NATIONAL COOPERATIVE
HIGHWAY RESEARCH PROGRAM REPORT **253**

APPLICATION OF DISAGGREGATE TRAVEL DEMAND MODELS

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM REPORT **253**

APPLICATION OF DISAGGREGATE TRAVEL DEMAND MODELS

W.B. TYE, L. SHERMAN, M. KINNUCAN, D. NELSON, and T. TARDIFF
Charles River Associates Incorporated
Boston, Massachusetts

RESEARCH SPONSORED BY THE AMERICAN ASSOCIATION OF STATE HIGHWAY AND TRANSPORTATION OFFICIALS IN COOPERATION WITH THE FEDERAL HIGHWAY ADMINISTRATION

AREAS OF INTEREST:

FORECASTING
(HIGHWAY TRANSPORTATION)
(PUBLIC TRANSIT)

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. DECEMBER 1982

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

Systematic, well-designed research provides the most effective approach to the solution of many problems facing highway administrators and engineers. Often, highway problems are of local interest and can best be studied by highway departments individually or in cooperation with their state universities and others. However, the accelerating growth of highway transportation develops increasingly complex problems of wide interest to highway authorities. These problems are best studied through a coordinated program of cooperative research.

In recognition of these needs, the highway administrators of the American Association of State Highway and Transportation Officials initiated in 1962 an objective national highway research program employing modern scientific techniques. This program is supported on a continuing basis by funds from participating member states of the Association and it receives the full cooperation and support of the Federal Highway Administration, United States Department of Transportation.

The Transportation Research Board of the National Research Council was requested by the Association to administer the research program because of the Board's recognized objectivity and understanding of modern research practices. The Board is uniquely suited for this purpose as: it maintains an extensive committee structure from which authorities on any highway transportation subject may be drawn; it possesses avenues of communications and cooperation with federal, state, and local governmental agencies, universities, and industry; its relationship to its parent organization, the National Academy of Sciences, a private, nonprofit institution, is an insurance of objectivity; it maintains a full-time research correlation staff of specialists in highway transportation matters to bring the findings of research directly to those who are in a position to use them.

The program is developed on the basis of research needs identified by chief administrators of the highway and transportation departments and by committees of AASHTO. Each year, specific areas of research needs to be included in the program are proposed to the Academy and the Board by the American Association of State Highway and Transportation Officials. Research projects to fulfill these needs are defined by the Board, and qualified research agencies are selected from those that have submitted proposals. Administration and surveillance of research contracts are the responsibilities of the Academy and its Transportation Research Board.

The needs for highway research are many, and the National Cooperative Highway Research Program can make significant contributions to the solution of highway transportation problems of mutual concern to many responsible groups. The program, however, is intended to complement rather than to substitute for or duplicate other highway research programs.

NCHRP REPORT 253

Project 8-13(2) FY '77
ISSN 0077-5614
ISBN 0-309-03423-x
L. C. Catalog Card No. 82-74118

Price: $12.40

NOTICE

The project that is the subject of this report was a part of the National Cooperative Highway Research Program conducted by the Transportation Research Board with the approval of the Governing Board of the National Research Council, acting in behalf of the National Academy of Sciences. Such approval reflects the Governing Board's judgment that the program concerned is of national importance and appropriate with respect to both the purposes and resources of the National Research Council.

The members of the technical committee selected to monitor this project and to review this report were chosen for recognized scholarly competence and with due consideration for the balance of disciplines appropriate to the project. The opinions and conclusions expressed or implied are those of the research agency that performed the research, and, while they have been accepted as appropriate by the technical committee, they are not necessarily those of the Transportation Research Board, the National Research Council, the National Academy of Sciences, or the program sponsors.

Each report is reviewed and processed according to procedures established and monitored by the Report Review Committee of the National Academy of Sciences. Distribution of the report is approved by the President of the Academy upon satisfactory completion of the review process.

The National Research Council was established by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and of advising the Federal Government. The Council operates in accordance with general policies determined by the Academy under the authority of its congressional charter of 1863, which establishes the Academy as a private, nonprofit, self-governing membership corporation. The Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in the conduct of their services to the government, the public, and the scientific and engineering communities. It is administered jointly by both Academies and the Institute of Medicine. The National Academy of Engineering and the Institute of Medicine were established in 1964 and 1970, respectively, under the charter of the National Academy of Sciences.

The Transportation Research Board evolved from the 54-year-old Highway Research Board. The TRB incorporates all former HRB activities and also performs additional functions under a broader scope involving all modes of transportation and the interactions of transportation with society.

Special Notice

The Transportation Research Board, the National Academy of Sciences, the Federal Highway Administration, the American Association of State Highway and Transportation Officials, and the individual states participating in the National Cooperative Highway Research Program do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of this report.

Published reports of the

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

are available from:

Transportation Research Board
National Academy of Sciences
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

Printed in the United States of America.

FOREWORD

By Staff
Transportation
Research Board

This report will be of interest to transportation analysts and policy planners in Federal Government, in state transportation departments, and in metropolitan planning organizations who have expertise in conventional urban travel demand analysis. Such persons will find guidance in using disaggregate travel demand forecasting models for policy analysis. These models are useful to the analyst studying alternative transportation modes, new technology, transit fare and service changes, traffic engineering improvements, and toll policies. Examples of models for work trip mode choice, and shopping trip mode choice, are provided for practitioners. Researchers will find in the appendixes exhaustive discussions on model development, the independence of irrelevant alternatives property, and a multinomial logit model that permits variations in taste across individuals.

The overall objective of this research was to develop operational travel demand forecasting models consistent with travel choice behavior and with coefficients estimated by use of data at the level of households or individual travelers. Such models are disaggregate models by definition. Early in the study it was determined that the three most critical issues in applying disaggregate models were: (1) whether the resulting models are transferable, (2) how to aggregate the model results for forecasting, and (3) whether the Independence of Irrelevant Alternatives property was an impediment to application of the resulting models. Findings on these critical issues together with a description of the state of the art in application of disaggregate models are given in this report. Under certain conditions and problem requirements (e.g., accuracy and area scale), a disaggregate model developed for one geographic area may be transferred to another area, but some adjustments to the model (e.g., recalibrating mode-specific constants), so that the model predictions match observed aggregate model splits, are necessary in most instances. The authors suggest that aggregation is best handled by segmenting travelers into similar groups on the basis of factors important to explaining travel behavior and its variability. The model would be applied to each segment and aggregated. With regard to the independence of the irrelevant alternative assumption, the authors state that the assumption can be a desirable and reasonable one for homogeneous traveler segments. For these issues, guidance is provided analysts in developing an approach to using disaggregate models for policy analysis.

The research was accomplished in three phases. Results for Plan I and Plan II are summarized in this report. Details are provided in (1) Disaggregate Travel Demand Models, Project 8-13: Phase I Report, Volumes 1 and 2 (February 1976), and (2) Disaggregate Travel Demand Models, Project 8-13(2): Phase II Report (May 1978). These reports are available on microfiche for $4.50 prepaid from the Transportation Research Board Publications Office, 2101 Constitution Avenue, N.W., Washington, DC 20418.

CONTENTS

ACKNOWLEDGMENTS

The research reported herein was performed under NCHRP Project 8-13 (2) by Charles River Associates Incorporated. Timothy Tardiff was the final principal investigator. He was preceded by William B. Tye, Vice President, CRA, who was the initial principal investigator. Other authors of the report were Michael Kinnucan, David Nelson, and Leonard Sherman. Other individuals who contributed significantly to the report were: Katherine O'Neil, Alexander Bardow, Harrison Campbell, Gerald Kraft, and Thomas Parody, all of Charles River Associates Incorporated. Special thanks are also extended to those who served the project as consultants in the areas of their special experience, including: Professor Daniel McFadden, Professor of Economics at the Massachusetts Institute of Technology; Professor Jerry Hausman, Department of Economics, Massachusetts Institute of Technology; Professor Robert Paaswell, Department of Civil Engineering, State University of New York at Buffalo; and Professor Peter Watson, Department of Economics, Northwestern University and Research Associate at the Transportation Center.

Michael Kinnucan, a young and brilliant contributor to the profession, died unexpectedly during this project at the age of 26. His assistance in the second phase of the research project was invaluable. This final report is dedicated to his memory by his associates at CRA.

APPLICATION OF DISAGGREGATE TRAVEL DEMAND MODELS

SUMMARY

Disaggregate models of urban area travel demand are by definition constructed using data at the level of the individual traveler or household. In contrast, aggregated data, such as zonal averages, have been commonly used in urban travel demand studies. Research in NCHRP Project 8-13 indicates that because disaggregate travel demand models are feasible methods to analyze many urban transportation planning alternatives and policy issues, they deserve wider application. They are especially relevant to the analysis of policy issues not satisfactorily considered with existing approaches. Examples include (1) forecasting the demand for a new mode, (2) analyzing the effects of transit fare and service changes, (3) determining the effect of alternative air quality control and energy conservation policies, (4) evaluating the impact of traffic engineering improvements, and (5) toll policies on the use of roads and other "low capital" policy issues.

The research described in this report demonstrates that disaggregate travel demand approaches offer considerable advantages over conventional approaches in many applications. Specific advantages include (1) reduced data costs; (2) improved ability to predict the effects of public policy on travel demand; (3) flexibility to meet different problems, needs, and response times; and (4) potential for improved transferability of model estimation results from one geographic area to another. Of course, in any particular application, the full extent of all of these benefits may not be realized.

The disaggregate approach to analysis of travel demand is a general framework for travel behavior analysis that can be used to meet the data available, the desired results, and the capabilities of the user. It is not a formula or standardized approach. No one disaggregate model should be specified for all situations, although standardized methodologies will evolve to meet analysts' needs for common problems.

The disaggregate method of analysis indicates that changes in the travel choice environment affect different market segments in substantially differing ways. Therefore, segregating the market into groups or segments characterized by homogeneous tastes and socioeconomic attributes is essential to good forecasting. Disaggregate approaches are an ideal analysis tool to account for differences in response among market segments. This study developed models of work and nonwork trip mode choice behavior.

This report incorporates much of the research findings of the earlier Phase I and Phase II project reports (*1*, *2*). These earlier reports should be consulted by the practitioner of disaggregate modeling because the findings of these lengthy reports are only summarized in the present report.

Early in the study it was determined that the three most critical issues in applying disaggregate models were (1) whether the resulting models are transferable, (2) how to aggregate the model results for forecasting, and (3) whether the Independence of Irrelevant Alternatives property was an impediment to application of the resulting models.

Findings on Transferability of Disaggregate Models

The transferability of disaggregate models from one group of individuals to another (e.g., from one geographic area to another) is one of the major potential advances over aggregate methods which disaggregate models may incorporate. The proposition that disaggregate models will be transferable is based conceptually on the fact that they attempt to capture the behavioral regularities of individuals. Evidence supporting transferability would be a major contribution to improved understanding of the usefulness of disaggregate demand models.

Despite the early optimism on transferability of disaggregate models, more recent evidence suggests that, at a minimum, adjustments to the models must be made before transferring the models from one geographic area to another. Furthermore, different models calibrated on different data have produced behavioral parameter estimates that are not consistent. However, these differences may be explained by factors other than inherent behavioral differences among people (different model specifications, variable definitions, etc.). Therefore, the evidence does not necessarily support the conclusion that a behavioral model is not transferable.

The wisdom of transferring an existing model depends greatly on the costs of making forecast errors. If a high level of accuracy is desired, a sample size of possibly as large as several hundred observations is highly recommended to test the reasonableness of transferring an existing model. If less accuracy is required, a forecast based on an existing model with recalibration of the ''mode-specific constant'' may well be acceptable. Transferability is most likely to be valid when transferring the model to a group of people choosing among a set of alternatives identical to the calibration data set. Furthermore, it must be remembered that while transferring a model saves on the cost of collecting new data to calibrate a new model, it places greater demands on the practitioner to understand fully the assumptions made when applying the model.

Findings on Aggregation of Disaggregate Models

The aggregation issue is closely related to the issue of the transferability of disaggregate models. One reason that aggregate demand models may not be transferable is that a group of individuals, on average, may not behave the same as an individual with average levels of the explanatory variables. Identical group averages of explanatory attributes may obscure substantially different distributions of attributes across individuals for two groups. Attempting to forecast on the basis of group means under these circumstances may lead to substantial errors (the ''aggregation error''). Calibrating a disaggregate model using data on individual behavior presumably would substantially reduce the data requirements and would also reduce the ''aggregation error'' in forecasting. However, to apply the model calibrated at the level of individuals, some procedure is required to aggregate forecasts to predict group behavior. Unless a workable procedure can be developed to conduct the aggregation process, the efficiency in use of data and accuracy of calibration of the disaggregate models may be dissipated in computationally burdensome and ''data-hungry'' aggregation procedures.

The findings on aggregation may be summarized as follows.

1. Aggregation bias in transportation forecasting is the error arising when a group of individuals is assumed to be responding to the same (aggregate or average) level of observed service or socioeconomic attributes when they are, in

fact, heterogeneous with respect to observed level-of-service and socioeconomic attributes.

2. Aggregation bias arises if attempts are made to use aggregate data to make predictions with (nonlinear) disaggregate models. The simplest method of making predictions of group behavior with disaggregate models is to make a single prediction using the group average values of the independent variables. However, the predictions made by this method will be erroneous if the model is nonlinear because in statistical terms, "the average of a nonlinear function is not equal to the function evaluated at the averages of the independent variables." The prediction error arising from the use of this method is a type of aggregation bias. The forecasting approach which employs group means as independent variables has been called the "naive method," or the "direct aggregation method."

3. Aggregation error also generally occurs if an aggregate model calibrated on one group is applied to another group composed of individuals whose decisions respond to changes in the levels of the observed attributes (the independent variables) in a manner different from the group in the calibration sample.

4. The most reliable method of making predictions with disaggregate models is to use the values for each individual in the forecasting model as independent variables and make the prediction over all individuals. This method has been referred to as the "enumeration method."

For many applications the use of market segmentation is recommended as an aggregation technique.

The findings presented in Appendix C make use of local transportation data and U.S. Census and other nationally available data to construct "market segments." In some respects, the market segmentation approach to applying disaggregate models is very similar to the familiar "cross-classification" method in trip generation forecasting. Both methods involve the construction of tables (cross-tabulations) where values of the variables thought to determine the relevant travel behavior are divided into categories. The intersection of a single category from one variable with a category from a second variable (and possibly with a category from a third variable, a fourth variable, etc.) defines a "cell" or market segment. For example, in this study a work mode choice model was applied to 12 market segments defined in terms of income (under $15,000 and over $15,000), distance from transit (under 3 blocks, 3 to 6 blocks, over 6 blocks), and autos per driver (less than 1, 1 or more). The purpose of the market segmentation approach differs from that of the cross-classification method, however. In the latter method, the number of trips per household per day, for example, is assumed to be correlated with the variables defining the cells in some unspecified fashion. That is, households falling in any cell are predicted to make a certain number of trips, with no explicit model that relates tripmaking behavior to the "explanatory" variables. In the market segmentation approach, on the other hand, the behavioral model is quite explicit and the cells are nothing more than a convenient way of disaggregating the population so that the model may be applied.

The market segmentation approach to aggregation is conceptually the simplest of the aggregation procedures that have been proposed for disaggregate models and is certainly the one that is most familiar to the transportation planner. Market segmentation hopes to minimize the aggregation error by constructing relatively homogeneous cells of persons reasonably similar in the observed attributes that determine travel choice (and presumably predicted behavior). Forecasting is accomplished by applying the model to each cell separately to predict behavior. The relative frequencies of the cells for the market segments are then used to weight the cells to predict aggregate behavior.

The market segmentation approach has been shown in many circumstances to eliminate much of the aggregation error. It is computationally simple (as long as the number of choice alternatives and explanatory variables do not become large), it can frequently rely on existing data, and it is similar to a technique already familiar to transportation planners. Although others have endorsed alternative aggregation approaches and the issue is by no means settled, the research in NCHRP Project 8-13 confirms that market segmentation is a feasible and reasonably accurate approach for many practical planning applications.

This report develops procedures using existing data sources to define the market segments and determine cell frequencies for use with existing models. Because existing data sources seldom report fully on the joint distributions of all explanatory variables in the model, this report develops an adaptive procedure. The approach uses local data on the relative frequency of values for separate explanatory variables (marginals) and data on joint distributions from other sources to estimate the joint cell frequencies in the forecast population. This procedure is known as ''marginal weighting.''

