context two predictions are important. They are (1) the prediction of choice shares after the change in travel service and (2) the prediction of change in choice shares due to the change in travel service.

A further set of prediction tests will be made using two additional subsets of data which differ from the calibration data set in one case by random selection only and in the other case by a geographically biased selection procedure. The purpose of these additional predictions is to identify any additional error in the choice model resulting from the attempt to transfer it to a different environment. Figure 5.2, Comparative Prediction Tests, relates the different sets of predictions to the types of error they include. Comparison of the differences between predictions is used to identify the errors which result from,

- Random variation in the choice model,
- Specification error associated with transference of the choice model to a different environment, and
- Bias error of the aggregation procedure.

*For a discussion of disaggregation of total error to errors in prediction and errors in observations in linear models see Cole (1969).

FIGURE 5.2

COMPARATIVE

PREDICTION TESTS

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

The following section describes measures which can be used to evaluate the performance of different methods and their component elements.

5.5 Error Measures for Use in Evaluation of Prediction Methods

The purpose of this section is to identify and describe the characteristics of measures which may be useful in evaluating the performance of different prediction procedures and to select a measure to be used for the evaluation of prediction methods. The selection of individual or sets of error measures should be based on their ability to represent the relative importance of different types and magnitudes and direction of errors in a way which reflects the importance of these errors to the policy decisions which will be based on these predictions.

Decision theory provides a structure for evaluating predictions based on the loss which may be incurred due to prediction errors in well defined decision situations. In such situations the proper weighting of individual errors is determined by analysis of the expected "loss" which will be incurred due to making less than optimal decisions due to errors in prediction. Application of these theories requires the specification of a loss function which distinguishes between different types of decision errors. The most important of these distinctions are (1) over investment rather than under investment and (2) the relative importance of different degrees of over or under investment. However, these methods cannot be used to evaluate prediction methods where the relationship to explicit decisions is not known. For this purpose it is useful to establish a parametric loss function which includes as special cases a wide range of loss functions which might be used in common decision situations. Such a parametric loss function should be able to take account of distinctions of the type described above. It will also be desireable to make distinctions among different subgroups of predictions. For example, in the prediction of mode volumes for the work trip, the errors in predictions of different modes might be considered separately. Alternatively, errors in volume predictions during the morning and afternoon peaks might be treated separately or both peak periods might be grouped together, but separately from off peak predictions.

The range of distinctions which should be included in a parametric loss function are:

1. Differentiation between over prediction and under prediction,
2. Differentiation between errors which do or do not exceed some range of nominal acceptable error,
3. Differentiation by the magnitude of the error,
4. Differentiation by the magnitude of the prediction which is made, and
5. Differentiation over groups of prediction elements which are more or less important in a specific decision context.

Each of these distinctions and the weighting associated with them must be applied to a basic loss measure. It is desireable for this loss measure to be independent of the units of prediction. The basic loss measure on which this parametric loss function is built is defined as,

$$BLM = \frac{P-A}{P} \tag{5.27}$$

where BLM ..... is the basic loss measure for the element of prediction,

P ..... is the predicted value, and

A ..... is the observed value.

This measure is, as desired, scale independent. Using this as a basic unit the parametric loss function is developed by applying parameters and structure which differentiate errors by the five characteristics itemized above.

To distinguish between errors of over prediction and under prediction the basic loss measure is defined for positive, negative and all errors as follows:

$$BLM_P = MAX(BLM,0)$$

$$BLM_N = MAX(-BLM,0)$$

$$BLM_T = \alpha_1 \; BLM_P + \alpha_2 \; BLM_N \tag{5.28}$$

where α_1 and α_2 are the weights applied to over and under predictions, respectively.

To distinguish errors which exceed some level of acceptable error, excess loss measures with the same positive versus negative weighting are defined by:

$$ELM_P = MAX(BLM_P - \beta, 0)$$

$$ELM_N = MAX(BLM_N - \beta, 0)$$

$$ELM_T = \alpha_1 ELM_P + \alpha_2 ELM_N \qquad (5.29)$$

where β ..... is the level below which prediction errors are acceptable.

Differential weighting by magnitude of error is accomplished by modification of the excess loss measure to:

$$ELM_P^{\gamma} \qquad (5.30a)$$

$$ELM_N^{\gamma} \qquad (5.30b)$$

$$ELM_T^{\gamma} \qquad (5.30c)$$

where γ..... indicates the importance of the magnitude of error as follows:

$\gamma=0$..... creates a count of errors greater than the acceptable level,

$\gamma=1$..... indicates the importance of error over the nominal level is proportional to its magnitude,

$\gamma=2$..... indicates the importance of error over the nominal level is proportional to the square of its magnitude.

Differential weighting by magnitude of the prediction is accomplished by introduction of a further modification to:

$$ELM_P^{\gamma} \cdot P^{\delta} \tag{5.31a}$$

$$ELM_N^{\gamma} \cdot P^{\delta} \tag{5.31b}$$

$$ELM_T^{\gamma} \cdot P^{\delta} \tag{5.31c}$$

where δ ..... indicates whether errors will be weighted by prediction magnitude (δ=1) or not (δ=0).

The combined effect of these modifications is included in a general loss measure defined by:

$$GLM(\gamma,\delta) = \sum_{w} (ELM_{Tw})^{\gamma} \cdot P^{\delta} \tag{5.32}$$

where w ..... indicates the prediction points over which the loss measure is determined.

The balance of this discussion is simplified by assuming that the error measure to be considered will not include any allowable level of error (β=0) and that positive and negative errors have equal importance (α_1=α_2=1) so that the general loss measure reduces to:

$$GLM(\gamma,\delta) = \sum_{w} \left(\left|\frac{P_w - A_w}{P_w}\right|\right)^{\gamma} \cdot P_w^{\delta} \tag{5.33}$$

The general loss measure represents the expected decision loss which would result from the set of prediction errors included in its development. As such the general loss measure is consistent with the structure of loss functions in decision analytic contexts. In the present case however, a measure is required which is independent of the number of prediction and which is scale independent of the weighting applied to the individual predictions. That is, the general loss measure must be transformed into an average loss measure for a typical element of prediction as,

$$ALM(\gamma,\delta) = \left[GLM(\gamma,\delta)/CF(\delta)\right]^{1/\gamma} \tag{5.34a}$$

where $$CF(\delta) = \sum_{w} P_w^{\delta} \tag{5.34b}$$

The resultant average loss measure may be disaggregated in terms of average errors and random errors. This is done by determining the average error:

$$AE = \sum_{w \in W} \left(\frac{P_w - A_w}{P_w}\right) P_w^{\delta} \Big/ \sum_{w \in W} P_w^{\delta} \tag{5.35}$$

Variability of prediction may then be measured in terms of the deviation of the error from the average error by,

$$BLM^{*} = BLM - AE \tag{5.36}$$

where BLM^{*} ..... is the adjusted basic loss measure, and

BLM ..... is the unadjusted basic loss measure defined by Equation 5.27.

When the importance of an error is proportional to the square of its magnitude the average loss measure may be defined in terms of the adjusted average loss measure:

$$ALM(\gamma, \delta)^{*} = \left[GLM(\gamma, \delta)^{*} / CF(\delta) \right]^{1/\gamma} \tag{5.37a}$$

$$GLM(\gamma, \delta)^{*} = \sum_{w} \left\{ \left| \frac{P_w - A_w - AE}{P_w} \right|^{2} \right\} \cdot P_w^{\delta} \tag{5.37b}$$

and the average loss by:

$$\left[ALM(\gamma, \delta) \right]^{2} = \left[(ALM(\gamma, \delta)^{*} \right]^{2} + AE^{2} \tag{5.38}$$

This is consistent with the disaggregation of random and average errors described by Theil (1966).

Selection of different values for the parameters described above (γ, δ) leads to a variety of loss measures including those most commonly used in decision analysis.

5.6 Summary

The purpose of this chapter 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 error generated in each of thes elements are described.

The process by which errors enter the process, interact with one another and are propagated to the aggregate prediction are analyzed. The important characteristics of the propagation process are identified.

An experimental design is suggested for use in the evaluation of the performance of component elements of a model structure. A strategy is outlined for the evaluation of aggregation procedures in the context of an applied prediction situation.

A parametric set of error measures is designed which may be used to evaluate the errors generated in an applied prediction situation.

CHAPTER VI

APPLIED PREDICTION ANALYSIS

6.1 Introduction

An aggregated prediction model structure was described in Chapter II and proposed for use in prediction of aggregate travel behavior. The aggregated prediction model includes three components: a disaggregate choice model, a representation of the distribution of choice variables and an aggregation procedure. Chapters III and IV explored and described the bias in prediction which is introduced by different aggregation procedures. These studies indicated the sensitivity of aggregation bias to the mean, variance and shape of the distribution of the net utility between pairs of alternatives in the aggregate group. They also indicated the differences in bias among the methods, the general ranking of methods in terms of their bias and the conditions under which the rank ordering might be inverted. These analyses also provided an indication of the magnitude of bias for different levels of variance in the distribution of net utility values for the binary choice case.

Sources of errors in the choice model and the values of independent variables were identified in Chapter V. The interaction of these errors and their propagation to aggregate predictions was also described.

The purpose of the application analysis described in this chapter is to explore the combined effect of errors from these different sources in a realistic prediction situation. The analysis identifies magnitudes of aggregation error for the prediction situation and places them in perspective with respect to other errors. The analysis also describes the relative performance of three aggregation procedures which have been proposed for use in prediction situations.

The prediction situation considered is the mode choice to work by breadwinners with a CBD workplace in the Washington Metropolitan area. Aggregate predictions are made at the level of the residential district. The chapter is organized in sections describing the data, the specification and estimation of the disaggregate choice model, a quantitative analysis of errors in parameters and independent variables, a description of the prediction groupings and a description of grouped data, a description of the approach to the analysis of aggregate prediction errors, an analysis of prediction errors in the base case and for policy alternatives and a summary of the results obtained.

6.2 Description of Data

The data used in this study are from the metropolitan Washington D.C. area. Socio-economic and trip making data come from the Washington Council of Governments 1968 Home Interview Survey. This survey was based on a random sample of households.

The sampling rate was 3% inside the capital beltway and 5% outside the beltway. In total, approximately 25,000 households were interviewed. The survey information included the socio-economic characteristics of the households, the characteristics of the individual in each household, and information on the trips made by each individual in the household. The socio-economic information included data on household composition, income, car availability, and the age, sex, race, education and occupation of members of the household, and a variety of other data items. The trip information includes a complete listing of every trip made by each person in the houshold during a 24 hour period. Trip information includes the purpose of the trip, mode of travel, origin and destination, time of day, etc. but does not include the route of travel.

Level of service data were obtained from an inventory of existing transportation facilities and services performed by the Washington Council of Governments and additional data assembled by R.H. Pratt Associates. These data were merged into a combined file, which related the level of service data to individual trips for home based work trips, by Cambridge Systematics Incorporated, as a part of a study of automobile ownership and mode to work choice [Cambridge Systematics, 1975]. The merged file included 18,455 households with one or more work trips. Although the data are reported on a disaggregated basis, that is, the socio-economic, travel choice and level of service data are reported for individual workers, the level of service data were actually developed from coded networks based on a traffic zone system and therefore represent some degree of aggregation.