The conclusions of this research regarding the market segmentation technique are summarized as follows:

1. Aggregation of predictions over market segments is performed by weighting the prediction for any particular market segment by the probability that an observation will fall in that segment, and then summing the weighted predictions.

2. Objectives in defining a market segmentation scheme for demand forecasting are: (a) to achieve a grouping where relatively similar values of the independent variables are grouped together; and (b) to minimize the number of market segments that can be feasibly manipulated for quick-reaction policy analysis. The first objective suggests that the market segments be defined by a cross-classification of the independent variables and the second objective suggests that only a very limited number of categories (say, between two and four) be created for each variable.

3. Market segmentation is not the only technique for applying disaggregate models, but it offers three attractive features for manual forecasting and sketch planning: it is intuitive to the policy maker, it facilitates the use of a wide range of data sources, and it can be accomplished manually without resort to analysis at the level of the entire transportation system.

The steps in applying the market segmentation technique to forecasting may be summarized as follows:

1. Select a disaggregate model to predict policy impacts or to forecast travel demand for planning purposes.

2. Determine how the data should be grouped into market segments. A particular grouping (or market segmentation) is defined both by the selection of which variables to segment and the determination of the number of categories to be created for each variable.

- The bases for deciding whether to segment on a variable are that variable's importance in explaining travel behavior and its variability.
- In determining the number of categories to be created for each of the classifying variables, a tradeoff exists in that the greater the number of categories, the greater the reduction in aggregation bias but also the more unwieldy the market segmentation technique becomes for a quick-response policy evaluation. A compromise must be struck between these two conflicting objectives.

3. For each market segment defined by the foregoing steps, obtain data on the probability that a trip will belong to that market segment (i.e., the proportion of all trips that fall into that market segment) and obtain data on the averages of the independent variables for all of the trips in that segment.

4. If the model is not calibrated on data from the population whose behavior is to be forecast, data on the actual choice shares within each of the market segments should be obtained if possible. These data may be used to adjust (or "update") the model to enhance its transferability to the current planning context.

5. Use the average value of the independent variables to compute the baseline travel choice shares for each segment as estimated by the (updated) model.

6. Weight the estimated choice shares for each market segment's proportion of all trips and combine these weighted choice shares to obtain the estimated aggregate baseline choice shares.

7. Adjust the average values of the independent variables (in each of the market segments) to reflect implementation of the policy or conditions at the end of the planning horizon. Use these to forecast the post-policy choice shares.

8. Weight and combine the post-policy choice shares for the various market segments as described for computing aggregate baseline choice shares to obtain aggregate post-policy choice shares.

The Independence of Irrelevant Alternatives Property

The multinomial logit (MNL) model predicts the probabilities of choice for an individual who confronts two or more alternatives. The Independence of Irrelevant Alternatives (IIA) property states that if, for example, two modes are available and a new mode is introduced, the ratio of the probabilities of the two old modes will be unchanged regardless of the probability of choice for a new mode. Applied to a group, the IIA property states that a new mode will capture an identical percentage of the market shares of the two existing modes. An extensive analysis of the IIA property of the multinomial logit model of travel demand has been conducted in the NCHRP Project 8-13 study.

An example of what troubled many critics of the IIA property is the classic "blue bus/red bus" case, i.e., the "new mode problem." Suppose, for example, that the blue bus and auto mode capture 50 percent of a given travel market. Assume that a new bus mode is introduced with exactly the same service attributes as the old bus mode except that the bus is painted a different color, red (to which the patrons are indifferent). It is expected that the true modal shares will now be ½, ¼, ¼ for auto, the first bus alternative, and the second bus alternative respectively. However, the ordinary multinomial logit (MNL) model will forecast that each of the three models captures one-third of the market, which is clearly a poor forecast.

The major conclusions of the analysis of the IIA property are:

1. The IIA property is not an inherent drawback to disaggregate demand modeling and is not presently an impediment to implementation of disaggregate demand models.

2. The IIA assumption may be reasonable or unreasonable, depending on the circumstances of the particular application. Therefore, diagnostic tests to determine whether the assumption is valid in a particular application have been designed as part of this study.

3. When the IIA assumption is unreasonable, the multinomial logit model

cannot be applied without error. The error may be large or small depending on the circumstances.

4. If the IIA assumption is invalid, corrective measures have been identified as part of this study to take the dependence into account.

Clearly, the IIA property is implausible if applied to any random group of decision-makers and mode alternatives. However, as discussed extensively in Appendix E, the MNL model itself applies only to a reasonably "homogeneous" market segment facing similar choice, i.e., there is an aggregation error if the simple MNL model is applied indiscriminately to forecast the behavior of a group composed of heterogeneous market segments. For homogeneous market segments where the MNL model is valid, the IIA will frequently be a reasonable and desirable assumption. The IIA assumption is common to most "share" models, aggregate or disaggregate (*3*).

Evaluation of Disaggregate Demand Models

Evaluation of the models calibrated in this study indicates the following conclusions: (1) they are consistent with prior intuitive expectations of travelers' preferences and economic theory of rational behavior; (2) statistically reliable disaggregate models can be estimated using data that are commonly available for urbanized areas, although extensive augmentation may be necessary for calibration of models for some choice alternatives; (3) the accuracy of the estimated models is reasonably high; (4) logit models of disaggregate travel choice can be calibrated using readily available user-oriented computer programs; (5) limited tests (conducted as part of this study and elsewhere) suggest that at least some parts of disaggregate demand models can be transferred from one area to another, either directly or after adjustment for known between-region differences.

CHAPTER ONE

INTRODUCTION AND RESEARCH APPROACH

Disaggregate travel demand models have been developed in recent years to respond to the transportation planner's needs for better techniques of demand analysis to address policy issues as they are understood today. The traditional methods of demand analysis that have widespread acceptance in the planning community have been found to be insensitive to important policy issues, e.g., energy and new technology (new mode) issues, and inappropriate for a small scale of planning, such as at the project and corridor levels. They are also too time-consuming to provide "quick response" answers in the short time frequently available for planning studies. Evidence on the development and application of disaggregate demand models in this study indicates that they are a workable approach to analyzing the new issues and they also offer potential for directly improving the conventional approaches.

NCHRP Project 8-13 was designed to develop, calibrate and evaluate the new disaggregate models and to recommend major conceptual and other improvements. Evaluation identified serious barriers to their use requiring a program of research. In the first phase of the study, models were developed and calibrated using existing urban transportation survey data, suitably augmented, to describe the travel choice environment (*1*). Although application of the model was not contemplated as part of the first phase of the study, exercises to illustrate their application and to identify possibly impediments to further application were conducted.

Because the first two phases of the research were limited to advancing model development that could be made without new sources of data, the research concentrated on model development and calibration with existing data. Thus in the first phase, research concentrated on improving the model of worktrip mode choice and testing its transferability between geographic areas; testing alternative specifications of model structure, especially for discretionary trips; improving the model's ability to incorporate more than two modes, with special emphasis on the Independence of Irrelevant Alternative property of the multinomial logit model; and the design of data collection programs to calibrate and apply the models.

Phase II (*2*) focused on eliminating impediments to the implementation of disaggregate models. A decision was made to address the Independence of Irrelevant Alternatives, aggregation, and transferability of logit models. Developments in trip chaining, attitudinal modeling, household-individual interactions, network equilibration, nonutility-maximizing models, time budgets, and other related topics were reviewed for integration into the disaggregate modeling approach, but were not the object of an independent development effort. This decision was based primarily on the fact that more conventional models, such as probit and logit, have a demonstrated record of achievement in actual transportation planning. Researchers interested in other approaches, e.g., "threshold" and "sequential" models, should consult references such as Gensch and Svestka (*4*).

The main text of this report is oriented to the general reader with an acquaintance of the issues currently facing the transportation analyst, but without experience in, or extensive familiarity with, the new disaggregate approaches. Extensive appendixes have been prepared on the technical details of the research, and they are directed to the expert who already has some knowledge of disaggregate modeling or desires to acquire a detailed working knowledge of some major issues in disaggregate demand modeling. Users who would like to learn more about urban travel forecasting should review the literature, especially the documents provided by the U.S. Department of Transportation (*5*). Rather than reproduce this voluminous literature, the present study assumes a familiarity with the fundamentals of conventional travel demand forecasting.

The vast bulk of the literature on disaggregate travel demand modeling is oriented toward the research community. As a result, greater and greater progress has been achieved in perfecting the techniques of disaggregate models, but at the expense of communicating with the practicing transportation professional. One of the objectives of this project is to provide an "entry point" for the practitioner.

In providing an "entry point" for transportation planners interested in applying disaggregate models, the report attempts to identify particularly salient recent advances in modeling by other researchers and to provide a discussion of advances that were specifically made as part of the research on this project. This document is not intended to be a complete guide to all elements of disaggregate modeling. In many cases the reader is referred to other references that address an issue in greater detail. The serious practitioner of disaggregate modeling is also strongly urged to review the reports of the Urban Travel Demand Forecasting Project, University of California at Berkeley (*6, 7*).

This project was undertaken in connection with a number of other simultaneous research efforts. NCHRP Project 8-12 produced two reports: *Travel Estimation Procedures for Quick Response to Urban Policy Issues* (*8*) and *Quick Response Urban Travel Estimation Techniques and Transferable Parameters: User's Guide* (*9*).

NCHRP 8-14 was designed to review and develop new procedures for understanding travel behavior. Phase I of NCHRP 8-14 developed a framework for exploring travel behavior based on the concepts of role and activity choices (*10*). Phase II (*11*) incorporated the concepts identified in Phase I into models of trip generation and activity time allocation. The study also included guidelines for applying the improved trip generation procedures for practical forecasting.

For the Federal Highway Administration, Talvitie et al. (*12*) explored a range of issues relevant to the development and application of disaggregate travel forecasting models. As such, their work is complementary to this project and should be useful for researchers and practioners interested in disaggregate travel demand models.

A fourth concurrent effort was the collection of a large new disaggregate data set from the Baltimore region. At the initiation of Project 8-13, it was contemplated that development of models using this data set would be a major element of the research plan. However, the data collection was beset by lengthy delays. As a result, the project emphasized the resolution of impediments to model implementation that could be accomplished without the new data. At the very close of the project, the data set became available and limited research was conducted with the remaining budget.

The final report of this project represents an abridgement of certain parts of the two interim reports. Readers who anticipate a substantial commitment to the use of disaggregate models should use the present volume in conjunction with the two interim reports. For example, the Phase II report contains an extensive example illustrating the market segmentation method of aggregation which is only summarized in the final report.

As a newly emerging discipline, disaggregate demand modeling suffers from a lack of consistent terminology and notation. To some extent this inconsistency also plagues this report because different parts of the report address different issues and are in some cases addressed to different audiences, calling for a different model specification and notation. The procedure used here is to attempt a clear definition of terminology in each section and to maintain that terminology within the section.

The report proceeds as follows. First, Chapter Two addresses the principal findings on disaggregate models, with emphasis on an exposition of the disaggregate approach; specification issues in mode choice, trip distribution, and trip generation; data collection; aggregation; transferability; and detection and correction of errors in the logit model. Chapter Three contains an appraisal of the results. Chapter Four presents the conclusions and suggestions for future research. Appendixes contain a review of the literature on aggregation, guidelines for applying market segmentation, an exposition of the models, and an elaborate discussion of the IIA property. Readers primarily interested in practical applications are directed primarily to Chapters Three and Four. The more technically oriented reader may want to emphasize Chapter Two and the Appendixes.

CHAPTER TWO

FINDINGS—THE STATE OF THE ART IN APPLICATION OF DISAGGREGATE MODELS

This chapter summarizes the status of the major issues a transportation planner will face when estimating and applying disaggregate models. The exposition proceeds first with the development of exactly what the practitioner is assuming when applying a disaggregate model to mode choice. The following sections describe model specification and estimation; statistical significance tests; models of worktrip and shopping trip mode and destination; disaggregate approach to trip generation modeling; and diagnosis and correction of errors in the logit model. Separate sections summarize findings on data collection and aggregation of disaggregate forecasts.

WHAT IS THE DISAGGREGATE APPROACH?

Appendix E develops, in detail, the behavioral assumptions incorporated in the disaggregate multinomial logit (MNL) model. In this section the fundamental assumptions are identified in a less rigorous, but more intuitive fashion as an introduction. In addition, the discussion covers the advantages of the disaggregate approach, evaluation with disaggregate models, elasticity formulation, and the Independence of Irrelevant Alternatives property. Before introducing the disaggregate approach, a brief comparison to conventional aggregate models is provided.

Comparison to Conventional Aggregate Models

Conventional travel demand models typically involve a four-step sequence: (1) trip generation (travel frequency); (2) trip distribution (where trips go); (3) modal split; and (4) route assignment. They are often called aggregate models because they explain the travel of a group of households or individuals, e.g., all households in a traffic analysis zone. Further, aggregate data are used in estimating the models. For example, the average household trip frequency for a zone may be a function of the average household size and average auto ownership levels in the zone. Note that although data are usually available at the household (disaggregate) level, they are averaged before estimating aggregate models.

In contrast, disaggregate models explain the travel of individuals or households directly. Therefore, data are used at the disaggregate level at which they are collected, rather than averaged into large aggregates.

In the following an overview of the disaggregate approach is provided.

Fundamentals

The logit model of individual choice behavior has been the most prominent methodology used in disaggregate travel demand models. The logit model assumes that each individual makes selections from among a set of alternatives, often referred to as the choice set. From that set he chooses the alternative he prefers. In making the selection, he assigns a utility value to each alternative. The utility of an alternative is a measure of the order of preference (e.g., if one alternative is more preferred, it will be assigned a higher utility). For modeling purposes the utility is composed of two components, a component based on observed attributes, often called the "representative utility," and an unobserved component, called the random utility component. The term "random utility model" is derived from the assumption that although the individual's choice is rational, an observer cannot predict a given individual's choice because of the influence of unobserved determinants of choice as reflected in the random component.

Mathematically, assume that each tripmaker assigns some utility to each of his travel alternatives. Let U_{it} be the utility of the i^{th} alternative for the t^{th} tripmaker. Further assume that each utility value can be partitioned into two components, a systematic component, or "representative utility," V_{it}, and a random component, ϵ_{it}, such that,

$$U_{it} = V_{it} + \epsilon_{it} \tag{1}$$

The systematic component V_{it} is that part of utility contributed by factors that can be observed and measured (the "representative utility") and the random component ϵ_{it} is the utility contributed by unobserved factors.

Tripmakers are assumed to choose the travel alternative that yields the highest utility. Thus, individual t will choose alternative i over alternative j if

$$U_{it} > U_{jt} \tag{2}$$

From Eqs. 1 and 2 it is clear that the alternative i is chosen if

$$V_{it} + \epsilon_{it} > V_{jt} + \epsilon_{jt} \tag{3}$$

or, equivalently, if

$$V_{it} - V_{jt} > \epsilon_{jt} - \epsilon_{it} \tag{4}$$

One cannot predict with certainty which alternative an individual will choose, i or j, because although V_{it} and V_{jt} can be estimated and compared, it cannot be determined with certainty if $(V_{it} - V_{jt})$ exceeds $(\epsilon_{jt} - \epsilon_{it})$. Instead, one seeks to determine the probability with which $(\epsilon_{jt} - \epsilon_{it})$ will be less than $(V_{it} - V_{jt})$. This is generally done by assuming that the ϵ_{it}'s are independently and identically distributed with the Weibull distribution. (See CRA (*13*), McFadden (*14*).) Based on these additional assumptions, it can then be shown that the probability that the t^{th} individual will choose the i^{th} alternative is given by:

$$P_t(i) = \frac{e^{V_{it}}}{\sum_{j=1}^{J} e^{V_{jt}}} \qquad (5)$$

in which

J = the number of alternatives (including the i^{th} alternative); and
e = the base of the natural logarithm.

Equation 5 is the well-known logit model.

It was stated above that V_{it} is the component of utility contributed by observed attributes. It is computationally convenient to assume that V_{it} is a linear combination of the observed attributes of the alternative i and individual t:

$$V_{it} = \sum_{k=1}^{K} X_{itk}\beta_k + \sum_{l=1}^{L} S_{tl}\alpha_l \qquad (6)$$

in which

X_{itk} = value of the k^{th} attribute of alternative i for the t^{th} individual;
K = total number of attributes of the alternatives;
β_k = parameter of the k^{th} attribute;
S_{tl} = l^{th} socioeconomic characteristic of individual t;
L = total number of socioeconomic characteristics; and
α_l = parameter of the l^{th} socioeconomic characteristic.

The logit model cannot be calibrated if any of the independent variables take on the same value across all alternatives for all individuals. Thus, attributes of the individual (such as income) that do not vary across alternatives enter the model by becoming part of the representative utility of only a subset of the alternatives. In the binary auto-bus case, for instance, income would be only part of the auto utility function. The income variable in the bus utility function is implicitly equal to zero. The decision to enter income into the auto utility function is arbitrary. If income were entered into the bus utility function, its coefficient would have the same absolute value but would have the opposite sign. Alternatively, attributes such as income can be part of each utility function through the use of interaction terms. For example, the travel costs of each mode are divided by income in several existing models.

Example

A specific example of Eqs. 5 and 6 is useful. Later in the chapter a work mode choice model, defined for two modes, auto driver and transit, is presented, with the following observed component of modal utility:

$$V_{it} = AD\,(-5.72 + 1.38\,HINC + 4.07\,APERW) - 0.117\,OVTT_i - 0.0348\,IVTT_i - 9.06\,C/INC_i \qquad (7)$$

in which

AD = a dummy variable which equals 1 for auto and 0 for transit;
$HINC$ = a measure of household income that equals 1 for incomes greater than $7,000 and 0 otherwise;
$APERW$ = the ratio of autos to workers in the household;
$OVTT$ = out-of-vehicle travel time, e.g. walking time, measured in minutes for the round trip;
$IVTT$ = in-vehicle travel time, measured in minutes for the round trip; and
C/INC = round-trip cost in dollars divided by the income code value presented in Appendix D.

Equation 7 illustrates several features of modal utility functions. First, $OVTT_i$, $IVTT_i$, and C/INC_i correspond to the X_{itk} of Eq. 6. Second, the socioeconomic characteristics, $HINC$ and $APERW$, only appear in the auto utility function (because AD = 1 for auto and 0 for transit). These variables correspond to the S_{tl} variables of Eq. 6. Third, income is also part of the interaction term C/INC. Finally, the constant term, −5.72, appears in the auto utility function and represents the effects of variables not included in the utility function, e.g., comfort of modes.

Table 1 presents hypothetical values for the variables in Eq. 7. Inserting these values in Eq. 7 yields:

$$\begin{aligned} V_{1t} &= 1\,[-5.72 + 1.38(1) + 4.07(1)] - 0.117(0) \\ &\quad - 0.0348(60) - 9.06(0.20) \\ &= -4.17 \end{aligned} \qquad (8)$$

$$\begin{aligned} V_{2t} &= -0.117(7) - 0.0348(110) - 9.06(0.10) \\ &= -5.553 \end{aligned} \qquad (9)$$

In this case mode 1 is auto driver and mode 2 is transit. Equation 5 is then used to estimate selection probabilities.

TABLE 1. CHARACTERISTICS OF WORK TRIP FOR HYPOTHETICAL INDIVIDUAL

	Socioeconomic Characteristics	Modal Attributes	
		Auto	Transit
Household Income	$10,000		
Autos	1		
Workers	1		
Round-trip Cost		$1.00	$.50
Round-trip In-vehicle Time		60 minutes	110 minutes
Round-trip Out-of-Vehicle Time		0	7 minutes
HINC	1		
INC*	5		
APERW	1		
INVTT		60	110
OVTT		0	7
C/INC		.20	.10

*Income code corresponding to $10,000 from Table D-15, Appendix D.

SOURCE: Charles River Associates, 1980.

$$P_t\,(\text{AUTO}) = \frac{e^{-4.17}}{e^{-4.17} + e^{-5.553}} \tag{10}$$
$$= 0.799$$

$$P_t\,(\text{TRANSIT}) = 1 - P_t\,(\text{AUTO}) \tag{11}$$
$$= 0.201$$

The selection probabilities indicate that an individual with socioeconomic and trip characteristics described in Table 1 would have a probability of about 80 percent of driving to work. Alternatively, about 80 out of 100 commuters with these characteristics would be expected to drive to work.

The logit model is often presented in the "log odds" format, where the log of the ratio of the probabilities of two alternatives can be expressed as a function of the difference in attribute levels of the alternatives:

$$\ln\left(\frac{P_{jt}}{P_{it}}\right) = \sum \beta_k (X_{jtk} - X_{itk}) \tag{12}$$

The conventional notation for the binary logit specification in odds ratio form may be derived by letting

$$P_1 = P\,(U_1 > U_2) \tag{13}$$

$$P_1 = \frac{1}{1+e^{-V}} \tag{14}$$

$$P_1 + P_1^{e^{-V}} = 1 \tag{15}$$

$$E^{-V} + \frac{1-P_1}{P_1} \tag{16}$$

$$\frac{P_1}{1-P_1} = e^V \tag{17}$$

$$\ln\left(\frac{P_1}{1-P_1}\right) = V = V_1 - V_2 = \beta_1(X_{11} - X_{21}) + \beta_2(X_{12} - X_{22}) + (\alpha_{11} - \alpha_{21})\,S_1 \tag{18}$$

This algebraic transformation of the logit equation is quite convenient. It has allowed researchers to simply estimate binary choice logit models using ordinary least squares regression packages since the dependent variable can be represented as the natural log of the ratio of selections between i and j for individual (or class) t, and the independent variables become the differences in values of attributes between alternatives i and j.

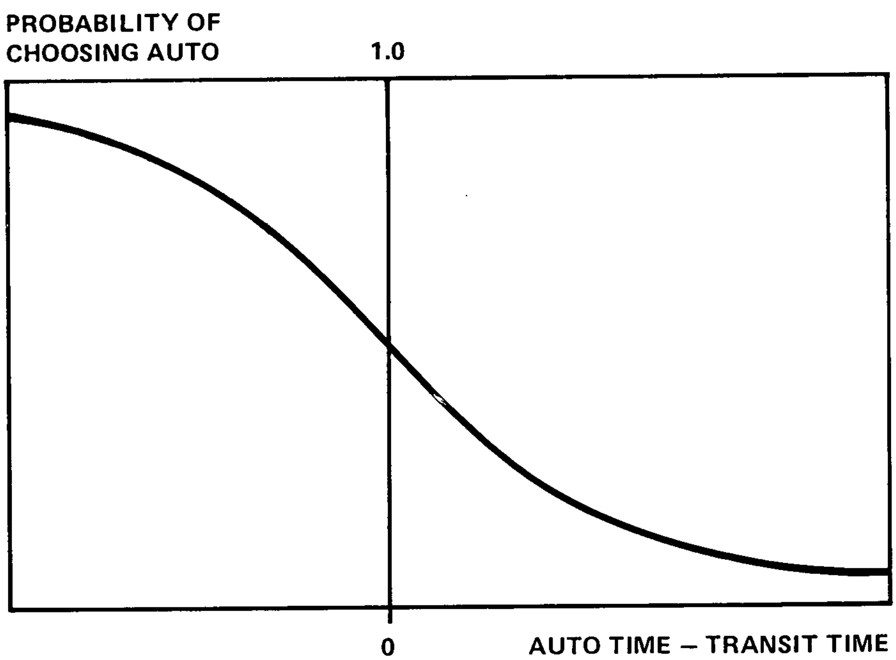
For negatively weighted attributes, such as time and cost, the probability of choosing an alternative will be a function of the attributes as shown in Figure 1. It should be noted that this plot is an S-shaped logistic curve.

Advantages of the Disaggregate Approach

Disaggregate approaches to understanding travel demand behavior are based on the assumption that since travel behavior originates with the decisions of the individual, improved understanding of the aggregate behavior of the population can be derived by improved understanding of the behavior of individuals. The trend toward disaggregation of demand is not limited to travel demand analysis—it is becoming standard practice in conventional economics, marketing, and many other fields.

The considerations that have prompted the use of disaggregate approaches in transportation demand analysis are the following.

1. *Economy of Data Collection*—Aggregation of data on individuals into group totals or averages, such as averages over travel zones or over metropolitan areas, loses the detailed information about the travel decisions of the individ-



SOURCE: Charles River Associates Incorporated, November 1980.

Figure 1. Response function assumed by a probabilistic choice model.

uals composing the groups. To calibrate models of group behavior, observations of many groups are required to obtain reliable estimation results. When the analysis is performed at the level of the individual, detailed information about his/her situation can be explicitly incorporated into the model and its estimation. Thus, with a given number of observations required for model calibration, many fewer individual observations are required when the data on individuals are not aggregated into groups. Furthermore, by avoiding the averaging or, equivalently, the aggregation process, the variability of the explanatory variables is much greater, making the estimation more reliable. Liou and Hartgen (*15*) found that disaggregate models permitted considerable savings in data costs as compared with conventional approaches. For these reasons, very substantial savings in data collection costs might be realized.

2. *Transferability*—Models that describe the behavior of aggregates of individuals are frequently not transferable from one group to another unless the size, composition, or other characteristics of the group are unchanged or controlled. Because models of individual behavior do not have this "aggregation problem," they are more likely to be transferable.

In most applications predictions of aggregates of individuals are necessary. In these cases the disaggregate models can be calibrated on data collected for the individual, and the level of aggregation (e.g., region, subregion, traffic analysis zone, corridor, etc.) can be taken into account in the analysis. The transferability property is particularly important in using the results of analysis in one area for predicting behavior in other geographic areas. Disaggregate approaches need not apply to an entire region, but can be used for subregions, corridors, or specific market segments. Models that are geographically transferable will also substantially reduce the cost of developing a new model to fit each particular situation.

3. *Policy Sensitivity*—Traditional aggregate demand models have not been sensitive to many public policy alternatives that affect travel behavior. The disaggregate detail on level of service and individual and household attributes can provide an improved understanding of the determinants of travel choices. Because disaggregate approaches are developed in terms of the behavior of the individual, the evaluation of public policy alternatives is enhanced. The disaggregate approach provides a natural framework for analyzing how a policy alternative affects the decision-making of the individual. If the policy effects are analyzed as they affect the individual, the transportation analyst's recommendations gain credibility because they are more intuitive. Moreover, disaggregate approaches are ideally suited to evaluating the impact of policies on different market segments or interest groups.

4. *Flexibility*—Disaggregate modeling is a method of analysis that is not a single model or a single "cookbook" approach. By the same token, it takes advantage of data and knowledge at hand and results of previous studies, whether the problem is long-range demand forecasting or short-range analysis of issues such as air quality and energy conservation alternatives.

Conventional urban transportation planning tools have been found satisfactory by many members of the planning community in meeting the needs for which conventional approaches were designed. However, a new generation of planning problems has emerged which requires improved knowledge of how public policy affects the use of existing facilities—fares and tolls, air quality control programs, energy conservation, exclusive bus lanes, and so forth. Disaggregate approaches can be designed to meet these new needs.

Evaluation with Disaggregate Choice Models

It is possible to evaluate the benefits resulting from a change in the transportation system with disaggregate travel choice models. The motivation for benefit measurement is similar to the approach used in standard cost-benefit analyses, i.e., the benefits derived from a policy change are a function of the differences of utilities with and without the policy.

Since it is assumed that individual utility is derived from the chosen alternative, but not the rejected ones, the benefit measurement involves the differences in maximum utilities (over the alternatives in the choice set) with and without the policy. However, the maximums cannot be observed because of the random component in the utility functions. Therefore, expected values must be used. Specifically, the benefit measure is

$$\text{Benefit} = \text{Expected Value } (Max\,_1U - Max\,_0U) \tag{19}$$

in which

$Max\,_1U$ = the maximum utility of the alternative with the policy change; and
$Max\,_0U$ = the maximum utility of the alternative without the policy change.

Ben-Akiva and Lerman (*16*) have shown that Eq. 19 has a very convenient form in the case of the multinomial logit model:

$$\text{Benefit} = \log \sum_{i=1}^{J_1} V_i - \log \sum_{i=1}^{J_0} V_i \tag{20}$$

That is, the benefit of a change of the transportation system is the difference in inclusive values with and without the change. (Inclusive value is discussed later in this chapter under destination choice modeling). Equation 20 can be used to evaluate the benefits resulting from changes in the choice sets available to individuals as well as changes in level-of-service variables such as travel times and costs. (Benefit measurement with the multinomial probit model, discussed later in this chapter, is also based on Eq. 19. The resulting expression for benefits is more complex computationally than Eq. 20 (the MNL benefit expression). Daganzo (*17*) gives a detailed discussion of the specifics.)

Elasticities

One use of disaggregate demand models has been to investigate the values and rates of substitution individuals use in making choice decisions. These values for specific observed attributes of alternatives are revealed by the utility parameters estimated in model calibration. Rates of substitution between attributes are implied by comparing estimated coefficients.

Among the more useful of these techniques for using logit models to analyze consumer values and develop quick response estimates of behavioral changes caused by policy shifts are elasticity approaches. A direct elasticity is the percent change in market share implied by a 1 percent change in an attribute of that mode. Similarly, a cross elasticity is the percent change in market share for alternative i implied by a 1 percent change in an attribute of another alternative j.

The elasticities of the probabilities with respect to each of the attributes in the respective utility (the "own" or "direct" elasticity) can be shown to be:

$$E^{P_i}_{X_{itk}} = (1 - P_i)\, \beta_k X_{itk} \tag{21}$$

in which

$E^{P_i}_{X_{itk}}$ = elasticity of P_i with respect to a X_{itk}, or the percent change in the value of the k^{th} attribute of alternative i for the t^{th} individual;
P_i = probability of selecting alternative i;
β_k = parameter of the k^{th} attribute; and
X_{itk} = value of the k^{th} attribute of alternative i for t^{th} individual.

The cross-elasticity of the i^{th} probability with respect to a change in the k^{th} attribute of the j^{th} alternative is:

$$E^{P_i}_{X_{jtk}} = -P_j\, \beta_k X_{jtk} \tag{22}$$

in which $-P_j$ = the negative probability of selecting alternative j.

It should be noted that this cross-elasticity is the same for all P_j, i.e., the cross-elasticities are identical for all alternatives. If a change in the attribute of one alternative improves the share of another by 10 percent, the same improvement occurs for a third alternative. This property is a manifestation of the Independence of Irrelevant Alternatives (IIA) property to be discussed below.

In joint or separable logit models (discussed below), the specification of direct- and cross-elasticities can be quite notationally complex. (See CRA (*2*), App. B, and Parody (*18*) for elasticity formulas.)

In applying the logit model results, elasticities must be interpreted cautiously. Equation 21 shows that the value of the level-of-service attribute heavily influences the elasticity. The value of the coefficient on walk time could be three times that of linehaul time, implying that the reduction in mode share resulting from a 1-min change in walk time would be three times that of an equivalent change in linehaul time. However, if linehaul time is ten times that of walk time, linehaul time will have the greater elasticity. Since elasticity is percentage change in market share divided by the percentage change in the attribute value, changes in attributes with relatively low values will tend to be big in percentage terms and have relatively low elasticities. Elasticity measures may leave much to be desired as a measure of the "importance" of an attribute in choice in disaggregate models.

The elasticity formula clearly shows that elasticities critically depend on the values of the attribute and probability where they are evaluated. For example, it is clear that as the value of an undesirable attribute approaches infinity, so does the elasticity. The property suggests that elasticities derived from nonlinear models are not generally transferable without properly accounting for the value of the attribute and mode share.

Independence of Irrelevant Alternatives

As a practical matter there are four key assumptions in arriving at the MNL model:

1. That individual behavior is random, as a result of unobserved determinants of behavior, but the relative shares of the choice alternatives can be predicted with the estimated model, based on the "representative utilities."
2. Within a group of individuals with identical observed attributes, there are no taste variations (stated differently, the model coefficients (β_k) are fixed and not random).
3. The random components of utility are independent across alternatives (the unobserved attributes of two alternatives vary independently and are as likely to be different as similar).
4. The random utility components and the attributes contributing to that utility are not correlated with the observed attributes.