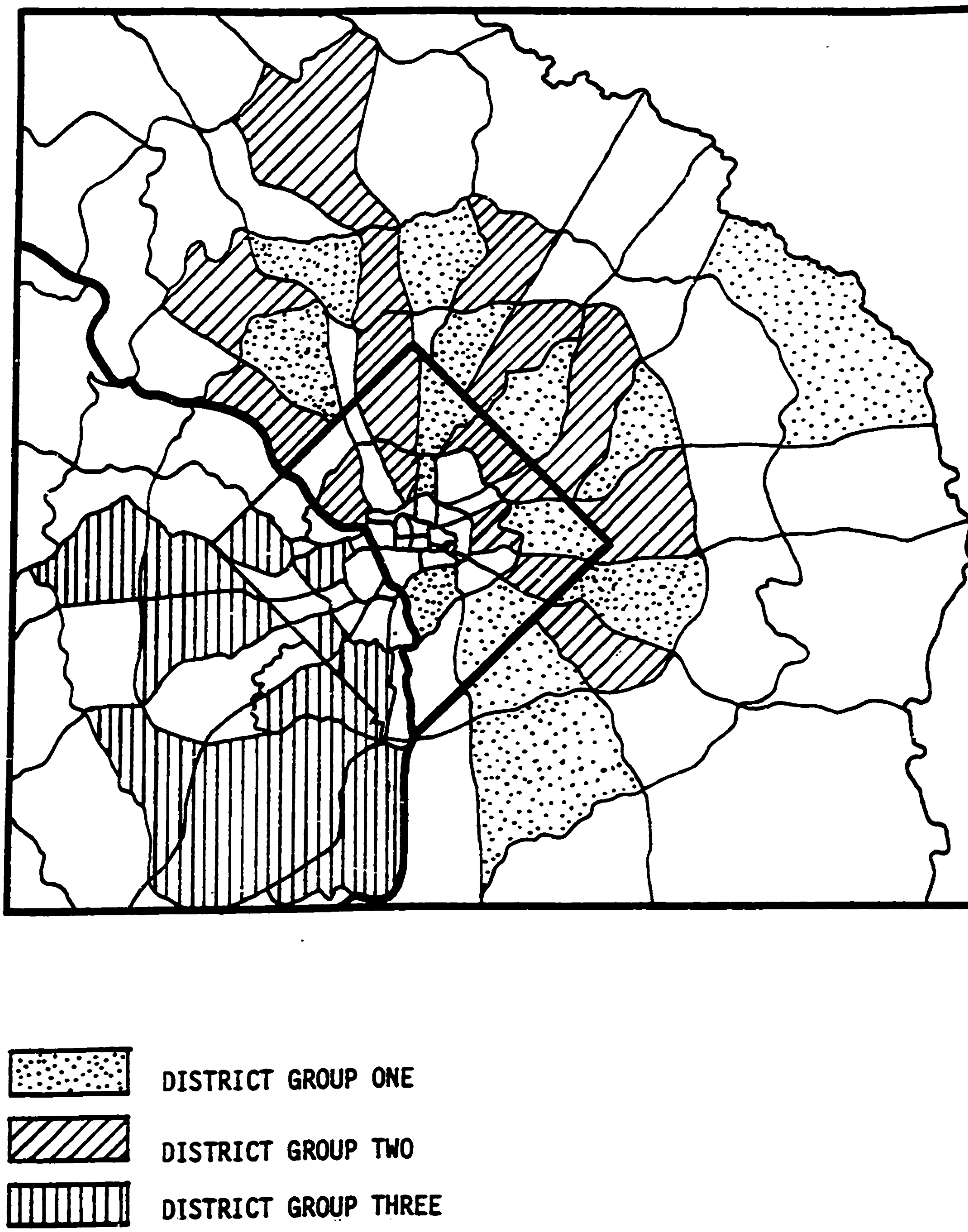
The data set is reduced to 2,965 records in this study. These records include households residing outside of the central business district and within the capital beltway in which the breadwinner made a home based work trip to the central business district by the drive alone, shared ride, or transit ride travel modes. To allow for later aggregate prediction and analysis, the data set was further reduced to 2,132 trips made by households resident in 45 districts (Figure 6.1, Selected Districts in Washington Metropolitan Area) which had 30 or more breadwinner trips to the CBD. This file was divided into three groups as follows:

1) 12 districts in Virginia, including 486 households,
2) 16 districts representing every second district in Washington and Maryland, including 772 households, and
3) 17 remaining districts in Washington and Maryland, including 874 households.

The purpose of this division of data is to allow estimation of the work mode choice model on one of the subsets of data and use of the estimated

FIGURE 6.1

DISTRICTS IN WASHINGTON METROPOLITAN AREA



model for prediction with all the subsets in order to consider the effect of transfering the estimated model between data sets within the metropolitan area.

6.3 Choice Model Specification and Estimation

This section describes the specification and calibration results of a disaggregate choice model developed for use in the evaluation of alternative aggregation procedures. The disaggregate choice model developed describes one part of the overall travel process, namely, the home to work mode choice.* To improve the potential to generalize the results of this analysis it was decided not to use a binary choice model. Instead a three mode model consisting of drive alone, shared ride (car pool) and transit alternatives was developed for work trips made by breadwinners with work place in the central business district.

It is assumed, in the study, that the mobility choices of job site, residential location and auto ownership are fixed. The only short-term choice left open for the work trip is the choice of travel mode. Therefore, given a new external change of travel environment, the model will predict only the relatively short run travel behavior of mode choice.

The behavioral unit is a worker (breadwinner). The model predicts the worker's choice of travel mode for his home to work trip. The following factors which influence the choice of mode to work are included in the model:

- Level of service by alternative travel modes from home to work and return,

*The model developed for this study is a modification of a work mode choice model developed by Cambridge Systematics (1975).

- Socio-economic characteristics of the worker and the household of which he is a member, and
- Special characteristics which describe the existence of special incentives for use of the shared ride mode.

The disaggregate choice function used in this study is a multinomial logit model. Its theoretical development and implications are described in Chapter II, Section 6. The form of the model is:

$$P_t(m:M_t) = \frac{e^{V_{mt}}}{\sum_{m'} e^{V_{m't}}} \tag{6.1}$$

where $P_t(m:M_t)$ ..... is the probability that individual t chooses mode m of the set of modes available to him, M_t, and

V_{mt} ..... is the utility of mode m to individual t.

The following modal definitions were used:

<u>Drive Alone</u> includes all reported car drivers less any who fall into the shared ride mode as defined below.

<u>Shared Ride</u> (Driver or passenger in a car with 2 or more persons) includes all auto passengers plus any auto drivers who reported they carried a passenger with them, or who reported they were in a car pool, or who reported they did not know whether or not they were in a car

pool.*

Transit includes all reported public transit (bus) trips.

The drive alone mode is available to breadwinners who have drivers license and live in households with one or more cars available. The shared ride mode is available to all tripmakers. The transit mode is available to residents of all districts where excess time for the trip to the CBD is less than thirty minutes. All breadwinners in this data set have the transit mode available to them.

The following variables are used in the model:**

1. Constant for Drive Alone (D_d)

$$= \begin{cases} 1 \text{ for drive alone} \\ 0 \text{ for other modes} \end{cases}$$

2. Constant for Shared Ride (D_s)

$$= \begin{cases} 1 \text{ for shared ride} \\ 0 \text{ for other modes} \end{cases}$$

3. Autos per Licensed Driver for Drive Alone ($AALC_d$)

$$= \begin{cases} \text{autos} \div \text{drivers for drive alone} \\ 0 \text{ for other modes} \end{cases}$$

---

*The last condition captures the people who were confused about the definition of a car pool. It is reasonable to assume that there could be confusion only if a passenger was carried but the respondent did not know whether or not to consider it a car pool. For example, intrafamily car pooling.

**The subscripts d and s denote variables which appear only in the utility equations for the Drive Alone and Shared Ride modes, respectively.

4. Auto per Licensed Driver for Shared Ride ($AALC_s$)

$$= \begin{cases} \text{autos} \div \text{drivers for shared ride} \\ \text{0 for other modes} \end{cases}$$

5. Out of Pocket Cost Divided by Income (OPTC/INC)

= round trip travel cost divided by household income in cents per thousand of dollars. The value for Shared Ride is the value for Drive Alone divided by 2.5.*

6. Total Travel Time (TTT) (in minutes)

= the sum of in vehicle travel time (IVTT) and out of vehicle travel time (OVTT). The value for Shared Ride is the value for Drive Alone plus a penalty of ten minutes for pick-up and drop-off based on data reported by Attanucci (1974).

7. Out of Vehicle Travel Time Divided by Distance (OVTT/DIST)

= out of vehicle round trip travel time divided by one way distance in minutes per mile.

8. Government Worker Variable for Shared Ride (GW_s)

$$= \begin{cases} \text{1 if the worker is a civilian employee of the federal} \\ \text{government for shared ride} \\ \text{0 if not so.} \end{cases}$$

9. Number of Workers in Household for Shared Ride ($NWORK_s$)

$$= \begin{cases} \text{number of workers for shared ride} \\ \text{0 for othermodes} \end{cases}$$

---

*The value 2.5 was assumed to be the average auto occupancy for Shared Riders. This value was obtained from data reported in a car pooling study by Attanucci (1974). It is used to reflect cost sharing among the riders.

The level of service variables describe the performance of the three modes. These are generic variables which are included with mode specific values for each of the mode alternatives.

Auto availability is included as two separate variables under the hypothesis that it has a differential effect on the probability of choosing each of the three modes. The government worker variable is a proxy for working in an organization which offers car pooling incentives. The number of workers is included to capture the effects of intra-family car pools.

The formulation of travel time variables assures that out-of-vehicle travel time will be represented as more onerous than in-vehicle travel time. The contribution of the travel time variables to the utility function (any mode) is expressed as:

$$\begin{aligned} V_m(\text{TTT},\text{OVTT}) &= b_6 * \text{TTT} + b_7 * \text{OVTT}/\text{DIST} \\ &= b_6 * (\text{IVTT} + \text{OVTT}) + b_7 * \text{OVTT}/\text{DIST} \\ &= b_6 * \text{IVTT} + \left(b_6 + \frac{b_7}{\text{DIST}}\right) * \text{OVTT} \qquad (6.2) \end{aligned}$$

Since the coefficients, b_6 and b_7 should both have negative signs, the importance of one minute of out-of-vehicle time will be greater than that of one minute of in-vehicle time.

The specification of the model is summarized in Table 6.1 and the estimation results are shown in Table 6.2. All the coefficients are significant and have the expected sign. The relationship between the

TABLE 6.1

MODEL SPECIFICATION

Variable Name	Coefficient	Variable Value in Modal Utility Function: Drive Alone	Share Ride	Transit
Drive Alone Dummy Variable	B_1	$D_d = 1$	0	0
Shared Ride Dummy Variable	B_2	0	$D_s = 1$	0
Autos ÷ Drivers, Drive Alone	B_3	$AALC_d$	0	0
Autos ÷ Drivers, Shared Ride	B_4	0	$AALC_s$	0
Out of Pocket Cost ÷ Income	B_5	OPTC/INC	OPTC/INC	OPTC/INC
Total Travel Time	B_6	TTT	TTT	TTT
Out of Vehicle Time ÷ Distance	B_7	OVTT/DIST	OVTT/DIST	OVTT/DIST
Government Worker, Shared Ride	B_8	0	GW_s	0
Number of Workers, Shared Ride	B_9	0	$NWORK_s$	0

TABLE 6.2

MODEL ESTIMATION RESULTS

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. Auto availability (Drive Alone)	$AALD_d$	3.64	0.38
4. Auto availability (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	$NWORK_s$	0.24	0.10

Summary of Statistics*

Number of Observations	874	L*(o)=-857.6
Number of Cases	1,495	L*(b)=-709.2

*The statistics are defined as follows:
1. Number of observations - the number of work trips in the sample.
2. Number of cases - the number of alternatives in excess of one per observation summed over all observations.
3. L*(o) - the value of the log of the likelihood function when all coefficients are zero.
4. L*(b) - the value of the log of the likelihood function when the coefficients take the estimated value.

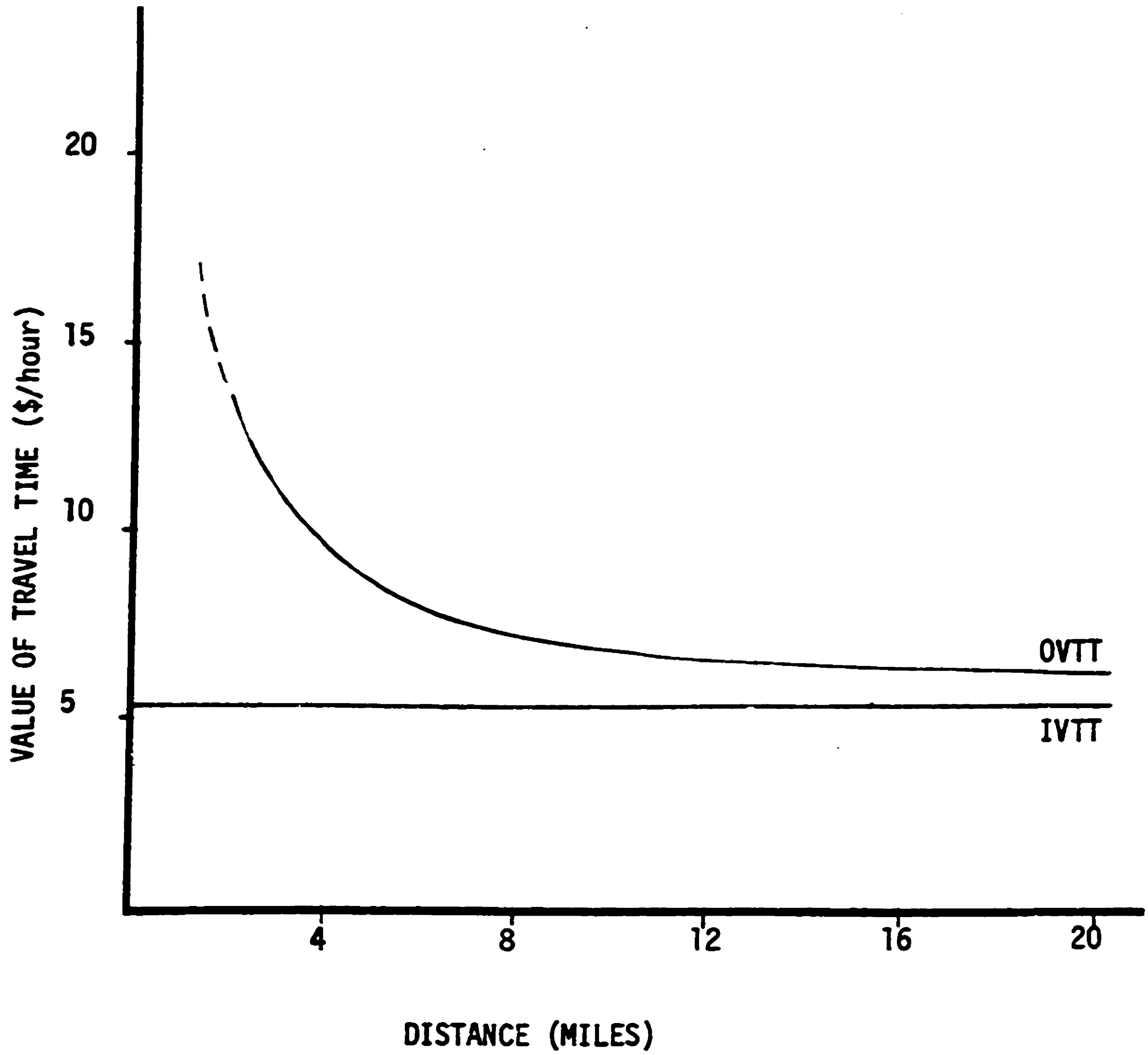
coefficients of the level of service variables is expressed in terms of the value of time as follows:

1) In-vehicle travel time (in dollars per hour) is valued at 0.51 times household income expressed in thousands of dollars. This is equivalent to approximately 100% of the hourly wage rate.
2) Out-of-vehicle travel time value exceeds the value of in-vehicle travel time by 1.65 times household income in thousands of dollars divided by the number of miles of one way travel between residence and workplace. This implies that the discomfort and/or inconvenience associated with out-of-vehicle time is discounted for longer trips.