Collectively, the last three assumptions produce the Independence of (or from) Irrelevant Alternatives property of the logit model. This property is the most controversial issue in disaggregate modeling. It may be demonstrated in several ways. For example, in the "log odds" ratio form (see Eq. 12), it is clear that the ratio of the share of two alternatives is not affected by the attributes of a third alternative. Consequently, if two alternatives have equal probabilities of being chosen in a two-way choice (e.g., $P_i = P_j = 0.50$), the introduction of a new "irrelevant" third alternative, k, with attributes identical to alternative j will cause all three alternatives to have equal market shares in a three alternative choice. This result is counterintuitive, since the new alternative, k, should only divert riders from the identical alternative j, producing shares of 50-25-25.

The most commonly used example of the property is a two-mode choice situation, auto and bus. Each has a 50 percent mode share. If a third irrelevant choice is added to the choice set by painting half the buses blue and the other half red, yielding a three-mode choice set, the MNL will illogically predict that the "new" mode will capture equal shares from auto and bus yielding a 33-33-33 mode split.

As a result of this example, the IIA problem is often referred to as the "blue bus/red bus" problem. The IIA property is also responsible for the fact that the cross-elasticity is the same for all modes in Eq. 22. Because of the IIA, if improvement of one mode causes a diversion of 10 percent of another mode's share, an identical 10 percent diversion must occur from a third mode if the ratio of the second and third modes' shares are to be independent of the first mode's attributes.

Most of the specification errors of logit models, and the increasingly sophisticated modeling alternatives designed to correct for these errors, result from a violation of the last three assumptions. Before turning to a discussion of these possible errors, it is helpful to first address specification issues that have arisen in mode choice analysis so that the reader may see how a model is developed in practice.

MODEL SPECIFICATION ISSUES FOR MODE CHOICE

A large amount of research has addressed the issue of appropriate specification of variables in disaggregate models. This section summarizes the most salient findings.

The specification of the logit model requires the identification of the X_{itk} and S_{tl} variables in Eq. 6. Once the variables are specified, a data set with the observations for each variable for individuals and their choices would permit statistical estimation of the model coefficients (the β_k's and α_l's) using one of the available statistical logit calibration packages.

McFadden (*7*) has conducted a survey of the variables that have been used in mode choice models and the conclusions are given in Table 2. Some of the more important variables are discussed below. This discussion is divided into two sections. First, the measurement and specification of socioeconomic characteristics are discussed. Also included under this heading is a treatment of the appropriate unit of analysis. Second, the measurement and specification of level-of-service variables (the attributes of alternatives) are discussed. This section also includes a discussion of attitude variables, alternative functional forms for explanatory variables, and the effects of specification errors.

TABLE 2. EXPLANATORY VARIABLES AFFECTING MODE CHOICE

Variables with critical explanatory power
Travel cost
On-vehicle time
Walk time
Transfer wait time
Transit initial headway
Number of persons in household who can drive
Determinants of alternative availability (e.g., ability to drive, auto required at work)
Wage
Variables with important explanatory power
Number of transfers
Respondent's relation to household head
Employment density at work location
Suburban or urban residence
Family composition
Variables with ambiguous explanatory power
Household income
Residential population density
CBD location of residence
Number of workers in household
Age of household head
Reliability of transportation mode
Perceptions of comfort, safety, convenience
Variables with low explanatory power
CBD work location
Sex of respondent
Age of respondent
Work status of household head
General attitudes toward privacy, delay, safety

SOURCE: McFadden (*7*)

Socioeconomic Characteristics

The most important socioeconomic characteristics affecting mode choice include financial considerations (e.g., income, wage, or wealth) and automobile availability (e.g., ownership and competition for the family car). Other relevant considerations in some choice settings can include employment type, lifecycle stage, age, and neighborhood.

Income and Wage Levels

Extensive research conducted as part of the Travel Demand Project at Berkeley indicates that after-tax wage is preferred over income as an indicator of the effect of financial considerations in mode choice (see McFadden (*19*)). Further, research by Train and McFadden (*20*) indicates that an acceptable specification is "cost ÷ wage," which has the effect of linearly relating the value of time to the wage rate. Their research indicated that this specification resulted in a somewhat better fit than did a specification where time was multiplied by the wage rate (which would have the effect of converting time into a money equivalent). However, the goodness of fit of the models differed only slightly, suggesting that the choice between specifications is essentially arbitrary.

Despite the superior conceptual appeal of the (after tax) wage rate (at least for worktrip mode choice), only family income is reported in many data sets; therefore, it must be used. Researchers also should be cautioned that the quality of the data of the income variable is often suspected to be poor. Many respondents give wrong answers to income questions or skip them on surveys. Finally, there is the issue of whether auto ownership, which is colinear with wage and income, is the true underlying determining factor in mode choice. In any event, financial considerations are theoretically important to individual travel decisions and generally should not be omitted from models.

With respect to transferring MNL models from one region to another, the importance of explicitly accounting for differences in traveler behavior among different income classes has a direct bearing on the validity of applying disaggregate models estimated on one group of data (e.g., from one city) to forecast travel behavior for another group of travelers. Ignoring income-specific travel behavior when it is important will result in parameter estimates that are dependent on the income distribution found in the estimation sample.

Auto Availability

Empirical research on disaggregate demand modeling has frequently found that household automobile ownership (AO) significantly influences worktrip mode choice (*21, 13*). There are two basic considerations in using household AO variables in mode choice models:

1. Travel decisions (e.g., mode choice) are not independent of household mobility decisions (e.g., residential location). As such, parameter estimates of AO variables in disaggregate mode choice models will probably be somewhat biased by these unobserved attributes.
2. It is not so much the number of autos in a household as the availability of an auto at the time of the trip that influences choices on how, where, when, and how often to travel.

Researchers appear to agree that possession of a driver's license is essential in defining those individuals actually having a choice between auto and transit. The views regarding auto ownership are mixed. While auto ownership variables have great explanatory power, the problem is that auto ownership is endogenous (i.e., is also explained by mode choice). Some argue that in light of this interdependency and the statistical problems it creates, auto ownership probably should not be employed as a variable because most of the effect is captured by income or wage variables. Others, however, have found that auto availability variables (such as autos per worker) have great explanatory power.

Research on this project has focused on understanding the distinction between automobile ownership and automobile availability, and how these factors influence travel behavior. In the short run, improvements in the performance of disaggregate demand models can be gained by improving their representation of auto availability effects. The larger question of modeling the interaction of household location, automobile ownership, and household travel (frequency, destination, time of day, and mode choice) remains as a long-range research issue.

In understanding the influence of automobile availability on travel behavior, it is useful to introduce the notion of competition for use of a car within the household. Generally, the greater the number of licensed drivers in a household, the greater will be the competition for use of the household's automobiles. Stated another way, as the competition for use of a household's automobile(s) increases, the probability that an auto is used for any given trip should decrease. This is particularly true for a household's worktrips where there is little flexibility on when the trip can be made.

This project assesses the effects of both worker and nonworker competition for use of a household's automobiles (see App. D). In this context, *worktrip competition* is referred to as the number of workers in a household who may make exclusive use of the household's auto(s) for their worktrip. *Nonworktrip competition* expresses the possibly mutually exclusive uses of the household's auto(s) between workers and nonworkers.

Based on models discussed in Appendix D, the estimation results are consistent with a priori expectations on the effects of auto availability. Both worker and nonworker competition for auto use within a household affect work mode choice. Moreover, it was found that the direct competition of two workers for one auto is more significant in determining work mode choice than indirect competition for household auto use by nonworkers.

Other Socioeconomic Variables

Many models for particular choice situations have found other socioeconomic variables to be useful predictors of mode choice. For instance, Ben-Akiva and Atherton (*22*) found employment type to be a useful variable in analyzing carpool incentives. Lifecylce stages can influence the amount of income available for transportation and the need for an auto at home. For instance, a young working couple may have considerable financial resources available for the comfort and convenience of automobile transportation and be relatively insensitive to costs. A suburban housewife may require an auto to get through her day, thereby successfully competing with her breadwinner husband for the car during the day. Age may be a relevant variable where walking and bicycling are included in the choice set.

Unit of Observation and Prediction—Household versus Individual

In formulating disaggregate demand models, the individual traveler generally has been considered to be the basic decision-making unit. However, household characteristics also may reflect individual behavior. Therefore, socioeconomic variables such as household income, number of autos owned by the family, or autos per licensed driver are frequently included in the traveler's utility function.

An alternative strategy could be postulated by assuming that travel behavior by individuals in a household is interrelated. As a consequence of these interdependent travel decisions, the household unit should be the primary element of analysis. For example, the use of the family automobile by the household head for a worktrip may reduce the likelihood of that mode for nonworktrips by other members of the family. On the other hand, if a shopping trip requires the transportation of bulky or heavy parcels, fewer automobiles may be available for worktrips.

The interdependency of travel decisions extends to trips for the same trip purpose. Many households have more than one worker. Hence, for a given number of automobiles owned by a family, it may be desirable to sequentially construct worktrip models for the primary worker and then for the secondary workers. The secondary worker models would include variables that assume values conditional on decisions made by the primary worker. For example, in calibrating such a model the choice of auto by the primary worker reduces by one the number of autos available for worktrips by secondary workers. In forecasting, an expected value based on weighted probability would be calculated. (See Ben-Akiva and Atherton (*22*) for an example of this approach for carpooling incentives.) A major problem with this approach would be empirically identifying primary and secondary workers. Moreover, it would be an even more complex problem to forecast primary and secondary work forces.

Family interactions in disaggregate models ordinarily have been handled by modeling the individual's choice behavior as affected by socioeconomic variables defined at the family level. The other alternative is to develop a hierarchy of choices so that mode choice for the secondary worker or shopper is dependent on the primary worker's worktrip mode choice.

Level-of-Service Variables

This section discusses issues in the specification of level-of-service (LOS) variables: travel time, wait time, and travel cost. The issues considered are the following:

- Generic versus alternative-specific.
- Level of aggregation.
- Network versus observed.
- Perceived versus objective.

Generic Versus Alternative-Specific

In a *generic* specification the estimated coefficient for a variable is restricted to taking the same value across alternatives. With an *alternative-specific* specification the restriction is lifted. A separate coefficient is estimated for each LOS attribute of each alternative.

The advantages of using generic LOS data in disaggregate mode choice models are well known: (1) generic LOS variables are consistent with economic utility theory; and (2) use of generic LOS facilitates demand forecasts of new choice alternatives.

The use of abstract commodity attributes in utility theory was introduced by Lancaster (*23*), and applied in practical applications to numerous aggregate (*24*) and disaggregate (*13, 25, 26*) travel demand model studies. Theoretically, the use of generic LOS variables is well founded. In a mode choice modeling framework, for instance, generic LOS representation assumes that an additional minute spent traveling on a bus is valued equally to an additional minute spent traveling by auto. Indeed, if such were not the case—if, for example, additional bus time is found to be more onerous than additional time spent in an auto—it is due to the effects of unobserved modal attributes omitted from the model (such as comfort, privacy, reliability, etc.). Thus, in a well-specified model that explicitly accounts for all attributes that significantly affect choice, the use of generic representations of LOS is justified.

In practice, however, it is generally not possible to ascertain a priori whether choice models are sufficiently well specified to justify the use of generic LOS variables. This project tested the validity of generic LOS representation (see App. D). The estimation results suggest that mode choice models are not able to distinguish significantly different traveler valuations of travel times and costs between auto and transit. In fact, the hypothesis that travelers' valuations of the LOS variables do *not* differ between modes was tested statistically. In the application in this report, the null hypothesis that the time and cost parameters do not differ between modes could not be rejected. Similarly, the modeling system developed by Cambridge Systematics (*27*) contains generic LOS variables.

How general are the findings of the NCHRP Project 8-13(2) study on the validity of using generic LOS variables in disaggregate choice models? In fact, the findings in this study are encouraging. In models that explicitly specify all significant influences on travel choice, one would expect that time or cost is valued "abstractly." The models estimated were admittedly weak in differentiating the comfort, safety, privacy, and other amenity characteristics between alternative modes. Nonethelsss, it was found that use of generic LOS variables is statistically justified.

However, these findings differ from the conclusions of McFadden et al. (*19*) on the use of generic versus alternative-specific LOS variables. McFadden et al. concluded that although the importance of in-vehicle time did not seem to vary for public transportation modes (bus and BART), auto in-vehicle time was valued differently from in-vehicle time for public transportation. Surprisingly, auto in-vehicle time was found to be more onerous than transit time.

The authors speculate that this apparent anomaly reflects the fact that transit in-vehicle time is not really very onerous, and that other separately specified variables in the model such as out-of-vehicle time may represent the unpleasant aspects of transit travel. In fact, transit travel time per se may be more relaxing than auto driving because the burden of driving is removed from the traveler. CRA also suggests that the most undesirable features of transit such as discomfort and schedule unreliability may have been captured in the alternative-specific constant term, since these features may not vary substantially with trip length.

Because the conclusion of McFadden et al. differs from the approach used in many disaggregate mode choice models where LOS variables are generic, further research on the issue of generic versus alternative-specific LOS variables would be useful in resolving the differences among previous studies.

Level of Aggregation

The general rule (see Train (*28*)) is that data generally should be disaggregated to the lowest level possible. Some examples follow:

- Time should be disaggregated into in-vehicle, walk, and transfer components.
- Variables should be specific to the individual decision maker (individual values are preferred over, say, zonal averages).

However, it should be noted that disaggregate LOS data may be difficult and costly to obtain, especially for alternatives, in the calibration data set and difficult to forecast for the forecast data. In many cases, hand-coded data differ significantly from network data. At minimum, it has a greater variance between respondents. These differences can have significant effects on the estimated model.

As disaggregate models have been calibrated and implemented, a considerable amount of attention has been given to the question of use of network zonal averages for LOS data versus data calculated specifically for the individual. An early CRA study (*13*) made a careful attempt at collecting LOS data specific to the individual traveler, but many studies since that data have been, by necessity, required to rely on network averages even when calibrating a disaggregate model. The disaggregate LOS data are obviously preferred, but tedious collection of such disaggregate data mitigates to some extent the purported data economy of disaggregate models. The evidence on use of disaggregate LOS data is discussed below.

Network vs. Observed Data

In this discussion, it is assumed that LOS data calculated specifically for the individual (observed data) correctly measures the characteristics of the travel alternatives facing the individual. The network data, which are zonal averages of the LOS variables, are used to approximate the observed variation. As a simplification, it is assumed that there is a linear relationship between observed and network data.

$$O_{ijt} = a_{ij} + b_{ij}N_{ij} + \epsilon_{ijt} \tag{23}$$

in which

O_{ijt} = the value of the i^{th} observed variable for the j^{th} alternative for the t^{th} individual;

N_{ij} = the corresponsing network variable;
a_{ij}, b_{ij} = coefficients; and
ϵ_{ijt} = a random error term.

Two questions can be raised. First, how accurately do network variables approximate observed variables? Ideally, one would want $a_{ij} = 0$, $b_{ij} = 1$, and $\epsilon_{ijt} = 0$, i.e., that the network and observed variables coincide. Talvitie and Dehghani (*29*) performed linear regression analyses of the form of Eq. 23 on several LOS variables for auto, bus, and BART modes. In many cases, correlation coefficients were low and a_{ij} and b_{ij} differed substantially from 0 and 1, respectively. This finding suggests that network data may not always closely approximate observed data.

The second question is how do models estimated with observed and network data differ? To analyze this issue, it is useful to express the observed variable as the sum of the corresponding network variable and a residual term, i.e.,

$$O_{ijt} = N_{ijt} + r_{ijt} \quad (24)$$

in which

$$r_{ijt} = a_{ij} + (b_{ij} - 1)\, N_{ijt} + \epsilon_{ijt} \quad (25)$$

Equation 24 shows that estimation of a model with network variables replacing the correct observed variables results in the exclusion of the residual, r_{ijt}. This is a form of specification error that results in biased model coefficients.

Tardiff's (*30*) results on specification errors can be used to analyze the nature of the bias. The random component, ϵ_{ijt}, in Eq. 25 causes a downward bias in the coefficients of the model, including the one corresponding to the LOS variable in question. If $b_{ij} \neq 1$, the residual term is correlated with the network variable (and possibly other variables in the model). This causes bias that can be either upward or downward depending on the direction of correlations between the residual and the other variables.

In general, therefore, the use of network data results in biased coefficients. This result is consistent with Horowitz's (*31*) finding, but seems to contradict the finding of McFadden and Reid (*32*) that when $b_{ij} = 1$, the use of network data does not result in biased coefficients. The apparent contradiction may arise from the fact that Tardiff and Horowitz assume maximum likelihood estimation, while McFadden and Reid assume the Berkson method (which uses the "log odds" formulation and least squares estimation). Tardiff has shown that the two estimation methods can yield different coefficients, even in very large samples, when important variables such as r_{ijt} are excluded from the model.

The estimation procedure assumed by McFadden and Reid requires that individuals with similar values on the observed LOS variables be grouped together before estimating the model. This would be difficult, if not impossible, to accomplish if network data were used. Maximum likelihood estimation has become the dominant method for travel choice modeling. Therefore, the findings of Horowitz and Tardiff are more germane to this issue.

Talvitie and Dehghani (*29*) and McFadden et al. (*19*) present empirical evidence on bias and the use of network data. The coefficients of models using network data are quite different from the corresponding coefficients of models using observed data, even though there is very little difference in goodness of fit. It is particularly interesting that network models indicate that out-of-vehicle time is consistently considered more onerous than in-vehicle time, while the models estimated with observed data do not.

"Perceived" Versus "Objective" Data

Considerable debate has ensued in the literature over whether "engineering" LOS data, based on sources other than the traveler, or data as "perceived" by the traveler are the appropriate variables.

Engineering LOS data are generally derived from computerized "skim trees" representing the travel times and distances between nodes in transportation network. These data are generally available for most urban areas and were developed during the highway building boom of the 1950s and 1960s. These data, however, may be outdated and may not accurately reflect the LOS for all individuals for all trips.

Perceived LOS data can be developed during the travel survey used to collect data on socioeconomic characteristics for model calibration. The argument for the use of "perceived" data is obvious. Since the traveler is responding to the factors he perceives, perceived data are "obviously" superior. On the other hand, if models are to be transferable, there must be an explicit mechanism for translating engineering data to perceived data. As a practical matter, most forecasters have only engineering data available, and developing sufficient perceptual data for both modeling and forecasting purposes would put an unacceptable burden on the interview process. Finally, there is the argument that perceptual data are engineering data weighted or discounted by the model coefficients and therefore should not be weighted again in the model calibration.

A major problem with perceived data is that travelers' perceptions of LOS on their alternatives to their chosen alternatives are likely to be poor. People have enough trouble estimating the time it takes them for the chosen alternative without having to guess how long it would take them by a rejected alternative.

Another major problem with respect to the use of perceived LOS data is the policy variables one can or one wishes to manipulate. Clearly, it is within the scope of public policy to change the objective level of service offered by transit or highway systems. Fares, tolls, headway, and congestion all can be externally controlled. Perceptions of LOS, which may or may not be closely tied to engineering LOS, are not as readily manipulable. Transportation planners have little experience in making the bus seem faster or the auto feel more expensive. Consequently, for these reasons the use of perceived LOS data for general planning applications is not recommended.

Attitude Variables

Attitude variables have been suggested for inclusion in travel choice models. Three types of attitude variables have been used in previous studies: (1) perceptions of the attributes of alternatives; (2) general feelings toward the alternatives; and (3) general lifestyle attitudes, e.g., perference for privacy, being on time, etc.

Empirical evidence on the usefulness of such variables is inconclusive. McFadden et al. (19) conclude that although particular attitudinal variables may have some explanatory power, attitudinal variables only marginally improve models already containing objective variables. At the other extreme is the finding of Dobson and Tischer (33) that perceptions of modal attributes are clearly superior to objective variables. A middle position is that of Gilbert and Foerster (34), who find that some attitudinal variables are quite useful and others are not.

If there is a strong relationship between attitudinal variables and travel behavior, the issue of causality arises. In order for attitudinal variables to be properly specified in travel choice models, the direction of causality must be from attitude to behavior, i.e., a change in attitude leads to a change in behavior. Empirical evidence on this issue suggests that the reverse causality of behavior influencing attitudes may be at least as important as the commonly assumed direction (35, 36).

In addition to these unresolved theoretical issues, models with attitudinal variables are more difficult to use for planning and analysis purposes. Not only must data on existing attitudes be collected, but forecasts on future values of the attitudinal variables must be made. The latter problem has received very little research attention. Therefore, because of the theoretical and practical difficulties with attitudinal variables, they are not used in this project.

Functional Form for Explanatory Variables

In most previous applications, the explanatory variables were entered directly into the representative utilities (i.e., utilities have been assumed to be linear functions of LOS variables and other characteristics). When the model is transformed into the "log odds" ratio form (Eq. 12), differences between characteristics of two alternatives are emphasized. However, the parameter estimates remain linear in form.

In early binary choice modeling studies, there was some discussion of whether differences were the most appropriate comparison of model characteristics. For example, Stopher and Lavender (37) compared differences, ratios, and logarithms of ratios in the "log odds" form of binary logit models and found only slight differences in goodness of fit. Similarly, Watson (38) argued for dividing the difference by the average time for the two modes. While there is some intuitive appeal for simple ratios and relative time differences, these comparison functions do not readily generalize to the multinomial choice situation.

The use of ratios as the comparison function introduces the problem that the model is sensitive to which alternative appears in the denominator of the ratio. The problem can be illustrated by a simple binary choice model with a single independent variable, cost. It is further assumed that there is a flat transit fare, so that transit cost is invariant.

If auto cost is used in the denominator, the resulting binary logit model is

$$\ln\left(\frac{P_A}{P_T}\right) = a\,\frac{\mathrm{COST}_T}{\mathrm{COST}_A} \tag{26}$$

Since transit cost is invariant, the log odds ratio varies with the inverse of auto cost. However, if transit cost is used in the denominator, the log odds ratio would vary directly with auto costs. This is clearly a different model. The difficulties with the ratio comparison are discussed in greater detail by Oum (39).

The logarithm of ratios comparison can be generalized to the multinominal logit case. The logarithms of the variables are used in the utility functions for the alternatives. More generally, variables can be transformed prior to entry in the utiilty function, i.e.,

$$V_{it} = \sum_{k=1}^{K} f_k\,(X_{itk})\,\beta_k + \sum_{l=1}^{L} g_l\,(S_{tl})\,\alpha_l \tag{27}$$

in which

- V_{it} = the representative utility of alternative i to individual t;
- K = the number of LOS attributes for alternative i;
- X_{itk} = the value of LOS attribute k for alternative i and individual t;
- β_k = utility parameter estimate for attribute k;
- f_k = any functional form or transformation of the value of X_{itk};
- L = the number of socioeconomic attributes of individual t;
- S_{tl} = value of socioeconomic attribute l for individual t;
- α_l = utility parameter estimate for attribute l; and
- g_l = any functional form or transformation of the value of S_{tl}.

Koppelman (40) has presented several alternative functional forms including linear [$f_k(Y) = Y$ and $g_l(Y) = Y$], logarithmic [$f_k(Y) = ln(Y)$ and $g_l(Y) = ln(Y)$], and power [$f_k(Y) = Y^{ck}$ and $g_l(Y) = Y^{cl}$] functions.

Koppelman estimated alternative work mode choice models with time and cost variables entered as linear functions, logarithmic functions, and power functions, respectively. Koppelman concluded that the power function specification was superior in terms of goodness of fit. These results should be interpreted cautiously, however. Even though there were statistically significant differences in goodness of fit, all specifications produced very similar goodness-of-fit measures. Further, since it can be shown that the linear and logarithmic functions are special cases of the power function, one would expect a superior fit for the power function.

Although there are usually not large differences in goodness of fit among models using alternative functional forms for the explanatory variables, there can be large differences in predicted choice probabilities, especially for applications beyond the range of existing data sets. For example, because of gasoline price increases, automobile operating costs are higher than those represented in data sets typically used to estimate existing models (early 1970s or before). Therefore, predicted modal splits at higher gasoline prices may be quite sensitive to how auto costs and other variables are entered into the utility functions.

The elasticity formulas presented earlier assume that LOS variables enter linearly. In the case of nonlinear transformations, these formulas become

$$E^{P_i}_{X_{itk}} = (1-P_i)\,\beta_k\,X_{itk}\,\frac{\partial f_k\,(X_{itk})}{\partial X_{itk}} \tag{28}$$

and

$$E_{X_{itk}}^{P_j} = -P_i \beta_k X_{itk} \frac{\partial f_k (X_{itk})}{\partial X_{itk}} \tag{29}$$

for the direct and cross-elasticities, respectively.

In the case of the logarithmic transformation, the elasticities are

$$E_{X_{itk}}^{P_i} = (1-P_i)\,\beta_k \tag{30}$$

and

$$E_{X_{itk}}^{P_j} = -P_i\beta_k \tag{31}$$

These formulas indicate that the elasticities are bounded, i.e., they range from zero to β_k in absolute value. In contrast, the elasticities previously presented (Eqs. 21 and 22) are unbounded.

The alternative elasticity formulas illustrate the possibility that alternative functional forms may yield very different predicted probabilities. For example, data from the Nationwide Personal Transportation Study (*41*) suggest that auto drive-alone mode splits do not vary substantially with trip length (*42*). Since auto operating costs are roughly proportional to trip length, this suggests that the direct cost elasticities of auto drive alone trips would not vary substantially with trip length if the logarithmic transformation is used. This, in turn, suggests that short and long trips are similar in their elasticity to auto cost increases. On the other hand, since the elasticity formula corresponding to the linear representation of cost directly increases with cost, long trips would be much more sensitive to auto cost increases than would short trips.

Effects of Specification Errors

One issue that must be addressed in specifying disaggregate logit models is the consequences of excluding behaviorally relevant variables or including behaviorally irrelevant variables. Tardiff's research (*30*) concludes that inclusion of superfluous independent variables does not affect the consistency of the estimation, but exclusion of behaviorally relevant variables can lead to biased estimates. Consistency is a property of estimation that, intuitively, means that the estimate gets closer and closer to the true value of the unknown parameter as the sample size increases. This bias would result in a situation where the excluded variable was correlated with the included variables (see App. E for discussion and examples), or more generally where there is a change in the structure of the random utility components. However, the author notes that if the correlation between included and excluded variables is constant over time, the underspecified model would provide useful forecasts despite the bias. An example would be where preference for transit was correlated with transit LOS. An improvement in transit service might, over time, lead to household location decisions that preserved the correlation and validated the forecast, even though the model coefficients would be biased.

MODEL ESTIMATION

Logit models are generally estimated by one of two alternative procedures. First, data on the calibration sample can be loaded into a maximum likelihood estimation computer package for logit models. (This is generally the preferred approach.) Second, the data can be preprocessed into the log odds format and the model estimated using a least squares multiple regression computer routine.

The maximum likelihood approach iteratively solves for the set of coefficients, β's and α's, which yields the representative utilities, V's, which generate the best fit to the observed pattern of choices in the calibration sample. The estimation package will iterate through the problem until the estimated coefficients reach a specified convergence criterion or the estimation completes a specified number of iterations. The least squares approach finds the set of coefficients that minimizes the sum of the squared errors between the predicted and observed log odds ratios.

In either case data are required on the unit of observation (individual or household) and each of its specific alternatives in the choice set. It should be recalled that variable coefficients may be specified as *generic* or *alternative-specific*. The generic specification requires the coefficient to take on the same value across all alternatives. For example, LOS attributes that are weighted the same across all modes are specified generically (e.g., in-vehicle travel time). Alternative-specific coefficients take on a different estimated value for each alternative or are restricted to entering only some representative utility functions. For instance, in some mode choice model specifications, auto availability is restricted to only auto alternatives, or transfer time is restricted to transit alternatives. If it is assumed that out-of-vehicle time is more onerous for the transit alternative than the auto alternative, separate coefficients for this variable can be estimated for each alternative.

STATISTICAL SIGNIFICANCE TESTS

This section summarizes the appropriate techniques for performing tests of the statistical significance of individual logit model parameter estimates, overall measures of model goodness of fit, and tests of hypotheses concerning linear combinations of model coefficients. Throughout the presentation here, it is assumed that the model parameters are derived from a maximum likelihood (ML) estimation technique. The statistical properties of ML estimators are discussed in detail elsewhere (see Theil (*43*) and McFadden (*14*)). However, a brief review of ML estimation here will serve to define the statistics employed in hypothesis tests on logit model parameter estimates.

The Maximum Likelihood Method

A likelihood function describes the probability of observing a given choice sample (that is, the sample used in estimating the coefficients in a logit model) when the distribution of the underlying random component of utility is known. For logit model estimation, the underlying distribution is assumed to be Weibull. (See Appendix E, Section 3, for a further discussion of this point.) In the context of logit model estimation, the likelihood function of a disaggregate sample may be written as

$$L(\theta) = \prod_{t=1}^{T} \prod_{i \epsilon C_t} P_{it}^{\delta_{it}} = \prod_{t=1}^{T} \prod_{i \epsilon C_t} \left[\frac{e^{x'_{it}\theta}}{\sum_{j \epsilon C_t} e^{x'_{jt}\theta}} \right]^{\delta_{it}} \tag{32}$$

in which

Π is the product notation, e.g., $\prod_{t=1}^{2} X_t = X_1 X_2$;

T = the number of observations in the sample t;

$i \epsilon C_t$ = denotation that alternative i is in choice set C_t;

C_t = the set of choices for individual t;

δ_{it} = 1 if alternative i were actually chosen by individual t;

P_{it} = the probability that individual t chooses alternative i as determined by the logit model

$$P_{it} = \frac{e^{x'_{it}\theta}}{\sum_j e^{x'_{jt}\theta}}$$

x_{it} = the set of K variables (a 1xK vector) included in the model; and

θ = a Kx1 vector $(\theta_1, \ldots \theta_k, \ldots \theta_K)$ of model parameters.

$L(\theta)$ = the likelihood function evaluated at θ.

It is desired to find estimates of the parameters θ which maximize the likelihood of observing the sample used in estimation. This leads to the specification of a set of K first order conditions:

$$\frac{\partial L(\theta)}{\partial \theta} = 0, \text{ for all } k \tag{33}$$

Normally, this maximization procedure is performed on a log transformation of the likelihood function,

$$L^*(\theta) = \ln L(\theta) \tag{34}$$

to simplify the computation of derivatives. McFadden (*44*) has shown that under most conditions the θ derived from the procedure previously outlined are unique and possess optimal asymptotic properties. Many of the hypothesis tests presented below employ comparisons of the likely functions for alternative θ.

Tests for the Statistical Significance of Individual Parameter Estimates

Theil (*43*) has shown that in large samples, minus the inverse of the matrix of second derivatives of L^* is the variance-convariance matrix. This allows for the computation of the standard errors of each parameter estimate. Thus, one may test the simple hypothesis that any given parameter differs from 0 with a t-statistic defined as the ratio of the parameter estimate to its standard error. Standard logit estimation packages give the standard error and t-statistic for each estimated coefficient.

Measures of Overall Goodness of Fit

Three different measures of how well a fitted logit model explains the variation in choice in an estimation sample may be employed: (1) coefficient of determination, (2) percent correctly estimated, and (3) goodness-of-fit test.

Coefficient of Determination

This measure is analogous to the squared multiple correlation coefficient (R^2) in the linear statistical model. Note from Eq. 32 that if a model perfectly predicts the choices of all individuals in a sample, the likelihood function will have the value 1. (In this case the term "predicts" is used in the sense that the chosen alternative is assigned a probability of 1. Actually, since a logit model is asymptotic to 0 and 1 probabilities in its tails, one can never precisely achieve a likelihood function value of 1.) At the other extreme, if all the parameters in a logit model are 0, the model predicts that all choices for any given individual are equally likely. In this case, the model does not explain choice variation and the corresponding value of the likelihood function is much smaller. (For example, in a binary choice logit model, the likelihood function with T observations and $\theta = 0$ would be $(0.5)^T$.) These observations suggest a natural measure of goodness of fit that can be used to compare alternative models estimated on the same data set:

$$\rho^2 = 1 - \frac{L^*(\hat{\theta})}{L^*(0)} \tag{35}$$

in which

$L^*(\hat{\theta})$ = log of the likelihood for the vector of estimated coefficients; and

$L^*(0)$ = the value of L^* for $\theta = 0$.