The values of in- and out-of-vehicle travel time are plotted as a function of distance for breadwinners from households with $10,000 annual income (1968) in Figure 6.2, Value of Travel Time. The values of travel time for households with higher or lower income levels may be scaled by the factor (Household Income · $10,000).

FIGURE 6.2

VALUE OF TRAVEL TIME
(INCOME = $10,000)



VALUE (IVTT) = 0.51* (INCOME IN THOUSANDS)

VALUE (OVTT) = $\left[0.51 + \frac{1.65}{\text{DISTANCE}}\right]$ * (INCOME IN THOUSANDS)

6.4 Quantitative Analysis of Errors in Parameters and Independent Variables

Three sources of prediction error discussed in Chapter V were (1) error in the parameters of the choice model, (2) errors in choice influencing variables, and (3) aggregation error. The effect of errors in parameters and errors in variables on the variance of individual probability estimates is examined by extension of Equations 5.6, 5.8 and 5.9 to the three mode choice situation as follows:

$$\mathrm{Var}(P_{mt}) = \sum_{m'} \sum_{m''} \left(\frac{\partial P_{mt}}{\partial V_{m't}}\right)\left(\frac{\partial P_{mt}}{\partial V_{m''t}}\right) \mathrm{Cov}(V_{m'}, V_{m''}) \tag{6.3a}$$

$$\mathrm{Cov}(V_{m'}, V_{m''}) = X'_{m'} A X_{m''} + b\ Y_{m'm''t}\ b' \tag{6.3b}$$

where P_{mt} ..... is the choice probability prediction for individual t choosing mode m,

$\frac{\partial P_{mt}}{\partial V_{m't}}$ ..... is the first derivative of the choice of mode m with respect to the utility value for mode m' which may or may not be equal to m,

$\mathrm{Cov}(V_{m'}, V_{m''})$..... is the covariance of the utility estimates for modes m' and m". When m'=m", this becomes the variance of $V_{m'}$. Either or both m' and m" may equal m,

$X_{m'}$..... is the vector of variable in the utility function for alternative m',

A ..... is the variance-covariance matrix of parameter estimates,

b ..... is the vector of parameter estimates, and

$Y_{m'm''t}$ ..... is the variance-covariance matrix of measurement errors for variables in the utility functions for alternatives m' and m" for individual t. m' and m" may or may not be the same and either or both may be the same as m.

First, assuming only errors in parameters, Equation 6.3b is reduced to:

$$Cov(V_{m'}, V_{m''}) = X_{m'}^{'} A X_{m''} \tag{6.4}$$

The estimated values of the variances and covariances of the model parameters are given in Table 6.3, Variance-Covariance Matrix of Parameter Estimates. Representative values of the variables(Table 6.4, Representative Values of Choice Variables) are used to estimate the variance and standard error of estimate of the mode choice probabilities.

The estimates of individual choice probabilities and errors due to errors in parameters are given in Table 6.5, Sample Choice Probabilities and Error Due to Errors in Parameters. The average value of the standard error due to errors in parameters is equal to 19% of the average choice probability.

To obtain an estimate of the corresponding error in probability estimate due to errors in variables an assumption must be made as to the actual error in variables. For the purpose of providing an example the assumption is made that measurement error in the level of

TABLE 6.3

VARIANCE COVARIANCE MATRIX OF PARAMETER ESTIMATES

D_d	D_s	$AALD_d$	$AALD_s$	$\frac{OPTC}{INC}$	TTT	$\frac{OVTT}{DIST}$	GW_s	$NWORK_s$
0.1318	0.0245	-0.1036	-0.0202	-0.0018	0.0007	0.0009	0.0013	0.0036
	0.0706	-0.0102	-0.0294	-0.0002	0.0004	0.0016	-0.0177	-0.0171
		0.1441	0.0325	-0.0004	-0.0000	-0.0007	-0.0001	-0.0037
			0.0560	0.0003	0.0000	-0.0007	0.0052	-0.0020
				0.0002	-0.0000	-0.0000	-0.0001	-0.0001
					0.0000	-0.0001	0.0000	-0.0000
						0.0031	-0.0005	0.0001
							0.0262	0.0006
								0.0107

Note: VALUE OF ZERO INDICATES ESTIMATE IS LESS THAN 0.00005 IN MAGNITUDE.

TABLE 6.4

REPRESENTATIVE VALUES OF CHOICE VARIABLES

	Variables*								
Mode	D_d	D_s	$AALC_d$	$AALC_s$	$\frac{OPC}{INC}$	TTT	OVTTD	GW	NWORK
Drive Alone	1.0	0	0.6	0	20.	60.	2.5	0	0
Shared Ride	0	1.0	0	0.6	8.	70.	2.5	0.4	1.5
Transit Ride	0	0	0	0	8.	100.	4.5	0	0

*Variables are defined in the text and in Table 6.1.

TABLE 6.5

SAMPLE CHOICE PROBABILITIES AND
ERROR DUE TO ERROR IN PARAMETERS

Mode	Estimated Probability	Variance	Std. Error	Std. Error ÷ Probability
Drive Alone	0.40	0.0024	0.049	0.12
Shared Ride	0.31	0.0015	0.038	0.12
Transit Ride	0.28	0.0077	0.088	0.31
AVERAGE	0.33	0.0039	0.062	0.19

service variables is 10% of the variable value. No correlation in measurement error for different level of service characteristics is assumed. However, the measurement error for each drive alone level of service variable is assumed to be perfectly correlated with the corresponding shared ride level of service variable.

The resulting probability estimates and error measures are given in Table 6.6, Sample Choice Probabilities and Error Due to Errors in Variables. The average standard error in choice probability estimate due to errors in variables is 11% of the average estimate.

The combined effect of errors in parameters and errors in variables is shown in Table 6.7, Sample Choice Probability and Combined Error Due to Error in Parameters and Error in Variables. The average standard error in choice probability due to both sources of error is approximately 21% of the average choice probability. The contribution to the standard error of choice probabilities, in this example, is greater for errors in parameters than from errors in variables. The true relative importance of these two sources of error depends on the actual magnitude of error in the prediction of independent variables.

The propagation of individual probability error to errors in aggregate share prediction is dependent on the covariance between the share prediction of pairs of individuals in the group and the number of individuals in the group. To see this Equation 5.5 is repeated here:

TABLE 6.6

SAMPLE CHOICE PROBABILITIES AND
ERROR DUE TO ERROR IN VARIABLES

Mode	Expected Probability	Variance of Probability	Std. Error of Probability	Std. Error ÷Probability
Drive Alone	0.40	0.0007	0.028	0.07
Shared Ride	0.31	0.0005	0.022	0.07
Transit Ride	0.28	0.0025	0.050	0.17
AVERAGE	0.33	0.0012	0.035	0.11

TABLE 6.7

SAMPLE CHOICE PROBABILITY AND COMBINED
ERROR DUE TO ERROR IN PARAMETERS AND ERROR IN VARIABLES

Mode	Expected Probability	Variance of Probability Due to			Std. Error of Probability	Std. Error ÷ Probability
		Errors in Parameters	Errors in Variable	Joint Errors		
Drive Alone	0.40	0.0024	0.0007	0.0032	0.057	0.14
Shared Ride	0.31	0.0015	0.0005	0.0019	0.044	0.14
Transit Ride	0.28	0.0077	0.0025	0.0102	0.101	0.36
Average	0.33	0.0039	0.0012	0.0051	0.071	0.21

$$\mathrm{Var}(S_m) = \frac{1}{T^2}\left[\sum_t \mathrm{Var}(P_{mt}) + \sum_t \sum_{t' \neq t} \mathrm{Cov}(P_{mt}, P_{mt'})\right] \quad (6.5)$$

where T ..... is the number of observations in the prediction sample.

For simplicity, the variance of share predictions is assumed constant across individuals and the covariance is assumed constant across pairs of individuals so that:

$$\mathrm{Var}(S_m) = \frac{1}{T^2}\left[T \cdot \mathrm{Var}(P_{mt}) + (T^2-T)\ \mathrm{Cov}(P_{mt}, P_{mt'})\right]$$

$$= \frac{1}{T}\ \mathrm{Var}(P_{mt}) + \frac{T-1}{T}\ \mathrm{Cov}(P_{mt}, P_{mt'}) \quad (6.6)$$

For large values of T, the share variance is approximated by:

$$\mathrm{Var}(S_m) = \rho\ \mathrm{Var}(P_{mt}) \quad (6.7)$$

That is, the variance in estimated shares declines with increasing prediction observations, T, to a minimum value which is determined by the individual probability variance and the correlation in probability variance for pairs of individuals. If the correlation in probability variance is high, the error in prediction of aggregate shares due to errors in the choice model may be large independent of other sources of error in prediction.

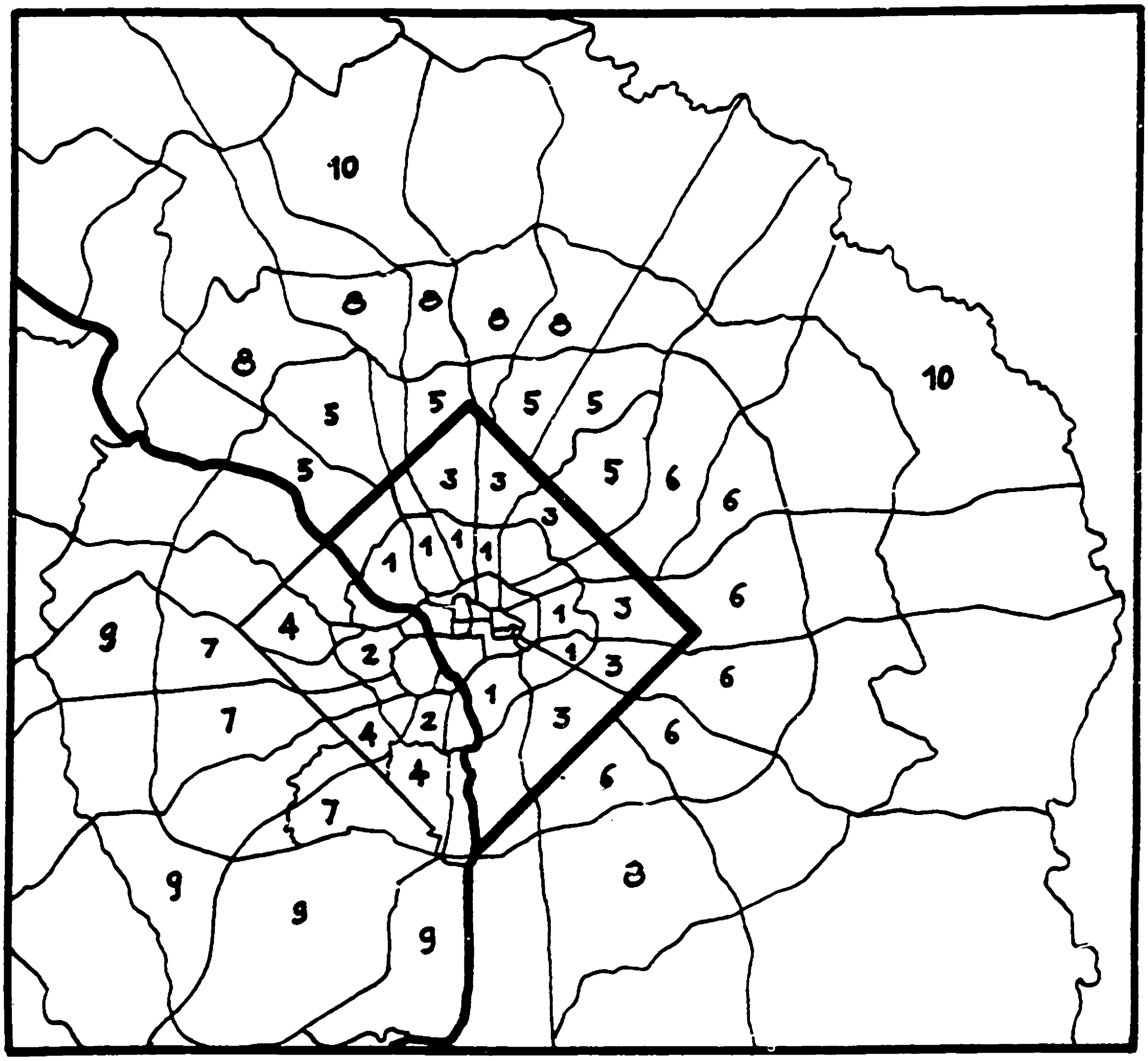
6.5 Prediction Groupings and Description of Data

For the purpose of prediction analysis the sample of breadwinners making work trips to the CBD is aggregated to the residence district of the trip maker. These districts are grouped in three sets described in Section 6.2 and illustrated in the map, Figure 6.1. Prediction analyses are made for each of these sets of districts, for all the districts combined in a single set, and for three additional levels of aggregation. The second level of aggregation is to ten super districts, each of which include two to seven districts. The super districts are shown in Figure 6.3, Super Districts in the Washington Metropolitan Area. The third level of aggregation consists of four groupings of individuals in rings around the CBD, as shown in Figure 6.4, Rings in the Washington Metropolitan Area. The fourth level of aggregation combines all of the individual trip makers into a single group.

Because of the geographically based definition of districts, super districts, and rings, these groups have different numbers of observations in the prediction units as well as differences in the characteristics of the observations. Table 6.8, Basic Data Set Characteristics, describes the range in the number of observations for districts (in three sets and overall), super districts, and rings and the range in the proportion of persons in the prediction groups who do not have the drive-alone alternative available to them (all individuals in the data set have both the shared ride and transit ride

FIGURE 6.3

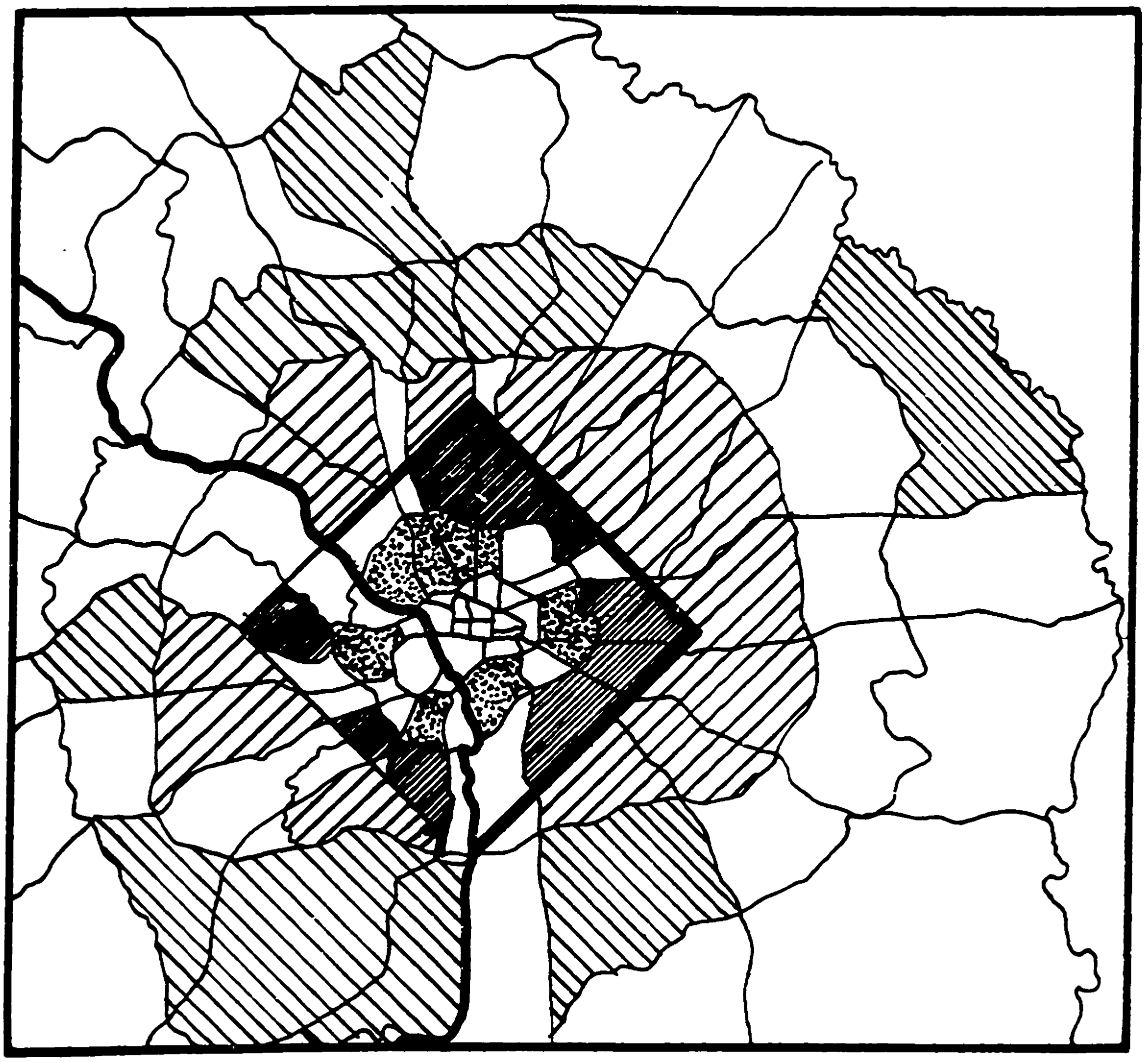
SUPERDISTRICTS IN WASHINGTON METROPOLITAN AREA



SUPERDISTRICTS NUMBERED 1 - 10

FIGURE 6.4

RINGS IN WASHINGTON METROPOLITAN AREA



RING ONE

RING TWO

RING THREE

RING FOUR

TABLE 6.8

DATA SET CHARACTERISTICS

AGGREGATE GROUP	GROUP SIZE			PROPORTION WITHOUT DRIVE ALONE ALTERNATIVE		
	Min.	Max.	Average	Min.	Max.	Average
DISTRICTS: SET ONE (16)	30	129	51	0	71%	29%
DISTRICTS: SET TWO (17)	28	93	48	0	59%	24%
DISTRICTS: SET THREE (12)	27	60	41	0	32%	8%
DISTRICTS: ALL (45)	27	129	47	0	71%	23%
SUPER DISTRICT (10)	84	400	213	2%	47%	23%
RING (4)	470	510	533	3%	41%	23%
REGION (1)	2132			23%		

alternative available). The importance of these two pieces of information will be apparent in the discussion in the following sections.

Table 6.9, Range of Observed Mode Shares, indicates the range in shares for each of the alternatives for the three sets of districts and four different levels of aggregation. The range in shares is not significantly different between district sets and does not narrow appreciably with higher levels of aggregation. The overall share choosing each alternative in the entire sample population is shown in the last line of the table.

The final characteristic of interest is the distribution of utility values for each of the prediction groups. The important characteristics of the distribution of utilities are fully described in terms of the distribution of the utility differences for each of the three pairs of choices: drive alone versus shared ride, drive alone versus transit ride, and shared ride versus transit ride. Table 6.10, Range of Distribution Characteristics for Net Utility Values, presents the range of mean, variance, skewness and kurtosis for each of the sets of prediction groups. The general effect of grouping to higher levels of aggregation is to reduce the range of values for each of these measures. The clustering of skewness and kurtosis values around zero and 3.0, respectively, indicates that the distributions become nearly normal as the number of observations increases. The higher levels of aggregation also eliminate extremely high and extremely low values for variance in the net distributions.

TABLE 6.9

RANGE OF OBSERVED MODE SHARES

AGGREGATE GROUP		DRIVE ALONE			SHARED RIDE			TRANSIT RIDE		
		Min.	Max.	Ave.	Min	Max.	Ave.	Min.	Max.	Ave.
DISTRICTS	SET ONE	.16	.61	.35	.10	.59	.28	.09	.73	.36
	SET TWO	.21	.75	.41	.13	.48	.28	.00	.65	.31
	SET THREE	.13	.67	.40	.17	.52	.34	.03	.51	.26
	ALL	.13	.75	.38	.10	.59	.30	.00	.73	.32
SUPER DISTRICTS		.24	.59	.38	.13	.46	.30	.05	.62	.32
RINGS		.25	.55	.38	.16	.36	.30	.10	.59	.32
REGION			.38			.30			.32	

TABLE 6.10

RANGE OF DISTRIBUTION CHARACTERISTICS FOR
NET UTILITY VALUES

A. Distribution of Drive Alone - Shared Ride Utility

Aggregation Group		Mean		Variance		Skewness		Kurtosis	
		Min	Max	Min	Max	Min	Max	Min	Max
DISTRICTS	Set One	-0.58	0.86	0.44	1.53	-0.83	0.79	1.63	4.72
	Set Two	-0.20	1.05	0.44	1.78	-0.73	0.92	1.80	4.79
	Set Three	0.09	0.75	0.38	1.00	-0.69	0.40	1.72	3.91
	ALL	-0.58	1.05	0.38	1.78	-0.83	0.92	1.63	4.79
Super Districts		-0.07	0.71	0.53	1.31	-0.62	0.37	2.22	3.86
Rings		0.04	0.61	0.60	1.26	-0.19	0.23	2.33	2.82
Region		0.33		1.07		-0.20		2.89	

TABLE 6.10

RANGE OF DISTRIBUTION CHARACTERISTICS FOR NET UTILITY VALUES

B. Distribution of Drive Alone - Transit Ride Utility

Aggregation Group		Mean		Variance		Skewness		Kurtosis	
		Min	Max	Min	Max	Min	Max	Min	Max
DISTRICTS	Set One	-1.90	1.87	0.82	3.36	-1.28	1.05	1.54	5.99
	Set Two	-1.24	2.01	0.53	4.07	-1.09	1.21	1.45	6.52
	Set Three	-0.24	1.89	1.01	2.76	-0.81	1.24	2.07	6.31
	ALL	-1.90	2.01	0.53	4.07	-1.28	1.24	1.45	6.52
Super Districts		-1.00	1.49	1.18	3.26	-0.73	0.42	2.00	4.89
Rings		-0.75	1.34	1.41	3.30	-0.24	0.22	2.19	4.02
Region		0.38		3.11		-0.32		3.06	

TABLE 6.10

RANGE OF DISTRIBUTION CHARACTERISTICS FOR
NET UTILITY VALUES

C. Distribution of Shared Ride - Transit Ride Utility

Aggregation Group		Mean		Variance		Skewness		Kurtosis	
		Min	Mas	Min	Max	Min	Max	Min	Max
DISTRICTS	Set One	-1.32	1.34	0.28	1.04	-0.82	0.92	1.86	4.08
	Set Two	-1.10	1.05	0.27	1.05	-1.00	0.80	2.14	4.41
	Set Three	-0.33	1.41	0.37	1.07	-0.49	0.67	2.28	4.46
	ALL	-1.32	1.41	0.27	1.07	-1.00	0.92	1.86	4.46
Super Districts		-0.93	0.87	0.41	1.04	-0.26	0.25	2.44	3.32
Rings		-0.78	0.73	0.66	0.86	-0.15	0.16	2.51	3.28
Region		0.05		1.06		-0.10		2.94	

6.6 Approach to the Analysis of Aggregate Prediction Errors

Four aggregated prediction models are included in the empirical analysis of prediction errors. The models are differentiated by the use of different aggregation procedures only. Each of the models are based on the disaggregate choice model described in Section 6.3. The variables used in each of the models are computed from the detailed data set. The aggregation procedures used are the enumeration, naive, statistical differentials, and classification procedures.*

The enumeration procedure consists of predicting the expected choice probability for each individual in an aggregate group and averaging across individuals to obtain the expected share for the group. Individuals who do not have the drive alone alternative available to them have an assigned probability of zero of choosing the drive alone alternative. This procedure introduces no aggregation error into the predictions obtained and is therefore used as a basis for the identification of the aggregation error of the other aggregation procedures.