Because R^2 in the linear statistical model is a measure of the improvement of a linear model over a model that contains only a constant term, the preceding definition of ρ^2, which has been commonly reported in previous studies, is not completely analogous to R^2. Tardiff (*45*) has discussed a revised definition of ρ^2, originally suggested by McFadden (*14*), which corresponds more closely to R^2.

$$\rho^2_c = 1 - \frac{L^*(\hat{\theta})}{L^*(C)} \tag{36}$$

in which $L^*(C)$ = the log likelihood for a model with all coefficients excluding alternative specific constants equal to zero. Like R^2, ρ^2_c lies between 0 and 1 with higher values connoting improving goodness of fit. The lower bound of the more commonly used ρ^2 equals 0 only when the proportions of the sample selecting each alternative are equal. When these proportions are unequal, the lower bound of ρ^2 can be relatively high (*45*). In these cases, relatively high values of ρ^2 could occur even when the independent variables are highly insignificant.

Percent Correctly Estimated

A second measure of how closely a model explains actual choice is the percent of cases in an observation sample where the highest predicted probability is for the alternative actually chosen. For example, in a binary choice model, one would assign a "correct" prediction to any individual whose chosen alternative was predicted to have a choice probability greater than 0.50. The deficiency with this goodness-of-fit measure is that it fails to distinguish between models that predict correct choices in the "tails" of the logit function (a desirable property) and models that predict choices near the midpoint of the probability range. Moreover, this measure does not lend itself to a rigorous test of significance in the sense of being able to assign a confidence level to the parameter estimates.

Goodness-of-Fit Test

This test overcomes the deficiencies with the percent correctly estimated measure cited above. It has been shown that a weighted sum of squared deviations of observations from predicted probability values is distributed as χ^2. Interested readers are referred to McFadden (*14*).

Tests of Linear Hypothesis

The log-likelihood ratio test is a useful technique for testing hypotheses on logit model parameter estimates. The following are examples of hypotheses that can be evaluated.

- Whether the set of estimated parameters θ significantly differs from 0. (In fact, neither a likelihood test nor a goodness-of-fit test, as previously described, can be used for this hypothesis. McFadden (*46*) has indicated that in small samples, the likelihood ratio test is more stable than the goodness-of-fit test.)
- Whether subsets of parameter estimates satisfy specified relationships. (An example here would be a test of the hypothesis that the parameter estimates of alternative-specific LOS do not specifically differ between modes.)
- Whether logit model parameter estimates are transferable from one data set to another.

Theil (*43*) has shown that for large samples, minus twice the logarithm of a likelihood ratio is distributed as chi-square:

$$\chi^2(q) = -2\ln\frac{L(\theta_o)}{L(\hat{\theta})} \tag{37}$$

in which

$\chi^2(q)$ = the chi-square statistic with q degrees of freedom (q is the difference in the number of coefficients between the null hypothesis and test hypothesis);
$L(\theta_o)$ = likelihood computed as a function of the K-element vector θ_o where q elements ($q < K$) take certain values corresponding to the null hypothesis; and
$L(\hat{\theta})$ = likelihood computed as a function of the maximum likelihood estimators $\hat{\theta}$.

Most standard logit packages provide the user with the log-likelihood ratio for each estimated equation.

An example of the log-likelihood ratio test follows.

Consider a test of the hypothesis that the coefficient of a travel time variable is the same for both auto and transit in a binary split mode choice logit model. To perform this test, two models are estimated. The restricted model embodies the null hypothesis assumption that the auto and transit travel time coefficients are equal. In the restricted case, the time coefficient for auto and transit would be specified as a generic variable that does not vary between alternatives. The unrestricted model estimates two separate time coefficients —one for auto and one for transit, as an alternative-specific specification. In this case, there is one constraint ($q = 1$) in the null hypothesis; that is, a single "generic" travel time parameter is estimated. Another way of thinking about calculating q is that there is a difference of one coefficient in the number of estimated parameters between the restricted and unrestricted models. If the computed value of $\chi^2(1)$ exceeds the critical value of $\chi^2(1)$ for a specified significant level, the null hypothesis can be rejected. When the null hypothesis is rejected, it is concluded that the coefficients for auto and transit time are different (the alternative-specific specification).

EXAMPLES OF CHOICE MODELS

The following section describes the estimation of some travel choice models. Models are included in the order of increasing conceptual complexity. They include worktrip mode choice, shopping trip mode choice, and destination choice. In addition, the section discusses multidestination tripmaking (trip chaining) and trip generation.

Binary Mode Choice

As part of this study, a binary choice model of worktrip mode split was estimated using disaggregate data processed from a 1967 Pittsburgh household interview survey and supplementary traffic network LOS data. Work mode split models were also estimated on data from the Twin Cities metropolitan area. The Twin Cities models are presented in Section 6 of Appendix D. The estimation sample consisted of 115 observations of auto driver and bus (with walk access) worktrips drawn from selected travel corridors in the Pittsburgh metropolitan area.

The mode choice model, shown in Eq. 7, included socioeconomic variables to account for differences in travelers' tastes and LOS variables to measure the relative impedances of the auto and transit modes. LOS was represented by three variables to measure the separate effects of in-vehicle travel time, *IVTT;* out-of-vehicle travel time (walk access), *OVTT;* and travel cost, C/INC_i, on mode split. The travel cost variable entered in the model was divided by income on the hypothesis that travelers with different incomes place different values on the cost of alternative modes in making their travel choices.

A separate income term, *HINC,* was also entered in the model as a "pure shift" variable. This socioeconomic variable reflects the hypothesis that all other factors equal, higher income travelers may more strongly prefer auto relative to lower income travelers. A second socioeconomic variable employed in the model was the auto per worker, *APERW,* term to assess the mode choice effects of competition among workers in a household for use of the household's automobiles. Finally, a mode-specific constant was entered in the model to capture the average effect of omitted attributes that may influence mode choice.

Estimation results of the mode choice model are summarized in Eq. 38 with the model displayed in a log odds formulation. The estimated coefficients precede each of the variable names. (The numbers below each coefficient represent t-statistics.)

$$\ln \frac{P\,(\text{auto})}{P\,(\text{transit})} = \underset{}{-5.72} + \underset{(1.44)}{1.38\,HINC} + \underset{(3.74)}{4.07\,APERW} \underset{(-2.18)}{-0.117\,OVT} - \underset{(-1.88)}{0.0348\,IVTT} \underset{(-4.28)}{-9.06\,C/INC} \tag{38}$$

P (auto) = probability of choosing auto;
P (transit) = probability of choosing transit;
HINC = 1 if household income exceeds \$7,000 per year; 0 if otherwise (alternative-specific variable entered in the auto utility function);
APERW = autos per worker (alternative-specific variable entered in the auto utility function);
OVTT = difference (auto minus transit) in out-of-vehicle travel time (in minutes);
IVTT = difference (auto minus transit) in in-vehicle travel time (in minutes); and
C/INC = difference (auto minus transit) in cost in dollars divided by income code (see Table C-15).

All the estimated coefficients have the theoretically correct sign and are significant at the 10 percent level (one-tailed significance test). As expected, the coefficient of the income shift variable is positive, suggesting that, all other factors equal, higher income travelers exhibit a preference for auto travel on their worktrip. The positive coefficient on the variable *APERW* suggests that the greater the number of autos relative to workers in a household, the greater the probability that a traveler will choose auto for the worktrip.

All the coefficients of the LOS terms are, as expected, negative. The relative magnitude of the coefficients of out-of-vehicle and in-vehicle travel time suggest that work travelers find walk access time more onerous (on the order of three to one) than in-vehicle riding. The negative coefficient on the travel cost term confirms expectation that the higher the cost of a mode, the lower the probability that mode will be chosen by a traveler. The values of time derived from the model for a traveler whose family income was \$8,000 per year (1967 dollars) were \$1.15 per hour and \$3.88 per hour for in-vehicle and out-of-vehicle travel times respectively. (See Appendix D for discussion of value of time estimates.)

The overall predictive ability of the estimated model was good. For 107 of the 115 travelers in the sample, the highest predicted mode selection probability corresponded to the mode actually chosen.

Models of Shopping Travel Choice

Shopping travel has many more choice dimensions than work travel behavior. Unlike work travel, frequency and destination are generally not fixed for the shopper. Consequently, a fully specified model of shopping choice behavior would consider at least frequency, destination, and mode choices. The underlying assumption of the MNL model is that travelers choose among alternatives so as to maximize their utility. In shopping (discretionary) travel, however, the number of choice alternatives is large. Unlike the short-run work trip analysis previously discussed, for shopping trips the choice of destination is no longer determinate, nor for that matter is the decision to make a trip at all. Thus, a truly behavioral demand model must be capable of characterizing the factors influencing the full range of choice—how often to travel, by what mode, and to what destination.

One approach to the model of shopping travel demand is a conditional probability formulation that splits each dimension of choice—frequency, f, destination, d, and mode, m —into a separate logit model specification. Ultimately, one is interested in predicting the joint probability of a household's choice of each fdm bundle, or in aggregate terms, the number of travelers by mode between each zone pair. However, using the laws of conditional probability, and assuming additive separability of utility to factor the simultaneous travel decisions of an individual into a series of separate choice models, one can always reconstruct the joint probability according to:

$$p\,(f,d,m) = p\,(m/f,d) * p\,(d/f) * p\,(f) \tag{39}$$

in which

$p\,(f,d,m)$ = joint probability of choosing trip frequency f, destination d, and mode m;
$p\,(m/f,d)$ = conditional probability of choosing mode m given f and d choices;
$p\,(d/f)$ = conditional probability of choosing destination d given f; and
$p\,(f)$ = marginal probability of choosing frequency f.

The conditional probability structure represented by Eq. 39 allows for a substantial savings in the number of parameters that must be estimated in any given model formulation, and facilitates identifying which explanatory factors influence the separate dimensions of travel choice.

Estimation of a shopping mode choice model is presented as follows.

The shopping trip mode choice model represents a conditional probability structure of auto and transit choice. The estimation sample consists of 140 observations randomly drawn from specific travel corridors in the Pittsburgh metropolitan area (see Section 1, App. D). As in the worktrip mode split analysis discussed earlier, mode choices were limited to auto driver and bus (with walk access). Auto was the mode chosen for 56 percent of the trips. The LOS data reflected travel conditions for the time of day the shopping trip was actually taken.

Estimation results for the shopping trip conditional probability mode choice model are summarized in Eq. 40. The model specification includes three LOS variables: (transit) walk access time, auto and transit in-vehicle travel times, and modal costs divided by an index of income. A pure income shift variable, *HINC*, and a term representing household autos per licensed driver were included as socioeconomic descriptors. Although the model specification of the shopping mode choice model is similar to the work mode choice model, no deliberate attempt was made to parallel the two.

$$\ln \left[\frac{P(\text{auto})}{P(\text{transit})}\right] = \underset{(-4.62)}{-6.63} + \underset{(2.48)}{2.16\,HINC} + \underset{(2.13)}{2.03\,APERDR} \qquad (40)$$
$$\underset{(-3.71)}{-0.34\,OVTT} \underset{(-2.02)}{-0.04\,INVTT} \underset{(-3.54)}{-13.50\,C/INC}$$

in which

P (auto) = probability of choosing auto;
P (transit) = probability of choosing transit;
$HINC$ = 1 if household income exceeds $7,000 per year; 0 otherwise;
$APERDR$ = autos per licensed driver;
$OVTT$ = difference (auto minus transit) in out-of-vehicle travel time (in minutes);
$IVTT$ = difference (auto minus transit) in in-vehicle travel time (in minutes); and
C/INC = difference (auto minus transit) in cost in dollars divided by income code.

The parameters of all the variables in the shopping mode choice model were of the correct sign and significant at the 5 percent level. The relative magnitudes of the parameters for transit walk ($OVTT$) and in-vehicle travel time are on the order of eight to one, suggesting that travelers find walk access considerably more onerous than in-vehicle travel time on a shopping journey.

As in the work mode choice model, the coefficient of the income shift variable was positive, indicating that, all other factors equal, high income travelers exhibit a preference for auto travel. Auto availability as represented by the term "autos per licensed driver" ($APERDR$) also positively influenced auto mode choice.

The model correctly predicted the mode choice of 92.86 percent of the travelers in the estimation sample. The overall goodness of fit of the model is good as indicated by the coefficient of determination, ρ^2, equal to 0.71. The coefficient of determination and other goodness-of-fit measures were discussed earlier in this chapter.

Destination Choice Models

In modeling choice of destination there may be more than two destination alternatives, and a simple binary choice model is no longer appropriate. However, the multinomial logit model can include any number of choice alternatives if the model assumptions are valid.

Trip decisions other than mode choice have not received the attention in disaggregate models that has been paid to mode choice problems. For that reason one is less certain of the variables to be contained in the equations for these other choices. Some preliminary work has been done, however, and one can obtain clues as to the important variables from this work.

Destination models using the MNL technique use some variables that are found in traditional gravity models. As in the gravity model, the two relevant classes of variables in destination choice probability models are "impedance" and "attraction" terms. Impedance is the utility associated with travel from each origin zone to each possible destination zone. It takes on a specific value for each origin-destination pair. Attraction is the utility of destination, which is specific to each destination.

In MNL models of destination choice, the impedance term has been called the "inclusive price" or "inclusive value." It will be recalled that the MNL model assumes that the representative utility of alternative i is given in Eq. 6. To be consistent with the logit specification, the inclusive value measure for each origin combines the utilities of travel to a particular destination by all of the available modes by taking the log of the sum of the exponentials of the utilities. The inclusive value travel from origin o to destination d, IV_{od}, say, is given by

$$IV_{od} = \ln \sum_{m=1}^{M} e^{V_{mod}} \qquad (41)$$

in which

IV_{od} = the inclusive value of destination d;
V_{mod} = representative utility of mode m for travel from origin o to destination d;
M = the number of available modes; and
ln = the natural logarithm.

The variables IV_{od} enter directly into destination equations respectively. This measure of the inclusive value of travel from o to d is often called the "log sum of the denominator." For a given origin-destination pair, it is the denominator of the logit mode choice equations.

To give an example, suppose that the exponentiations of the expected utility of auto and bus to the first destination (the e^{V_m} respectively) are 2 and 3. Summing them one gets 5 and ln(5) = 1.609. Similarly, for the second destination assume that the exponent of utility for the two mode choices are 2 and 4; ln(6) = 1.792. The values, 1.609 and 1.792, would enter the utility function of the first two destination choice alternatives for the destination choice model. This procedure assumes, of course, that the modeler already has a mode choice model that he has transferred to or estimated for the region where he is calibrating the destination choice model. This mode model is used to get the origin-destination specific values of V_m in order to calculate the inclusive value or impedance.

With this definition of inclusive value, the higher the value of the travel time and cost to destination i, the lower the inclusive value. Note that this formulation of the inclusive value is expected to have a sign opposite from that used in CRA (*1*). A more appropriate title is therefore "inclusive value," rather than the earlier "inclusive price." Thus, the coefficient of the inclusive value in the destination choice model should have a positive sign.

The "attraction" of destinations in a shopping trip destination choice model would measure the level of shopping opportunities at each destination alternative. In CRA (*1*) this variable, EMP_i, was defined as the retail employment in destination (traffic) zone i expressed as a fraction of total regional retail employment. However, such a specification cannot lead to a transferable and consistent modeling specification for an MNL model. The reason is that the attraction variable is defined as percent of the total, meaning that the size of the variable is dependent on the city size. However, the IIA property requires that the relative shares of two destinations be independent of any third destination's attributes, thereby causing an inconsistency.

For the MNL model to be specified, the definition of the difference in the attraction measure between the two destinations should not be dependent on the number of destinations.

The variable "retail employment in the destination zone as a percentage of total retail employment in the Pittsburgh region" (in CRA (1)) clearly does not have this property, as a simple example will illustrate. The logit model predicts the log odds ratio of the probability of the first destination to the second based on the differences in the attraction measure. Suppose they are the only two destinations and the first employs 200 persons and the second employs 100. The difference in the values of the independent variables is 0.66 − 0.33 = 0.33. The addition of a third choice alternative with 100 employees causes the difference in the variables to take on a value of 0.5 − 0.25 = 0.25 and changes the log odds ratio of the two destinations, thereby making it inconsistent with the IIA property. The attraction variable must be somehow normalized so that it can apply to any subset of the choice set, i.e., it does not depend on the size of the choice set if the logit model is applied. Of course, even if the attraction variable were defined consistent with the IIA, it would still remain to be proved that the IIA be applicable to the joint destination choice/trip frequency decision.