The naive procedure consists of using average values for the independent variables for each prediction group in the disaggregate

---

*Procedures of integration are excluded as their expected computational requirements for extensions to the large choice sets which will be included in multidimensional choice situations are expected to make them infeasible for practical use. Further research on this point is suggested.

choice model and using the computed probabilities as the expected shares for the prediction group. An adjustment is made to the initial share estimates to take account of the proportion of individuals who do not have the drive alone alternative available to them. The formulation for this adjustment is:

$$S_{DA} = S_{DA}^* : \; S_A \tag{6.18a}$$

$$S_{SR} = S_{SR}^* \; \frac{1-S_{DA}}{1-S_{DA}^*} \tag{6.18b}$$

$$S_{TR} = S_{TR}^* \; \frac{1-S_{DA}}{1-S_{DA}^*} \tag{6.18c}$$

where S_{DA}, S_{DA}^* ..... are the adjusted and initial predicted share for the drive alone alternative,

S_{SR}, S_{SR}^* ..... as above for shared ride,

S_{TR}, S_{TR}^* ..... as above for transit ride, and

S_A ..... is the proportion of the population which has the drive alone alternative available to it.

<u>The statistical differentials procedure</u> modifies the prediction obtained by the naive procedure to take account of the variance and covariance in the distribution of utility values in tne aggregate group as described in Chapter II and III. The initial share estimates are adjusted, as in the naive case, to take account of the proportion of

individuals who do not have the drive alone alternative available to them.

The classification procedure consists of first classifying the population into two groups according to whether they have the drive alone alternative available to them, and second, classifying that portion of the population which has the drive alone alternative available into groups which have or do not have competition for use of the automobile. That is, individuals from households where there is a smaller number of automobiles than licensed drivers are segregated from those in households where there is essentially one automobile available for each licensed driver. The selection of automobile availability as the variable for use in classification is based on the fact that variance in this variable contributes the greatest variance to each of the pairwise net utility distributions. The other variables in the choice model are included at their average values for the entire prediction group.

The analysis of error in prediction is first concerned with predicting the travel choice behavior for the situation which prevailed during the period when the data was collected. This is referred to as the base case. The analysis of base case prediction error is undertaken in two parts. First, the aggregation error in prediction is isolated by comparing the predictions of the naive, statistical differentials and classification procedures to those of the enumeration procedure. This is done for the three groupings of districts described in the preceding section, for all of the districts combined and for super districts, rings, and the region as a whole. Next, the total error is considered by comparing the predictions by each of these three procedures and the enumeration method to the observed shares. The observed differences include the effect of errors in the choice model and variables, errors in the observed shares and, if they exist, errors due to transferability (application of the choice model to a situation or data set different from that in which it is estimated) in addition to aggregation errors.

The performance of the different prediction procedures in predicting choice behavior under three different policy changes is also considered. These policy changes are:

1) Provision of shared ride incentives for all CBD workers,

2) Elimination of all public transportation fares, and

3) Increase in public transportation service to reduce both in and out of vehicle time to one half of its current values.

For each of these cases both the point predictions obtained by each of the prediction procedures and the difference predictions, that is the difference between the base case and the change case, are examined.

Prediction errors, both for aggregation error alone and for total error, are summarized in terms of the weighted root mean square error in terms of the difference in prediction and the reference value (enumeration prediction for analysis of aggregation error or observed share for analysis of total error) divided by the prediction. This error measure is equivalent to the average loss measure defined in Chapter V with errors weighted by the square of their magnitude and by the magnitude of the prediction with which they are associated.

The error measures are computed for the complete set of predictions (predictions for three modes for each prediction group in the set of interest). In some cases the error measures are reported separately for each mode. In addition, the error measures are sometimes further disaggregated to average error and standard deviation of the error around the average. These errors are related to total root mean square error by:

$$RMSE^2 = AE^2 + SD^2 \tag{6.9}$$

where RMSE ..... is the root mean square error,

AE ..... is the average error, and

SD ..... is the standard deviation of the distribution of errors around their average value.

6.7 Analysis of Aggregate Prediction Errors for the Base Case

This section considers the aggregation error and total error in prediction for different aggregate groups. The aggregation error of the naive, statistical differentials and classification procedures is isolated by comparing the predictions by these procedures to those obtained by the enumeration procedure. Since each of these predictions contains the same model errors, the differences between them represents errors due to aggregation only. The error measures associated with aggregation error for the three different groups of districts defined earlier are presented in Table 6.11, Base Case Aggregation Error for Three Groups of Districts. The naive and classification procedures have similar error measures for each of the three groups. The statistical differentials procedure has relatively high error for group one, lower error for group two, and much lower error for group three. In all three groups, the classification procedure has the least prediction error. The naive procedure has the second least prediction error for groups one and two while the statistical differentials procedure has the second least predictive error for group three. Analysis of the characteristics of the data in the three district groups (Tables 6.8 to 6.10) indicates that the most significant difference among them is that district group three has a much lower proportion of individuals without the drive alone alternative (approximately 8.5%) than either of the other groups (29% for district group one, 24% for district group two). This relationship between the error by the statistical differentials

TABLE 6.11

BASE CASE AGGREGATION ERROR*
FOR THREE GROUPS OF DISTRICTS

PREDICTION GROUP	AGGREGATION PROCEDURE		
	NAIVE	STATISTICAL DIFFERENTIALS	CLASSIFICATION
DISTRICT GROUP ONE	7.7	16.2	3.3
DISTRICT GROUP TWO	8.3	11.8	3.1
DISTRICT GROUP THREE	8.2	6.2	3.5

procedure and the proportion of individuals who do not have the drive alone alternative available to them is explored further in this section.

Table 6.12, Base Case Aggregation Error for Four Levels of Aggregation, presents measures of aggregation error for the naive, statistical differentials, and classification procedures for observations grouped to 45 districts, 10 super districts, 4 rings and a single region wide group. The aggregation error of the different methods is independent of the level of aggregation.* This is due to the fact that the higher levels of aggregation do not result in a major increase in the variance of the net utility distributions** and, to the extent that increases in variance occur, they are partially offset by a reduction in skewness of the distributions. The higher error measure for the region as a single group is consistent with the increase in the average variance of the net utility distributions which results from aggregation over the wide range of tripmakers and level of service characteristics which are included in the total region group.

Additional insight into the structure of errors by the different procedures is obtained by disaggregating the error measures for the district level prediction by modes and by average error and variations in error. This is done in Table 6.13, Base Case Aggregation Error by

---

*Error measures for the region are based on a single observation and are, therefore, not reliable.

**This is consistent with Fleet and Robertson's (1968) observation that 80% of the variability of selected socio-economic characteristics is due to within zone differences.

TABLE 6.12

BASE CASE AGGREGATION ERROR FOR
FOUR LEVELS OF AGGREGATION

PREDICTION GROUP	AGGREGATION PROCEDURE		
	NAIVE	STATISTICAL DIFFERENTIALS	CLASSIFICATION
DISTRICTS (45)	8.1	12.9	3.3
SUPER-DISTRICTS (10)	8.5	11.8	2.9
RINGS (4)	8.4	12.2	2.8
REGION (1)	10.2	13.5	1.3

TABLE 6.13

BASE CASE AGGREGATION ERROR
BY MODE FOR DISTRICTS

MODE	ERROR TYPE	AGGREGATION PROCEDURE		
		NAIVE	STATISTICAL DIFFERENTIALS	CLASSIFICATION
DRIVE ALONE	AE	+5.7	+1.1	+0.8
	SD	3.7	3.4	2.3
	RMSE	6.8	3.6	2.4
SHARED RIDE	AE	+2.0	+10.5	-1.9
	SD	2.4	11.0	3.1
	RMSE	3.1	15.2	3.6
TRANSIT RIDE	AE	-9.6	-12.5	+0.6
	SD	7.2	12.1	3.9
	RMSE	12.0	17.4	3.9
ALL	AE	6.5	9.4	1.2
	SD	4.8	8.8	3.1
	RMSE	8.1	12.9	3.3

Note:

AE = AVERAGE ERROR
SD = STANDARD DEVIATION OF ERROR
RMSE = ROOT MEAN SQUARE ERROR

Mode for Districts. The error for all modes (average error, standard deviation of error and root mean square error) is the square root of the average squared value of the corresponding error for each mode. The aggregation error for this prediction situation appears partially as an average over-or under-prediction for different modes and partially as a deviation from this average error. Predictions with large measures of average error generally have large measures of standard deviation of the error also. With one exception (prediction of the drive alone share by the statistical differentials procedure versus the naive procedure) the ranking of procedures for each error element is the same as for total error.

The fact that the shared ride errors by the statistical differentials procedure are predominantly positive and the transit ride errors are predominantly negative suggests that the poor performance of this method is due to an inability to satisfactorily predict the split between the shared ride and transit ride. District by district analysis of prediction errors indicates that the large errors in predicting shared ride and transit ride shares by the statistical differentials procedure occur in those districts which have a high proportion of individuals who do not have the drive alone alternative available to them. This further suggests that the correspondence between high aggregation errors of predictions by the statistical differentials procedure and the proportion of individuals without the drive alone alternative observed earlier is due to the errors in predicting the binary mode choice between shared ride and transit ride when the drive alone alternative is not available.

Errors in prediction other than aggregation error are identified by comparing predictions by the enumeration procedure to the observed shares. The errors in the enumeration procedure include stochastic error in the choice model and variables, error in the observed share and, if it exists, error due to application of the model in a situation different from the one in which it was calibrated.

Error in observed shares is inversely proportional to the square root of the number of observations in the prediction groups (Section 5.2).Stochastic error in the choice model is proportional to the same factor to the degree that errors among individuals are uncorrelated (Section 5.3 and 6.4). Error due to transferability is independent of the number of observations in the prediction groups.

The combined effect of error in observed shares, model and variable error and error of transferability is shown in Table 6.14, Base Case Error by the Enumeration Procedure. The differences in error for the three groups of districts is presumably due to both errors of transferability and differences in the average number of observations in districts in the three groups. This question will be examined shortly.

The magnitude of error declines with increasing levels of aggregation. This is a direct result of the increase in the number of observations in each prediction unit from an average of about 50

TABLE 6.14

BASE CASE ERROR
BY THE ENUMERATION PROCEDURE

PREDICTION GROUP	DRIVE ALONE	SHARED RIDE	TRANSIT RIDE	ALL MODES
DISTRICT GROUP ONE	18.4	34.4	22.8	25.6
DISTRICT GROUP TWO	18.7	32.2	30.5	27.1
DISTRICT GROUP THREE	22.9	41.1	37.2	32.9
ALL DISTRICTS	19.8	35.4	28.7	27.8
SUPER DISTRICTS	12.3	24.9	21.6	19.7
RINGS	5.4	18.0	20.1	15.4
REGION	1.6	7.1	4.1	4.5

at the district level to 200 at the super district level to 500 at the ring level. Figure 6.5, Model and Observation Error Related to Average Size of Prediction Group, indicates the relationship between model and observation error and the number of observations in the average group. The fact that the magnitude of error does not decline with the square root of the average number of observations indicates that there is correlation between the errors in individual choice probability estimates (as discussed in Sections 5.2 and 6.4). This suggests that for reasonably sized prediction groups (less than 200 to 500 per prediction), the errors in share prediction due to errors in the choice model and variables is significant. The existence of error at the region level is attributable to application of the model to a systematically different data set (the region wide sample) from that on which it was estimated (a geographically selected sample).

The significance of such transferability errors may be examined by use of the data in Table 6.15, Base Case Total Error for District Groups by the Enumeration Procedure. In this table the root mean square error for each mode and for all modes combined is disaggregated by average error and standard deviation of the error. The differences in errors for the three groups of districts are partially due to the average error in mode prediction and partially to differences in the standard deviation of the errors. The increase in standard deviation is consistent with differences in the average number of observations per district in each group.

FIGURE 6.5

MODEL AND OBSERVATION ERROR
RELATED TO AVERAGE SIZE
OF PREDICTION GROUP

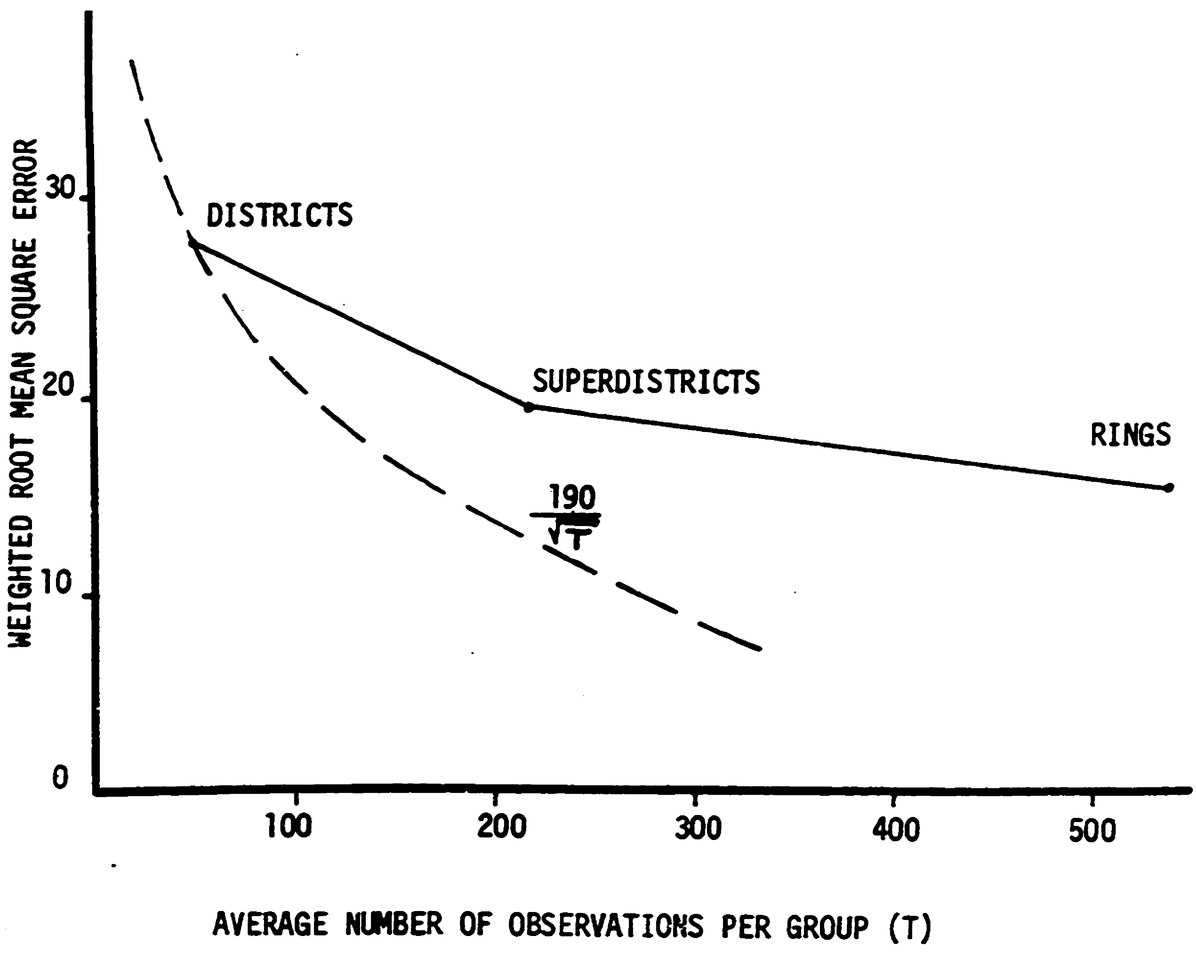


TABLE 6.15

BASE CASE TOTAL ERROR
FOR DISTRICT GROUPS
BY THE ENUMERATION PROCEDURE

PREDICTION GROUP	ERROR TYPE	DRIVE ALONE	SHARED RIDE	TRANSIT RIDE	ALL MODES
DISTRICT GROUP ONE	AE	0	0	0	0
	SE	18.4	34.4	22.8	25.6
	RMSE	18.4	34.4	22.8	25.6
DISTRICT GROUP TWO	AE	-4.9	-7.7	+11.3	8.3
	SD	18.0	31.2	28.3	25.8
	RMSE	18.7	32.2	30.5	27.1
DISTRICT GROUP THREE	AE	+12.3	-18.5	-1.1	12.8
	SD	19.3	36.7	37.2	30.6
	RMSE	22.9	41.1	37.2	32.9
ALL DISTRICTS	AE	+1.6	-7.1	+4.1	4.5
	SD	19.7	35.3	28.4	27.3
	RMSE	19.8	35.4	28.7	27.8

Note:

AE = AVERAGE ERROR
SD = STANDARD DEVIATION OF ERROR
RMSE = ROOT MEAN SQUARE ERROR

The average error in share predictions by the enumeration procedure for district groups two and three is due to the application of the model estimated on data from district group one (which has zero average error) to these groups. The magnitude of the average errors for individual modes are small compared to the standard deviation of the errors for the individual districts.

Total error in prediction for the base case is identified by comparing the predictions by the naive, statistical differentials and classification procedures to the observed share. Table 6.16, Base Case Total Error for Four Levels of Aggregation, presents the overall error measures for these procedures and for the enumeration procedure for predictions at four levels of aggregation. The total errors reported are of similar magnitude for all of the aggregation procedures. The aggregation error of the naive and classification procedures is completely submerged in the errors in the observed shares and in the choice model. The fact that the error measures for these methods are less than those for the enumeration procedure are due to the fact that the average errors in these procedure happen to offset, in part, the average error due to transferability in the enumeration procedure. The aggregation error of the statistical differentials procedure is large enough that it results in an increase in total error of this procedure over the total error of the enumeration procedure.

To increase our understanding of the effect of the number of observations in the prediction unit on reported error, without the confound-

TABLE 6.16

BASE CASE TOTAL ERROR
FOR FOUR LEVELS OF AGGREGATION

PREDICTION GROUP	AGGREGATION PROCEDURE			
	ENUMERATION	NAIVE	STATISTICAL DIFFERENTIALS	CLASSIFICATION
ALL DISTRICTS	27.8	27.7	31.2	27.2
SUPER-DISTRICTS	19.7	19.1	24.8	18.6
RINGS	15.4	15.4	22.4	14.0
REGION	4.5	7.5	9.2	3.3

ing effect of different levels of aggregation, error measures for predictions of districts and super districts disaggregated into two groups in each case for larger versus smaller districts and larger versus smaller super districts are given in Tables 6.17, Base Case Aggregation Error Disaggregated by Size of Prediction Group. The aggregation error reported in Table 6.17 is not substantially different between the larger and smaller districts or the larger and smaller super districts except for the statistical differentials procedure which has a much lower error for small super districts than larger super districts.*

The total error given in Table 6.18 is substantially larger for small versus large districts and small versus large super districts. For all of the groupings in this table, the error from observations and from the disaggregate choice model dominate the aggregation error so that the performance of the enumeration, naive and classification procedures are not differentiable while in some cases the statistical differentials procedure has higher total error.

Throughout this analysis the errors in prediction by the classification procedure have been consistently low. It is useful to analyze classification procedures further to determine the reason for this extremely good performance. One element of the classification procedure used is to group individuals according to whether or not they have the

---

*This observation is again associated with a low proportion of individuals who do not have the drive alone alternative in small super districts.

Table 6.17

BASE CASE AGGREGATION ERROR DISAGGREGATED
BY SIZE OF PREDICTION GROUP

A. Districts

Method	Naive	Statistical Differentials	Classification
Large Groups[1]	8.2	13.5	3.2
Small Groups[2]	7.8	10.4	3.4

B. Super Districts

Method	Naive	Statistical Differentials	Classification
Large Groups[3]	8.9	13.0	3.0
Small Groups[4]	6.7	4.5	2.3

[1] 14 Districts with 451 observations averaging 32.2 observations per district.

[2] 31 Districts with 1681 observations averaging 54.2 observations per district.

[3] 4 Super districts with 481 observations averaging 104.5 observations per super district.

[4] 6 Super districts with 1714 observations averaging 285.7 observations per super district.

Table 6.18

BASE CASE TOTAL ERROR DISAGGREGATED BY
SIZE OF PREDICTIVE GROUP

A. Districts

Method	Enumeration	Naive	Structural Differential	Classification
Large Groups	25.3	25.9	29.5	24.7
Small Groups	35.8	33.7	36.8	35.3

B. Superdistrict

Method	Enumeration	Naive	Structural Differential	Classification
Large Groups	17.1	16.0	23.6	15.7
Small Groups	27.9	28.6	29.2	27.6

drive alone alternative available to them. The importance of the proportion of individuals having the drive alone alternative has been mentioned with regard to the performance of the statistical differentials procedure. To determine the relative importance of the classification according to the availability of the complete or partial choice set compared to classification according to automobile availability, two additional classification procedures are proposed. Classification procedures two is based on automobile availability only. It groups individuals according to whether the ratio of autos to licensed drivers is less than 0.25, greater than 0.25 but less than 1.0, or equal to 1.0. Classification procedure three groups individuals solely in accordance as to whether they have the drive alone alternative available to them. Classification procedure one, previously described, groups individuals according to whether they have the drive alone alternative available to them and, if they have the alternative available, according to whether the ratio of automobiles to number of licensed drivers is less than 1.0 or equal to 1.0. All of the other variables are included at their mean value for the entire group for all three procedures.

The error measures for these three classification procedures with respect to aggregation error and total error are given in Table 6.19, Base Case Aggregation and Total Error for Different Classification Procedures for Three Levels of Aggregation. Classification procedure one, has the lowest aggregation error of the three procedures. Classification procedure three which depends on selection by choice set avail-

TABLE 6.19

BASE CASE AGGREGATION AND TOTAL ERROR FOR DIFFERENT CLASSIFICATION PROCEDURES FOR THREE LEVELS OF AGGREGATION

A. Aggregation Error

Prediction Group	Aggregation Procedure		
	Class One	Class Two	Class Three
Districts	3.3	9.9	5.2
Super Districts	2.9	8.4	5.2
Rings	2.8	7.8	5.3

B. Total Error

Prediction Group	Aggregation Procedure			
	Enumeration	Class One	Class Two	Class Three
Districts	27.8	27.2	29.0	26.7
Super Districts	19.7	18.6	21.0	17.6
Rings	15.4	14.0	17.8	12.2

ability alone has the second lowest error measure and classification procedure two which classifies according to automobile availability only has a much higher error measure. Classification procedure two has error measure values which are approximately equal to those for the naive procedure. The total error for each of the three methods is related to the level of aggregation and implicitly to the average size of the prediction group. The two classification procedures which include the choice set availability, classification procedure one and classification procedure three, have total error which is approximately the same as that of the enumeration procedure. Procedure two which classifies only according to automobile availability has higher total error measures than the enumeration procedure. These results confirm the importance of knowledge of the choice set availability in the prediction process.

6.8 Analysis of Aggregate Prediction Errors for Policy Changes

The purpose of this section is to examine the effect on the previously described aggregation errors of shifting the choice shares to different points on the choice function. An analysis of errors is made for the prediction of the new shares and the changes in shares which result from the policy change considered. Three different policy changes are considered. They are:

1) Provide shared ride incentive to all trip makers,

2) Reduction of transit fare to zero, and

3) Reduction of in and out of vehicle transit times by one half.

The new predicted share for each of the policy changes considered and the share for the base case estimated by the enumeration procedure are shown in Table 6.20, Choice Share for Three Policy Changes. The share values vary from 39% in the base case to 35% for change one and change three for the drive alone mode, from 36% for change one to 23% for change three for the shared ride mode and from 30% for change one to 42% for change three for the transit ride mode.

The aggregation error for the change predictions is given in Table 6.21, Aggregation Error in Share Prediction for Three Policy Changes for the naive, statistical differentials and classification procedures. The classification procedure has the least error and the naive procedure has the second least error for all the changes and the base case at each level of aggregation. As observed for the base case

TABLE 6.20

CHOICE SHARES FOR THREE POLICY CHANGES

PREDICTION SITUATION	DRIVE ALONE	SHARED RIDE	TRANSIT RIDE
CHANGE ONE	.35	.36	.30
CHANGE TWO	.37	.26	.37
CHANGE THREE	.35	.23	.42
BASE	.39	.28	.33

TABLE 6.21

AGGREGATION ERROR IN SHARE PREDICTION FOR THREE POLICY CHANGES

PREDICTION SITUATION	LEVEL OF AGGREGATION	AGGREGATION PROCEDURE		
		NAIVE	STATISTICAL DIFFERENTIANS	CLASSIFICATION
CHANGE ONE	DISTRICTS	9.0	14.2	3.0
	SUPER DISTRICTS	9.6	12.8	3.1
	RINGS	9.5	12.7	3.1
CHANGE TWO	DISTRICTS	7.1	12.3	3.4
	SUPER DISTRICTS	7.4	11.6	3.0
	RINGS	7.2	11.8	3.0
CHANGE THREE	DISTRICTS	5.3	10.7	3.2
	SUPER DISTRICTS	5.3	9.7	2.8
	RINGS	4.9	9.9	2.9
BASE CASE	DISTRICTS	8.1	12.9	3.3
	SUPER DISTRICTS	8.5	11.8	2.9
	RINGS	8.4	12.2	2.8

(Table 6.12) the error measures are independent of the level of aggregation. The errors are of different magnitudes for the three different policy changes. The different prediction situations in order of increasing error are : change three, change two, base case and change one. There is no way to examine total error for these three situations as there are no observed shares corresponding to these policies.

The aggregation errors for prediction of the differences between shares before and after each of thepolicy changes are presented in Table 6.22, Aggregation Error in Difference Predictions for Three Policy Changes. The error measures in this table are greater than any error measures in the preceding tables. This is primarily due to the change in the denominator from a share prediction, generally between 20 to 40 percent, to a change in the share prediction, generally under eight percent in magnitude. The magnitude of error by the naive and classification procedures is independent of level of aggregation for each set of difference predictions. However, the statistical differentials procedure has much greater error at the district level than at higher levels of aggregation.