In short, the researcher must be careful to make the impedance variable and the attraction variable consistent with the MNL framework, unless there is a desire to depart from that framework to correct a known violation of the IIA. Ordinarily, this is accomplished by making the impedance term equal to the "log sum of the denominator" from the mode split equation and appropriate constraints are placed on the specification of the attraction variable, as described below.

Sensitivity of the Destination Choice Model to Partition of the Destination Set

An early study, Watanatada and Ben-Akiva (*47*), appropriately defined and constrained the attraction variable. The attraction variable that entered the utility of each destination was 1.0 ln (Q_d), that is one times the log of the attraction variable in the destination. Although this specification may seem unusual, it can be shown that it provides both for consistent aggregation of zones and for consistency with the IIA property.

It is desirable that forecasts with destination choice models be invariant to the aggregation of zones; i.e., if two zones are added, the forecast of aggregate share should be the same as the sum of the individual forecasts.

In a properly specified logit model, if a zone is partitioned, the separate shares should sum to the total of the aggregated zone. A further desirable property is that scalar changes in the size of two zones should not affect the log odds ratio.

It is also clear that different variables should be treated differently after aggregation. If a homogeneous zone is divided into two, the two separate zones should have the same transportation times and costs, but half the share.

The log transformation of the attraction measure has the property that the market shares of two competing destinations are insensitive to a scalar change in the attraction measures for both alternatives since $\log(\lambda A) = \log\lambda + \log(A)$, and $\log(\lambda)$ nets out of the difference in attractions.

This property is helpful for transferability because it ensures that a scalar change in attraction measures across zones does not affect relative shares. This specification also satisfies the other requirements mentioned above. The aggregate share of two zones is invariant to partitioning between the zones if the coefficient is 1 on the log of attraction. This may be demonstrated by assuming that there are three zones and

U_j = utility of j^{th} zone,
V_j = "representative utility" of j excluding the attraction term,
A_j = attraction of the j^{th} zone, and

$$U_j = V_j + \alpha \log A_j + \epsilon_j \qquad j = 1,2,3 \tag{42}$$

Then assume $\alpha = 1$, and

$$P_i = \frac{e^{V_i + \log A_i}}{\sum e^{V_j + \log A_j}} = \frac{e^{V_i}A_i}{\sum e^{V_j}A_j} \tag{43}$$

Now suppose that some of the attraction of zone 2 is transferred to zone 1 and all other variables are unchanged. Let ΔA_i be the change in attraction. Will the total share of the two zones remain unchanged? Let P'_1 and P'_2 be the new shares:

$$P_1' = \frac{e^{V_1'}(A_1' + \Delta A)}{\sum e^{V_j}A_j} \tag{44}$$

$$P_2' = \frac{e^{V_2'}(A_2' - \Delta A)}{\sum e^{V_j}A_j} \tag{45}$$

$$V_1' = V_2' \quad \text{(no change in other attributes in zones)} \tag{46}$$

$$P_1' + P_2' = \frac{e^{V_1'}A_1' + e^{V_2'}A_2'}{\sum e^{V_j}A_j} \tag{47}$$

Thus, the sum of probabilities will be invariant to any shift in attraction between zones that are otherwise homogeneous (up to the total amount in each zone since attraction cannot be negative).

It may be proved that the elasticity of a destination's choice probability with respect to the attraction variable approaches 1.0 as the probability approaches zero and approaches zero as the probability approaches 1.0 if the coefficient of the log of attraction is 1. (This is a direct consequence of Eq. 30.)

The SIGMO study (*48*) represents one attempt to apply this consistent approach to the destination modeling problem. Interestingly, the coefficient on the log of the attraction variable was not constrained, but nevertheless was calibrated to be very close to 1.0, which is consistent with the IIA property for destination choice.

Multiple Attraction Measures

In the preceding discussion, it was assumed that only one attraction variable appears in the utility function. It is possible that more than one attraction variable is used. For example, both employment and population are used as attraction variables in the Metropolitan Transportation Commission nonhome based trip distribution models (*49*). Using analysis similar to that shown in Eqs. 42 to 47, Eq. 42 is generalized to the case of more than one attraction variable as follows:

$$U_j = \overline{V}_j + \log \sum_{i=1}^{I} a_i A_{ij} + \epsilon_j \quad (48)$$

in which

A_{ij} = the i^{th} attraction variable for the j^{th} zone; and
a_i = the corresponding coefficient.

It should be noted that Eq. 48 is not linear in parameters. Therefore, modification of existing MNL software would be necessary for estimating a model of this form.

Ben-Akiva et al. (*49*) propose that a utility function of the following form also results in appropriate aggregation.

$$U_j = \overline{V}_j + \log A_{1j} + \sum_{i=2}^{I} a_i \log A_{ij}/A_{1j} \quad (49)$$

They indicate that the first attraction variable, A_{1j} (population in their models), is analogous to the single attraction variable discussed earlier. The remaining attraction variables, divided by the first variable, are viewed as similar to the level-of-service variables in $\overline{V}_j$. That is, Ben-Akiva et al. suggest that although the remaining attraction variables might introduce some aggregation error, the resulting errors are not fundamentally different from aggregation errors introduced by the level-of-service variable.

However, as Dunbar (*50*) notes, the utility function in Eq. 49 does not result in consistent aggregation, i.e., the sum of the predictions for two contiguous zones does not equal the prediction for the aggregation of these two zones. This can be shown with an example. Let A_{1j} be population and A_{2j} be employment (with no other attraction variables). Then

$$P_j = \frac{(A_{1j})^{1-a_2} (A_{2j})^{a_2} \exp \overline{V}_j}{\sum_k (A_{1k})^{1-a_2} (A_{2k})^{a_2} \exp \overline{V}_k} \quad (50)$$

Consider two zones with the following characteristics:

$$A_{11} = 100, A_{21} = 0, A_{12} = 0, A_{22} = 100, \text{ and } \overline{V}_1 = \overline{V}_2.$$

Equation 49 predicts no trips for zones 1 and 2. However, since $A_{11} + A_{12} = 100$ and $A_{21} + A_{22} = 100$, the model predicts a positive number of trips for the zone aggregating zones 1 and 2.

Summary

As noted earlier, trip attraction measures in disaggregate choice models are frequently not unlike the measures normally used in aggregation models, such as type and amount of land use (measured by space, etc.), or on levels of activity (population, employment, etc.). The difference is that aggregate models first estimate trip attraction ("trip ends"), which then becomes the independent variable in the trip distribution phase. Disaggregate models directly calibrate distribution using the attraction measure, eliminating the trip attraction phase.

Although both modeling approaches use some index of the magnitude of the opportunities for the trip purpose as measures of attraction, disaggregate approaches also may include measures of the perceived strength of the attraction. Readers interested in a survey of issues in destination choice modeling for aggregate and disaggregate models should refer to Jones (*51*). The author reviews the various measures of attractiveness, such as zonal retail employment in shopping trips or selling space, and subjective or perceived measures of attractiveness. Much of this research, however, has not as yet been integrated into the urban transportation planning process.

Multidestination Tripmaking

Early travel choice models were based on simple trip definitions, e.g., single trip link between home and work on a home-work-home roundtrip. Actual travel patterns, especially for nonwork purposes, can be more complex. In recognition of this fact, a number of recent models of multidestination travel have been developed, both in the multinomial logit context and with other approaches.

In analyzing nonwork travel, an important distinction between simple and multidestination travel patterns is that the latter involve nonhome-based trip links, i.e., trips between two nonhome destinations. Therefore, some choice modeling studies have developed separate destination choice models for home- and nonhome-based trips. For example, Ben-Akiva et al. (*49*) developed separate home-based and nonhome-based nonwork destination choice models for use in the Metropolitan Transportation Commission (MTC) forecasting system. They also proposed procedures for deriving aggregate zonal home-based and nonhome-based trip distribution from the disaggregate models.

Lerman (*52*) and Lerman et al. (*53*) used a somewhat similar approach. They estimated separate joint mode/destination choice models for nonwork trips originating at home and nonhome locations. These models were combined with probability distributions for the amounts of time spent at destinations in a semi-Markov model of trip chaining. The models are applied with Monte Carlo simulation techniques. (Monte Carlo methods involve the generation of simulated data based on information on the distributions of key variables. Examples include: (1) generation of a simulated sample from Census data input; (2) generation of an individual daily travel pattern based on the distribution of times spent at travel destinations and travel choice models; and (3) generation of the distribution of mode choice probabilities in a population for purposes of estimating aggregate mode shares.)

Other approaches to analyzing multidestination travel are not based explicitly on simple trip links. Adler and Ben-Akiva (*54*) developed an MNL model of complex travel patterns. In the model, alternative simple and complex daily travel patterns were defined as the alternatives. The advantage to this approach is that interactions among trips taken during a day are accounted for in the definition of the alternatives. Models based on single links assume temporal independence among travel decisions. The major disadvantage of the approach is the difficulty in defining the alternatives in the choice set. The number and type of possible daily travel patterns is large. This fact inhibits the use of the model for prediction purposes.

Horowitz (*55*) has developed a set of models for multidestination nonwork travel. The models include: (1) an MNL destination choice model; (2) a sojourn frequency model (a sojourn is a simple trip to a nonhome destination); (3) a model of the ratio of tours per sojourn (a tour is a roundtrip originating at home which may involve multiple nonhome destinations; and (4) a model relating vehicle kilometers traveled to the spatial characteristics of tours. The sojourn frequency model was a nonlinear equation that incorporated interactions among sojourns made during different times in a day. Estimation of the tour to sojourn model suggested that transportation and household variables do not explain variations in tours per sojourn, i.e., the model was simply a constant ratio.

Findings on Trip Generation

Research in NCHRP Project 8-13 has confirmed that the logit model in its present state of development is not well suited for calibrating trip generation models. This deficiency was early recognized as a weakness of the 1972 CRA study when the trip generation model performed poorly in an application (see Spear (*56*) for discussion). In Phase I (*1*) of this study, an effort was made to improve the logit trip generation model by correcting the "inclusive price" term to conform to the logit specification by use of the "log of the denominator." However, this correction was not sufficient to deal fully with the problem, namely the IIA property of the MNL appears to be a fatal defect in its use for trip generation.

The weakness of an MNL model for trip generation can be seen by reference to the specification of the logit trip frequency equation, Eq. 51. Because of the IIA, if the choice set is expanded, predicted trip frequency will increase. In a joint destination and frequency model, the new destination will attract equally from the other destinations and the no-trip alternative. In a separable model the inclusive value term will increase with the addition of a new destination alternative. Both of these circumstances caused trip frequency to increase by adding new destination alternatives to the choice set, an undesirable property. This phenomenon implies the following two possible situations where the IIA assumption may be troublesome for forecasting trip frequency:

1. MNL model calibration and prediction for trip frequency will be extremely sensitive to the specification of the relevant destination choice set. For example, the separable frequency model infers the sensitivity of trip frequency to LOS by assuming that a greater number of destination choices generate greater "accessibility" and more trips. Unless the specification of the destination choice set accurately reflects the relevant choice set, cross-section model estimates will be unreliable.
2. The transferability of the model will be highly sensitive to the reasonableness of the IIA assumption as the number of destination alternatives expands. If the model is transferred to a new city with twice as many destinations of comparable utility as the old city, trip frequency will increase due to the IIA. The reasonableness of this relationship remains untested.

The problem may be illustrated by reference to the trip frequency equation in the NCHRP Project 8-13 Phase I report (*1*);

$$\ln \frac{P(\text{frequency} = 1)}{P(\text{frequency} = 0)} = \underset{(1.37)}{0.14 P_f} + \underset{(1.79)}{0.071\, INC} + \underset{(1.93)}{0.77\, DMW} \tag{51}$$

in which

P_f = inclusive price of travel for frequency = 1;
INC = income code; and
DMW = number of licensed drivers minus number of workers in the household.

Despite the fact that the model appeared reasonably to predict frequency for the calibration sample, the results appear to be highly suspect as a candidate for a transferable model because the attraction variable for destinations, which enters into the inclusive price term, incorporated a nontransferable definition inconsistent with the logit model, as discussed above.

The sensitivity of trip frequency to the number of destinations can be illustrated with a simple example, using a separable model. Suppose that there are three destinations with equal attractiveness and utility V_d. The inclusive value variable will be $\ln 3e^{V_d}$. If the number of destinations doubles, the effect on trip frequency will be due to an increase in inclusive value to $\log 6e^{V_d}$. The effect on trip frequency will depend on the probability of taking a trip. The effect will be greatest when there is a low probability of taking a trip in the three destination case. (This is because of the elasticity properties of the logit model.)

Clearly, if the IIA property does not apply, and the trip frequency model should not depend arbitrarily on the number of alternatives specified for destination choice, some procedure for "normalizing" the inclusive price term should be applied. While this approach is intuitively attractive, its theoretical justification in terms of the logit model specification is not obvious.

The weaknesses of the discrete choice models of the logit type for the trip frequency (or trip generation) equation has generally necessitated the use of other model specifications. For example, traditional cross-classification and linear regression models (see SIGMO (*48*) for example) have been used in combination with logit models of the other choices. These other modeling approaches have their drawbacks as well. First, it is difficult to incorporate a measure of accessibility or "generalized cost" into the model, especially for the "no-trip alternative." (What was the trip not taken whose accessibility is to be measured, and how is accessibility to be measured for grouped data?)

A second problem is that of the "limited dependent variable." In travel diary data such as that used to calibrate most disaggregate travel models, there is a preponderance of zero values. For instance, many or most respondents report no shopping trips. Consequently, the data are clustered at a limit of the distribution of the dependent variable. This clustering at the limit of the distribution violates one of the principal assumptions of ordinary least squares (OLS) regression. OLS assumes there is no relationship between the values of the dependent variable and the error term of the estimated equation. The consequence of ignoring this assumption is that the OLS parameter estimates have a probable bias toward the concentration of values at the limit. The relative impact of the independent variables on the dependent variable is underestimated in those cases were the dependent variable is above the limit. (See Tobin (*57*); McKelvey and Zavonia (*58*).)

This general state of dissatisfaction with available disaggregate modeling techniques has led to an effort to develop alternatives (see for example, Ruygrok and Van Essen's presentation of the Poisson model (*59*)). Nevertheless, the matter is presently in a state of development and considerable differences exist regarding the most appropriate methodology. A maximum likelihood technique developed by Tobin (*57*) to model consumer behavior in buying consumer durables is tailored for estimating a model where the dependent variable is limited from taking an unbounded range of values. The technique, which has shown to be a special case of a more generalized method of solving single equation relationships with limited dependent variables (Johnson (*60*)), iteratively solves for the vector of coefficients which maximizes the log likelihood subject to the constraint of the lower limit of zero on the dependent variable and the constraints imposed by the data. Under fairly general conditions it can be demonstrated that the resulting maximum likelihood estimates of the coefficients are consistent, asymptotically normal, asymptotically unbiased, and more efficient than any other estimator (see Beals (*61*)). This approach can be used to model choice behavior for sporadic events such as shopping trips.

Sheffi (*62*) has developed a model for estimating choice probabilities among alternatives that can be naturally rank ordered. Examples include trip generation (number of trips) and household auto ownership. The key behavioral assumptions are that: (1) selection of a higher ranked alternative implies that lower ranked alternatives have been previously chosen, e.g., a household that has made three trips has previously decided to make one and two; (2) the utilities of alternatives increase monotonically up to the chosen alternative; and (3) the utilities of alternatives ranked higher than the chosen alternative are smaller than the utility of the chosen alternative. In order to facilitate analytical tractability, Sheffi also assumes that differences in utilities are uncorrelated.

These assumptions and the standard logit assumption on the distribution of the random component of the utilities result in the choice probabilities being a product of binary logit choice probabilities. Specifically,

$$P_i = p\,(i > i + 1) \prod_{j=1}^{i} p\,(j > j - 1) \tag{52}$$

in which

P_i = the probability that the i^{th} ranked alternative is selected;

$p\,(j > j - 1)$ = the binary logit probability that the j^{th} alternative is preferred to the previous alternative;

$p\,(i > i + 1)$ = the binary logit probability that the i^{th} alternative is preferred to $i + 1$ alternative; and

$\prod_{j=1}^{i}$ = the product of P_j from 1 to i.

A standard binary logit computer program can be used in estimating a model of this type. In the estimation, each individual or household can be thought to contribute $i + 1$ "observations." Each "observation" is one of the binary comparisons in Eq. 52. Sheffi used the procedure in estimating a trip generation model for households with elderly members.