An alternative view of the prediction of the effect of policy changes is to use the prediction of differences as a means of predicting the new shares rather than predicting the shares directly. In this case, the appropriate basic error measure is the error in difference prediction divided by the new share prediction. The corresponding error measures are presented in Table 6.23, Aggregation Error in Share Prediction

TABLE 6.22

AGGREGATION ERROR IN DIFFERENCE PREDICTION FOR THREE POLICY CHANGES

PREDICTION SITUATION	LEVEL OF AGGREGATION	AGGREGATION PROCEDURE		
		NAIVE	STATISTICAL DIFFERENTIATION	CLASSIFICATION
CHANGE ONE	DISTRICTS	15.7	47.4	13.1
	SUPER DISTRICTS	13.4	31.0	11.9
	RINGS	13.0	28.5	11.9
CHANGE TWO	DISTRICTS	20.6	14.6	8.1
	SUPER DISTRICTS	20.2	7.4	7.5
	RINGS	20.5	6.7	7.4
CHANGE THREE	DISTRICTS	16.2	24.3	6.4
	SUPER DISTRICTS	16.6	14.4	6.4
	RINGS	17.3	14.5	6.1

TABLE 6.23

AGGREGATION ERROR IN SHARE PREDICTION BY THE INCREMENTAL APPROACH FOR THREE POLICY CHANGES

PREDICTION SITUATION	LEVEL OF AGGREGATION	AGGREGATION PROCEDURE		
		NAIVE	STATISTICAL DIFFERENTIATION	CLASSIFICATION
CHANGE ONE	DISTRICTS	2.5	3.4	2.3
	SUPER DISTRICTS	2.4	2.4	2.2
	RINGS	2.4	2.2	2.2
CHANGE TWO	DISTRICTS	1.5	1.2	0.6
	SUPER DISTRICTS	1.4	0.5	0.5
	RINGS	1.4	0.5	0.5
CHANGE THREE	DISTRICTS	2.6	2.0	1.0
	SUPER DISTRICTS	3.1	3.0	1.4
	RINGS	3.3	3.0	1.3

by the Incremental Approach for Three Policy Changes. The aggregation errors in this case are substantially lower than the errors of direct prediction of the new shares given in Table 6.21. As observed pervously, the aggregation error is relatively unaffected by changes in the level of aggregation but is different for the three different policy changes. This result strongly supports the concept of incremental prediction.

6.9 Results of Analysis of Aggregate Prediction Errors

This empirical analysis investigated both the aggregation error and the total error of alternative prediction methods. The aggregation errors identified conform with expectations developed in the preceding mathematical and simulation analyses with respect to the importance of the distribution of utilities in the aggregate prediction groups.

The aggregation error for each of the methods used is independent of the level of aggregation. The effect of increases in the average variance of net utilities at higher levels of aggregation (which is expected to cause an increase in aggregation error) appears to be offset by a compensating reduction in skewness. The combined affect of these changes is neutral with respect to aggregation error at different levels of aggregation (Tables 6.12 and 6.21).

The classification procedure has low aggregation error for all of the situations considered. This is consistent with the expected robustness of the classification procedure indicated by the low average and maximum absolute aggregation bias and low average absolute unit bias, for different levels of variance and skewness identified in the simulation analysis (Tables 4.1 to 4.6). The classification procedure has consistently low levels of aggregation bias for predictions for each of the modes (Table 6.13). The aggregation error for the classification procedure varies widely, however, for different types of classification. Classification procedures which take account of differences in the choice set available to different individuals predicted substantially better than procedures which did not include this classification element (Table 6.19).

The aggregation error for the naive procedure was generally in the range of eight percent of the prediction (Tables 6.12 and 6.21). This is about the same as the classification procedure based on auto availability alone but about twice as high as the aggregation error for classification procedures which took account of differences in the available choices.

The statistical differentials procedure generally had higher levels of aggregation error and was more erratic (Tables 6.12 and 6.21) than the other procedures. Aggregation errors using this procedure are particularly sensitive to the proportion of individuals in the prediction group which did not have the drive alone alternative available to them.

Similar results are obtained for prediction of the base case and prediction for three alternative situations which represented possible changes in transportation operating policies. Error measures for the prediction of changes in shares resulting from changes in transportation operating policies are larger as a percentage of the difference predicted (than the errors in share predictions reported as a percentage of the share predicted) and more variable especially for predictions based on the statistical differentials procedure (Table 6.22). Use of the prediction of differences to obtain share predictions for a changed situation by an incremental procedure results in considerably lower aggregation error in revised share predictions (Table 6.23).

The total error in prediction is examined to place the aggregation error in perspective with respect to other sources of error. The measurement of total error in prediction is partially confounded by error in the observed share as an estimate of true shares. The error in observed shares combined with error in the choice model generally overwhelm the error from the aggregation procedures. This makes it impossible to identify significant differences between the prediction of errors of the enumeration, naive and classification procedures. The statistical differentials procedure, however, had larger total errors than the other procedures.

The magnitude of the combined error from the choice model and the observations (error measures for the enumeration procedure which has no aggregation error) varies, as expected, with the average number of observations in the predictive group from 28% at the district level to 20% at the super district level to 15% at the ring level of aggregation. Compared to this the aggregation error is approximately 3% for the classification procedure, 8% for the naive procedure and 12% for the statistical differentials procedure for all of the levels of aggregation considered. The errors from the choice model and the observations are primarily stochastic (Table 6.15). The aggregation error for the naive and statistical differentials procedures is primarily the result of a consistent average error or bias in the share predictions.

This empirical study confirms the existence of aggregation error which, measured as a proportion of the magnitude of prediction, ranges from about three to thirteen per cent for different aggregation procedures. The study also indicates that the magnitude of error due to error in the choice model and error in observations is generally larger than the aggregation error. The interaction of these errors is such as to make differences in the aggregation error less than eight or ten percent indistinguishable with respect to their impact on total prediction error.

Prediction of choice shares for modified situations (three policy changes with respect to shared ride or transit ride level of service) indicates (1) the level of aggregation error for each method is different for different situations and (2) the relative ranking of the naive, statistical differentials and classification procedures are the same for the different situations. The prediction of differences in shares before and after a policy change contains much higher error (in terms of error in difference prediction per unit of difference prediction) than the prediction of shares themselves (Table 6.22). However, use of difference predictions as an incremental approach to prediction of revised shares has substantially lower error than the direct prediction of revised shares (Table 6.23).

In summary, the results of this empirical study are:

1. Aggregation error in share prediction of the naive and statistical differentials procedures is substantial.
2. Aggregation errors are relatively independent of the level of aggregation (given an initial degree of aggregation).

3. Aggregation error of the statistical differentials procedures is extremely sensitive to the proportion of the aggregate group which does not have the complete choice set available.
4. Aggregation error of the naive procedure (and presumably the statistical differentials procedure) may be reduced substantially by classification of the prediction group in accordance with differences in the set of available choices.
5. Aggregation error for prediction of shares for three different policy situations is similar in magnitude to the aggregation error for the base case.
6. Aggregation error in the prediction of shares for a policy change situation by the incremental procedure (base case plus difference) is substantially lower than the aggregation error in direct prediction of the changed shares.
7. Errors in observed shares and errors in choice model parameters and variables are large relative to aggregation errors for prediction at levels of aggregation with less than 200 observations in each prediction group.

CHAPTER VII

CONCLUSIONS AND RESEARCH DIRECTIONS

7.1 Conclusions

This research study investigates the issues associated with prediction of aggregate travel demand based on models of individual choice behavior. The study focuses on the sources of error in prediction with an aggregated disaggregate choice model.

The results indicate that it is feasible to predict aggregate travel demand using an aggregated prediction model which is based on a disaggregate model of travel choice behavior. The results also indicate that the errors in prediction by the procedures considered may be significant*. Care must be taken to assure that these errors do not exceed acceptable limits in the context of particular prediction situations.

Errors in prediction result from bias and/or random error in the disaggregate choice model and the prediction of choice influencing variables as well as bias error which is introduced by approximate aggregation procedures.

Theoretical and simulation analyses demonstrate that the bias error due to aggregation is dependent on the distribution in the prediction

---

*No attempt is made in this study to compare predictions based on aggregated disaggregate models to predictions based on models estimated with aggregate data. A study by Watson (1975) indicates that errors in prediction with models based on aggregate analysis are larger than those based on disaggregate analysis.

group of net utilities between pairs of alternatives in the choice set. Specifically, aggregation error is affected by the mean (or location on the choice function), variance and shape of the distribution of net utilities. These analyses indicate that the aggregation errors which result from use of the naive procedure (assumed homogeneity of the prediction group) are substantial for some distributions of net utility. These analyses also indicate that the aggregation error associated with the naive aggregation procedure can be substantially reduced by alternative aggregation procedures based on relatively limited knowledge of the distribution of variables in the prediction group.

An empirical study of the prediction of mode share for breadwinners making a work trip to the central business district is described. The expected aggregation bias of share predictions for a wide range of situations (different mode share proportions and levels of aggregation) is less than ten percent of the predicted values. In all of the situations considered the expected aggregation error is reduced to under four percent by classification of the prediction group according to differences in choice set availability.

Contrary to expectations, predictions based on the statistical differentials procedure have greater aggregation error than those based on the naive procedure in almost all of the prediction situations. The high levels of error by the statistical differentials procedure are associated with predictions for groups with a large proportion of persons who do not have all alternatives available to them. This observation and the error reduction gained by classification in accordance with choice set availability indicates the importance of this characteristic.

Use of an incremental approach to the prediction of expected choice shares due to a change in transportation service (expected share is determined by adding the predicted change in share to the original share) reduces the aggregation error by the procedures studied to less than one-half of the error in direct prediction of changed shares.

Aggregation error for each of the aggregation procedures studied appears to be independent of the level of aggregation or the number of observations in the prediction group. Other sources of error (error in choice model and choice variable predictions) are sensitive to the number of observations in the prediction group. It is impossible to identify the error in prediction from these sources as error measurements are confounded by error in the observed shares which are the natural base for the comparison of predictions.

The expected error from all these sources, in the case studied, is approximately 28% of the predictions for groups with an average of fifty observations, 20% for groups with 200 observations and 15% for groups with 500 observations. (The corresponding expected error in observed shares is approximately 20%, 10% and 7%, respectively). These errors are large enough to overwhelm the error due to aggregation so that total error in prediction with the naive or classification procedures is similar to that for the enumeration procedure. The key conclusions of this study are:

- It is feasible to predict aggregate travel demand using aggregated models of individual choice behavior,
- The observed error, for the prediction of mode choice to work, is substantially less than the theoretical maximum of average error,
- The error in prediction by the naive procedure may be substantially reduced by use of classification based on choice set availability or use of an incremental prediction procedure.
- The contribution of aggregation error to total error in prediction given existing error from other sources is negligible for both the naive and classification procedures,
- Estimates of total prediction errors are large enough to require that they be reported with their corresponding predictions.

7.2 Approach to Prediction with Aggregated Prediction Models

The results of this study indicate that errors in prediction due to aggregation vary in magnitude depending on the characteristics of the aggregate group for which predictions are required. In the empirical case described the aggregation error of the naive procedure is considerably less than the average error per unit of prediction based on the simulation analysis of aggregation error. The empirical study also indicated the importance of differences in choice set availability on the aggregation error of different aggregation procedures.

These results suggest some guidelines which may be used in approaching a specific prediction situation. These guidelines identify situations which result in large magnitudes of aggregation bias and suggest procedures for reducing the aggregation bias when it occurs.

The characteristics which lead to large magnitudes of aggregation bias are:

- Location of the distribution near the point of maximum curvature on the choice function (this is at a mode share split of approximately 20% or 80% for the binary choice case),
- Large variance in the distribution of net utility between pairs of alternatives (variance values greater than 1.0 to 1.5 are "large"),
- Differences in choice set availability, and
- Extreme deviations from normality in the distribution of net utility between pairs of alternatives.

To some extent these conditions, especially large variance and differences in choice sets can be avoided in the construction of prediction groups. That is, the need to perform classification (or use other than the naive procedure for aggregation) can be avoided by designing prediction groups which are reasonable homogeneous with respect to characteristics which influence the choice for which the prediction is required.

The conditions which are identified as contributing to large bias are partially inter-related. That is, variance greater than the indicated range has a small affect on the bias when the share split is near 50%. Similarly, location of the distribution near the point of maximum curvature has little effect on aggregation bias when the variance in the net utility distribution is extremely low.

When the conditions identified cannot be avoided three approaches are suggested for reduction of the aggregation error which is expected to occur. These are:

- Classification of the prediction groups to increase their homogeneity (the most important element of classification is according to differences in choice set availability, beyond this classification should be according to those variables which contribute most to the variance in the distribution of net utility),
- Use of incremental procedures to predict the travel choice shares which occur as a result of transportation policy changes,

- Procedures for modifying the initial predictions based on yet to be developed bias reduction rules; these rules will make average adjustments in predictions based on the bias expected for the conditions of the prediction (Landau, 1975).

These suggestions for identifying situations which may have large bias and taking action to reduce the bias are based on a limited range of study. These rules should be refined and extended as additional experience with aggregated prediction models is obtained.

7.3 Implications of the Analysis Results for the Transportation Planning Process

The transportation planning process is oriented to the selection of transportation policies from among a wide range of policies which may be considered for implementation. A major element in the evaluation and selection among alternative policies is the prediction of the effects of these alternatives on transportation system performance and on the external environment. Predictions made in the past have been criticized for their failure to be sensitive to differences in the alternatives under consideration.

This study demonstrates that it is feasible to use disaggregate choice models which are sensitive to level of service characteristics to predict aggregate travel demand.

The analysis also identifies the potentially important magnitudes of error in prediction. The joint result of more sensitive predictions reported with uncertainty provides benefits in terms of better decision making but calls for increased effort in structuring and applying the decision process. Considerations of risk and uncertainty can be dealt with explicitly and periodic updating of predictions and decisions should be part of the ongoing planning process (Neumann, 1975).

7.4 Research Directions

This study explores the issues of using disaggregate travel choice behavior models to predict aggregate travel demand. The analysis indicates that this is a feasible approach at least for the prediction of mode shares for the home based work trip. The results of the study also indicate a range of research which should be undertaken to verify and extend these results and to clarify related issues which arise in the prediction of travel demand.

There is a need for more extensive theoretical analysis of the sources of error in prediction, their interaction and propagation. Specific research areas are:

- Methods for improving the precision of parameter estimates in the choice model. These include consideration of improvements in model specification with respect to model form and selection of choice variables.
- Methods for determining the precision of the predictions of choice variables. This requires analysis of the models used for prediction of choice variables and identification of the errors in these models.
- Analysis of the interaction of errors from different sources taking account of structural interactions which may increase or reduce the prediction error.
- Analysis of the errors in multi-dimensional prediction situa-

tions with consideration of the differences in errors which result from the direct prediction of joint shares rather than recursive prediction.

There is a parallel need for <u>applied analysis of errors in prediction</u> to supplement and verify the results of theoretical analyses and identify special problems which may arise in application. Such efforts should include:

- Verification of the results of the present study in different prediction situations, that is, for different choices and in different geographic areas.
- Extension of the study approach to ex post (independent variables predicted rather than obtained from observed data) predictions.
- Extension of the study approach to prediction of multi-dimensional choice shares analyzing the affect on both total error and aggregation error of sequential versus direct prediction.
- Modification of the prediction procedures to take account of differences in choice set availability.

There is no existing basis for comparing the results of this analysis to predictions by conventional methods. Explicit tests should be constructed to <u>make comparisons of prediction error using aggregated prediction models and aggregate models based on analysis of aggregate data</u>. Comparisons should be made for prediction of the data on which the models are estimated (Watson, 1975) and for prediction of travel demand under different sets of conditions. Analysis of errors in prediction should identify total error and, where possible, identify the source of major error components.

The model specification used in this study is relatively simple. The effect of improvement in model specification on prediction errors should be determined in different contexts. Explicit attention should be directed toward the identification of relationships between choice model estimation statistics and error measures in aggregate prediction.

Research should be directed toward refinement of the guidelines suggested in Section 7.2 for identifying situations of large bias and the development of heuristics for the reduction of aggregation error in those situations where it occurs. Attention should be directed to identify the information required to undertake bias reduction procedures and methods for obtaining the required information.

Prediction of small choice shares requires special attention. The theoretical and simulation analysis of aggregation error indicates that this error as a proportion of the prediction is high for small choice shares. To keep this in perspective, the standard error in choice prediction due to error in the choice model as a proportion of the prediction and the standard error of observed shares as a proportion of true share are also high for small shares. Specific research should be directed to the development of procedures for improving the prediction of small choice shares.

The affect of different data collection approaches should be explored with respect to their impact on model estimation and prediction analysis. Specific issues to be researched are:

- The affect of different sampling procedures on the validity of estimates of parameters of disaggregate choice models.
- The affect of low rate sampling and the resultant error in observed choice shares on the estimates of aggregate shares used for calibration of aggregate model parameters.
- The need for data to be used in determining the distribution of choice variables.

More generally, the error analysis approach used in this study suggests the use of a decision analysis approach to the development of future research priorities. To the extent that research effort will be expended on improvements to prediction modelling capabilities the potential for error reduction and the research and application cost required to obtain such error reduction should enter into both the research priorities decision and the allocation of effort for applied model building.

Finally, the recognition of error in predictions makes apparent a new element of uncertainty in the transportation planning process. Developments in the planning and programming process should account for this additional element of uncertainty.

BIBLIOGRAPHY

Adler, T., and M.E. Ben-Akiva, "A Joint Frequency Destination and Mode Choice Model for Shopping Trips", paper submitted to the Highway Research Board, Washington, D.C., August, 1974.

Allen, R.G.D., Mathematical Economics, MacMillian and Co. Ltd., London, 1956.

Atherton, Terry J., Approaches for Transferring Disaggregate Travel Demand Models, M.S. Thesis, Dept. of Civil Engineering, M.I.T., February, 1975.

Attanucci, John P. Analysis of Car Pooling Behavior and the Formulation of Car Pool Incentive Programs, S.M. Thesis, Dept. of Civil Engineering, M.I.T., Cambridge, Mass. 1974.

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

Ben-Akiva, Moshe E., and Martin G. Richards, "A Disaggregate Multi-Modal Model for Work Trips in the Netherlands," presented at the 54th Annual Meeting of the Transportation Research Board, Washington D.C., 1975.

Brand, Daniel, "Theory and Method in Land Use and Travel Forecasting," Highway Research Record No. 422, Highway Research Board, Washington, D.C., 1972.

Brand, Daniel, "Travel Demand Forecasting: Some Foundations and a Review," in Urban Travel Demand Forecasting, Highway Research Board Special Report 132, Washington, D.C., 1973.

Bureau of Public Roads, Urban Transportation Planning: General Information and Introduction to System 360, June 1970.

Cambridge Systematics, Inc., The Effect of Transit Service on Automobile Ownership, Research for U.S. D.O.T., Office of the Secretary, Contract DOT-OS-30056, 1975.

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," in Economic Forecasts and Expectations (Mincer,J., ed.), Columbia University Press, 1969.

Cox, D.R., "Some Procedures Connected with Logistic Qualitative Response Curve," Research Papers in Statistics: Festschrift for J. Neyman (F.N. David, ed.) Wiley, London, 1966.

de Donnea, F.X., The Determinants of Transport Mode Choice in Dutch Cities, Rotterdam University Press, Louvain, Belgium, 1971.

Difiglio, D., and M.F. Reed, Jr., "Transit Sketch Planning," presented at the 54th Annual Meeting of Transportation Research Board, Washington, D.C., 1975.

Fleet, C.R. and S.R. Robertson, "Trip Generation in the Transportation Planning Process," Highway Research Record, No. 240, Highway Research Board, 1968.

Green, H.A.J., Aggregation in Economic Analysis, Princeton University Press, Princeton, New Jersey, 1964.

Johnson, N.L., and S. Kotz, Distributions in Statistics, Volume I, Discrete Distributions, Wiley & Sons, New York, 1969.

Kanafani, Adib K., "An Aggregate Model of Trip Making," Transportation Research, Vol. 6, pp. 119-124, Pergamon Press, 1972.

Kannel, E.J., and K.W. Heathington, "The Temporal Stability of Trip Generation Relationships," paper prepared for presentation at the Annual Highway Research Board Meeting, Washington, D.C., August, 1973.

Kassoff, H. and H. Deutschman, "Trip Generation: A Critical Appraisal," Highway Research Record, No. 297, Highway Research Board, 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 Acacemic Bookstore, Helsinki, Finland, 1968.

Landau, Uzi, Strategic Analysis Methods in Transportation Systems Planning, Ph.D. Thesis in progress, Dept. of Civil Engineering, M.I.T., Cambridge, Mass. 1975.

Lave, Charles A., "A Behavioral Approach to Model Split Forecasting," Transportation Research, Vol. 3, pp. 463-480, 1969.

Lerman, Steven R., Towards Behavioral Modelling of Urban Residential Location, Ph.D. Thesis, M.I.T., Dept. of Civil Engineering, May, 1975.

Lerman, Steven R. and Moshe E. Ben-Akiva, "A Disaggregate Behavioral Model of Automobile Ownership," presented at the 54th Annual Meeting of the Transportation Research Board, Washington, D.C., January, 1975.

Lisco, Thomas E., The Value of Commuters' Travel Time - A Study in Urban Transportation, Ph.D. Dissertation, University of Chicago, 1967.

Liou, Peter S., Gerald S. Cohen, and David T. Hartgen, "An Application of Disaggregate Mode Choice Models to Systems-Levels Travel Demand Forecasting," presented at the 54th Annual Meeting of the Transportation Research Board, Washington, D.C., January, 1975.

Luce, R.D. and P. Suppes, "Preference, Utility and Subjective Probability," Handbook of Mathematical Psychology. R.D. Luce, gen. ed., John Wiley and Sons, New York, 1965.

Manheim, Marvin L., "Fundamental Properties of Systems of Demand Models," Discussion paper T70-1, Dept. of Civil Engineering, M.I.T. Cambridge, Mass., 1970.

Manheim, Marvin L., "Practical Implications of Some Fundamental Properties of Travel-Demand Models," Highway Research Record No. 422, Highway Research Board, Washington, 1972.

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

Martin, B.V., F.W. Memmott, and A.J. Bone, Principles and Techniques of Predicting Future Urban Area Transportation, M.I.T. Press, 1961.

McCarthy, G.M., "Multiple Regression Analysis of Household Trip Generation - A Critique," Highway Research Record No. 297, Highway Research Board, 1969.

McFadden, Daniel and Fred Reid, "Aggregate Travel Demand Forecasting from Disaggregated Behavioral Models," presented at the 54th Annual Meeting of the Transportation Research Board, Washington, D.C., 1975.

McGillivray, Robert G., Binary Choice of Transport Modes in the San Francisco Bay Area, Ph.D. Thesis (Economics), University of California, Berkeley, 1969.

Neumann, Lance A., The Limitations of the Comprehensive Plan: An Implementation Strategy, Ph.D. Thesis in progress, Dept. of Civil Engineering, M.I.T., Cambridge, Mass. 1975.

Pearson, Karl, "Contributions to the Mathematical Theory of Evolution," Philosophical Transactions of the Royal Society of London, 1895.

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

Stopher, Peter R., "Probability Model of Travel Mode Choice for the Work Journey," Highway Research Record, No. 283, Highway Research Board, 1969.

Stopher, Peter R. and J.O. Lavender, "Disaggregate Behavioral Travel Demand Models: Empirical Tests of Three Hypotheses," Transportation Research Forum Proceedings, 1972.

Stopher, Peter R. and T.E. Lisco, "Modelling Travel Demand: A Disaggregate Behavioral Approach, Issues and Applications," Transportation Research Forum Proceedings, 1970.

Tahir, Nadeem and Mark Hovind, "A Feasibility Study of Potential Feeder Bus Service for Homewood, Illinois," Proceedings of the Transportation Research Forum, 1973, pp. 553-570.

Talvitie, Antti, "Aggregate Travel Demand Analysis with Disaggregate or Aggregate Travel Demand Models," Proceedings of the Transportation Research Forum, 1973.

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

Theil, Henri, Linear Aggregation of Economic Relations, Amsterdam, North Holland Publishing Co., 1955.

Theil, Henri, "A Multinomial Extension of the Linear Logit Model," International Economic Review, 1969.

Theil, Henri, Principles of Econometrics, John Wiley and Sons, New York, 1971.

Tversky, A., "Elimination by Aspects: A Theory of Choice," Psychological Review, Vol. 79, No. 4, July, 1972, pp. 281-299.

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

Warner, S.L. Stochastic Choice of Mode in Urban Travel: A Study in Binary Choice, Northwestern University Press, Evanston, Illinois, 1962.

Watson, Peter L., "Comparison of the Model Structure and Predictive Power of Aggregate and Disaggregate Models of Intercity Mode Choice," Transportation Research Record, No. 527, Transportation Research Board, 1975.

Westin, Richard B., "Prediction from Binary Choice Models," Journal of Econometrics, April, 1974.

BIOGRAPHICAL SUMMARY

Frank Sanford Koppelman was born in Newark, New Jersey on May 16, 1937. He received a Bachelor of Science Degree in Civil Engineering from the Massachusetts Institute of Technology in 1959. He received a Master of Business Administration Degree with Distinction from Harvard University in 1961.

After two years of military service and eight years of professional experience, Koppelman returned to M.I.T. to undertake advanced study in the Transportation Systems Division of the Department of Civil Engineering. His doctoral thesis is entitled "Travel Prediction with Models of Individual Choice Behavior".

Koppelman's experience includes four years in the field of international marketing, plant location and investment.

He subsequently worked as a Transportation Engineer for the Tri-State Regional Planning Commission in the New York Metropolitan Area. In this position he was responsible for the development of a simplified system sensitive approach to the prediction of travel demand in major sub-regional areas. While with the Commission, Koppelman organized and directed a unit responsible for developing systematic analyses of the performance of large scale public service systems.