Sheffi's model is an attractive alternative to multinomial logit and linear regression for problems in which alternatives can be rank ordered. In particular, it does not have the IIA property, which implies, for example, that if a household that currently makes four trips is restricted to three or fewer trips, the choice probabilities for these lower ranked trips increase proportionately. (It might be expected that the probability of making three trips increases much more.) The model also can be viewed as a solution to the truncated distribution problem of linear regression.

It should be noted, however, that Sheffi's model does not address the problem of the no-trip alternative in a joint trip frequency, trip distribution model. In its present form, it is a model that can address only problems where alternatives can be rank ordered. The joint trip frequency, trip distribution problem combines rank ordered and nonranked alternatives. Therefore, the problems of trip generation that result from the IIA property of the joint model cannot be solved with Sheffi's approach.

In closing, it is worth noting that the same problems of IIA that plague trip generation will also often plague models of auto ownership. If a family prefers two cars to one car for unobserved reasons, it probably also will prefer one car to no car for the same unobserved reasons.

FINDINGS ON DATA COLLECTION

Initial research on disaggregate demand models was heavily oriented toward reducing the heavy data requirements for estimating aggregate models. Since those early expectations, considerable advances have been realized in knowledge of data requirements for demand modeling. The promise of disaggregate models for reducing data requirements has been largely realized. However, application of disaggregate models has demonstrated that accurate forecasting can be quite voracious in its appetite for data. Although savings on data collection costs can be realized for model calibration, there is an increasing awareness that accurate forecasting may require detailed data on the underlying distribution of the explanatory variables.

Confidence intervals for the estimated disaggregate model parameters are readily available from the model calibration programs. Reasonably "tight" intervals often can be achieved with relatively small data sets (say 200 observations), a result which has been the basis of much of the purported benefits of disaggregate models. Confidence regions for the forecasts of choice probabilities are more complex because the relations are nonlinear. Horowitz (*63*) proposed several means of estimating these regions for the logit model, and Daganzo (*64*) proposed procedures for the probit model. Researchers have reported fairly large confidence intervals for some forecasting applications, a caveat that should be considered in applying the results of a disaggregate model to forecast actual splits. Since the magnitude of the error is dependent on the square root of the sample size, fairly large increases in sample size may be required to appreciably affect the magnitude of the error.

There is no one answer to the issue of the sample size required for model calibration. Richards and Ben-Akiva (*65*) have estimated that a sample size of 200 to 500 is reasonable, but samples as small as 50 to 70 have produced reasonable results. A discussion of the advantages and disadvantages of different data collection approaches is contained in U.S. Department of Transportation (*66*). Given the expense of data collection, it is highly likely that most applications of disaggregate models would require the use of existing data collected for the purpose of using aggregate models.

The third phase of this project focused heavily on the analysis of an experimental data set expressly designed for the calibration of disaggregate travel models. One of the principal findings of this work relates the design of data sets for disaggregate demand model calibration. Based on this experience, the collection of omnibus data sets for the analysis of a wide variety of choice situations is not endorsed. CRA's experience with the Baltimore Disaggregate Data Set is described in detail in Appendix D. In order to achieve the widely advertised claims of economy in data collection and manipulation, the data sets for model calibration should be of a more modest scope, limited to only one or two choice dimensions. Surveys that focus on many choice situations are costly and troublesome to administer and manipulate for model calibration.

"Choice-Based Samples" as a Means of Economizing on Data Costs

Research on the use of "choice-based samples," i.e., samples where the probability of sample selection depends on the actual mode or destination selected (such as an on-board survey or screenline survey), by Lerman et al. (*67*) has improved the usefulness of existing data. Home interview survey (HIS) data are normally stratified on the independent variables of demand models. As a result, no special weighting of the observation is needed. Normally, such stratification is done with the objective of improving the efficiency of the estimate.

A choice-based sample is stratified by the dependent variables. For instance, a choice-based mode choice sample could consist of transit riders and auto drivers. However, unless proper estimation procedures are followed, simply calibrating a logit model on two choice-based subsamples will not yield valid parameter estimates.

Lerman et al. provide a mechanism for splicing two or more choice-based samples. Such a mechanism is particularly interesting because on-board surveys can mitigate many of the problems (e.g., low response by certain modal segments) and costs of HISs. One caveat, however, is that the home interview may provide valuable information on the distribution of the explanatory variables, which is critical to aggregation and forecasting. These data would have to come from another source in the absence of an HIS. Further, the splicing mechanism proposed by the authors requires knowledge of the aggregate splits, which may not be known without an HIS. This is not to say, however, that a hybrid choice-based approach is not feasible.

Another significant advantage of choice-based sampling procedures accrues in instances where one or more important alternatives are rarely chosen by the general population. For instance, if only 4 percent of the population is in a corridor carpool, a random sampling approach to collecting data for a ridesharing model would not be efficient because many household surveys would be required to yield a sufficient number of carpoolers. It would be more data-efficient to stratify the survey on the basis of mode choice and to sample more intensively from the carpool population.

The weighting factor recommended by Lerman et al. (*67*) is the "probability of that choice from a random sample/probability of that choice from the choice-based sample." This finding is based on the use of Bayes Theorem. This approach is somewhat similar to the Atherton and Ben-Akiva (*68*) approach for using a small sample to update a transferred model calibrated on data from another area.

It should be noted that this weighting procedure is generally applicable to choice models, not just the MNL. In particular, it can be used with the MNP model. Daganzo (*17*) discusses the use of choice-based sampling with the MNP.

This approach calls attention to the fact that more sophisticated sampling designs could pay big dividends in travel forecasting. Simple random sampling does not provide sufficiently large numbers of observations of rare events (e.g., transit ridership), thereby reducing the efficiency of the estimation process. A number of interesting research projects have recently addressed the issue. Before beginning data collection, the researcher should consult these reports (see CRA (*69*) and Daganzo (*70*)).

Special Data Sets—Repeated Trials, Before-and-After Data, and the Ordered Logit Model

Special estimation problems arise where the researcher has more information than just the choices at an observed single point in time. The proper estimation technique for "repeated trials" using the same levels of independent variables has been addressed by McFadden (*14*). Daganzo and Sheffi (*71*) have addressed the issue of panel data, i.e., sequential sampling of individuals over time where the independent variables vary, such as with "before-and-after data."

Beggs et al. (*72*) addresses the issue of how to use information, not only on the most preferred alternative, but also the order of preference among alternatives using an "ordered logit" model in an application of disaggregate modeling to demand for electric automobiles. The ordered logit model uses ranked data by respondents on their order of preference for an array of choice alternatives. One of the advantages of this technique is its high efficiency and relatively low data costs. On the other hand, its predictive validity has not yet been demonstrated. Daganzo (*17*) presents the MNP analog to the "ordered logit" model, although this model apparently has not been used empirically.

Comment on Data Requirements for Implementation and Forecasting

In order to understand the data required for implementation and forecasting, first consider the steps in forecasting using the disaggregate approach and the "market segmentation" method of aggregation which is described in the next section:

1. Identify the population whose choices are to be forecasted as a result of the policy change.
2. Disaggregate the population into market segments according to the relevant choice set, levels of independent variables, differences in valuation of choice attributes and other segmentation criteria as required by the behavioral model. Identify the level of the explanatory variables for each of the market segments.
3. Calibrate a disaggregate model on the subject population or adjust a model calibrated on other data to account for nontransferability, if any, between the calibration population and the forecast population.
4. Modify the explanatory variables to reflect the policy change and evaluate the impact on the selection probabilities.
5. Aggregate the forecast selection probabilities to obtain estimates of aggregate shares for each of the choice alternatives.

The list looks rather forbidding, but Appendix C illustrates a method for using disaggregate models with data typically available to urban transportation planners.

DIAGNOSING AND CORRECTING ERRORS IN THE MNL MODEL

Sources of Errors in Disaggregate Models

Two recent papers by Horowitz (*73, 74*) have provided an interesting survey of sources of error in estimating disaggregate models, methods of diagnosing these errors, and techniques for correcting the errors. Although the list of potential errors, and their consequences if uncorrected, is rather discouraging ("large enough to destroy the practical value of a model"), considerable progress has been made in diagnosing and correcting these errors. In principle, the errors are generally susceptible to correction, although in some cases the costs of doing so can be quite large. The potential sources of errors reviewed by Horowitz may be catalogued as follows:

1. Choice of the wrong model form or behavioral assumption (i.e., the assumption of utility-maximizing behavior when travelers actually respond to hierarchies, thresholds, etc.).
2. Statistical sampling errors in estimation.
3. Inclusion of irrelevant explanatory variables.
4. Violation of the Independence of Irrelevant Alternatives (IIA) property of the MNL model (the blue bus/red bus problem).
5. Assumptions of fixed parameter values in a population that is characterized by taste variations.
6. Omission of behaviorally relevant variables.
7. Errors in data, especially use of zonal average LOS data rather than disaggregate data.

Horowitz concludes from his tests that the most important errors, in order of seriousness, are: (1) item 7—zonal averages for LOS; (2) item 6—omission of behaviorally significant variables; and (3) item 5—taste variations (random parameters). Curiously, the IIA problem was found to be less worrisome, at least as far as estimation of model parameters is concerned. According to the author's findings, violation of the IIA as a result of the blue bus/red bus problem is likely, for many applications, to cause errors comparable to or smaller than errors resulting from ordinary sampling error with samples of less than 1,100. Where the IIA error is significant enough to affect the results, it appears to be large enough to be detected. The effect of omitting a behaviorally relevant variable from the specification depends on the relationship between the omitted and included variables (see Tardiff (*30*)). If a relevant but omitted variable is correlated with included variables, the estimated coefficients for the included variables will be biased. If the omitted variable is not related to other explanatory variables, there is no bias but rather less precision in the predictive poser of the model. The author finds that improved data and model specification (such as the random parameter probit) can correct the truly serious errors. As with any model, it appears that the user should be fully aware of the types of errors that can occur, the means of detecting them, and procedures for correcting the model to account for the problem.

The consequences of a correlation between an observed attribute and an unobserved attribute are discussed in detail in Appendix E. There it is shown that the result is a bias in the estimated model coefficient. For example, the appendix notes that if people with tastes predisposed to transit (the unobserved attribute) locate near transit lines, time and cost (the observed attributes) elasticities will be biased upward. Chan's study (*75*) of elasticities indicates that time series elasticities are about one-half those of cross section, a finding that may be explained by the cross-section bias resulting from correlation of the unobserved attributes and independent variables.

Methods of Diagnosing IIA Errors in Logit Models

A new approach to addressing the IIA problem in the MNL model was pioneered in Phase I of NCHRP Project 8-13 (*1*). A great deal had been said in the literature about the serious consequences of a violation of the IIA assumption, but very little had been done to evaluate how serious the problem is in practice or to provide a means to detect situations where the IIA was unreasonable. To this end, a serious effort was made first to itemize the various sources of a violation of the IIA and to develop diagnostic tests to permit the practitioner to determine whether a violation has occurred. The results of this research are included in Appendix D and in the Phase I report.

Since that report, Horowitz (*73*) has evaluated alternative means of diagnosing errors in the logit model. The following tests have been proposed to diagnose violations of the IIA:

1. Tests based on conditional choice proposed by McFadden, Tye, and Train (*76*) and in CRA's Phase I NCHRP Project 8-13 report (*1*).
2. Tests based on the universal logit model proposed by McFadden, Tye, and Train (*76*).

3. Tests based on extrapolation, first discussed in CRA's Phase I report (*1*) and developed and applied by Horowitz (*73*).

4. Likelihood ratio tests to distinguish between the logit and probit models and tests of significance of the estimated parameters in the probit model that measure departures from IIA (see Horowitz (*77*) and Hausman and Wise (*78*)).

In a very recent paper, Hausman and McFadden (*79*) have made two new theoretical contributions to diagnosing violations of IIA. First, they have developed a test statistic for the tests based on conditional choice (no such statistic was developed in Phase I of this project). Second, they have proposed statistical tests for distinguishing between the logit model and the nested logit model, which is described in detail later in this chapter.

The assumption that the unobserved, or random, components of utility are independent and identically distributed (IID) with the Weibull distribution is necessary and sufficient to derive the MNL model (see App. D for details). Therefore, all violations of the logit model assumptions also will violate the IID assumption and the IIA property. Therefore, errors in applying the MNL model often may be detected as a violation of the IIA property and corrected by making alternative assumptions regarding the random utility functions.

Horowitz's (*73*) findings were that the most powerful tests are a test against a probit model and the McFadden, Tye, and Train (*76*) test against the universal logit model. The tests based on conditional choice were found to be variable in their power to detect violations, but their intuitive appeal remains. Horowitz's results represent an extension of the diagnostic tests developed in Phase I of NCHRP Project 8-13 (and reported in an appendix of this report).

Williams and Ortuzar (*80*) use a somewhat similar approach in analyzing possible model specification errors. In addition to potential violations of the IIA, they also examine specification errors in Horowitz's (*73*) first category. In particular, the authors consider three types of decision rules other than utility maximization over a common set of attributes and alternatives: (1) a decision rule using only a subset of possible attributes; (2) a decision rule using only a subset of possible alternatives; and (3) a decision rule hypothesizing habitual choice behavior, i.e., choice of an alternative increases the probability of choosing that alternative in the future.

The approach used to analyze these potential specification problems involves three steps: (1) generation of simulation (Monte Carlo) data reflecting a choice process other than MNL; (2) estimation of an MNL model using the simulated data; and (3) comparison of the predictions of the resulting MNL model with the predictions of the "true" model.

This approach can indicate the conditions under which the MNL is a reasonable approximation of alternative decision rules. Not surprisingly, prediction errors can be quite large in some cases. For example, the authors find that the MNL, which is based on tradeoffs among attributes, does not produce good predictions when the "true" decision rule is lexiographic; i.e., choice is based on dominance on a single attribute. Like Horowitz, the authors conclude that the IIA need not be a serious problem because there are relatively straightforward modificiations to the MNL that mitigate the problem. (The most prominent modification is the nested logit model, which is described later in the chapter.)

As a result of concern about the IIA property and other consequences of violation of the logit model assumptions, research has focused on two approaches to address the problem. The first is the diagnostic test approach first developed as part of Phase II of the present project and applied in McFadden, Tye, and Train (*76*). This approach focuses on continued application of the logit model with safeguards designed to detect departures from the model's assumptions. This approach has been further developed by Joel Horowitz (*73*).

The alternative approach, pursued by Hausman and Wise (*78*), Albright, Lerman, and Manski (*81*), Bouthelier and Daganzo (*82*), and McFadden's (*83*) generalized extreme value (GEV) model, is to develop more general models that do not employ the IIA assumption. The most widely applied alternative is the multinomial probit model with both taste variations across the population (variable coefficients) and dependent random utility components (red bus/blue bus).

Continued Application of the Logit Model

The IIA property is the result of three MNL assumptions: (1) the model coefficients are fixed, i.e., there are no taste variations; (2) the random components of utility are independent across alternatives; and (3) the random utility components are uncorrelated with the observed attributes. If diagnostic tests indicate that the IIA does not apply, a possible strategy is to respecify a logit model so that the violations of the three assumptions are remedied.

Taste Variations

Taste variations can be addressed in two ways. First, socioeconomic characteristics can be included in the utility function as described earlier. Second, the estimation sample can be segmented based on socioeconomic and other characteristics and separate choice models can be estimated for each market segment. Taste variations are represented by different values for the model coefficients for different market segments. This approach has been used by Stopher and Lavender (*37*), Recker and Golob (*84*), and Kitamura (*85*).

Correlation of Unobserved Attributes

Violation of the second assumption occurs when the unobserved characteristics of two or more alternatives are correlated. The "nested logit" model, which is discussed in detail later, has been proposed as a means for remedying violations of the second assumption. For example, Sobel (*86*) applied nested logit analysis to a mode choice model with six alternatives: walk, bicycle, moped, transit, auto driver, and auto passenger. In one specification, the first three modes were assumed to be similar in their unobserved characteristics; i.e., the random components of their utility functions were assumed to be intercorrelated. A logit mode choice model was estimated for these three modes and an inclusive price variable was constructed. This variable was included with the characteristics of the other three modes in a second MNL model. A second specification also grouped the two auto modes as similar alternatives. Sobel (*86*) provides more details on these examples and the procedures for estimating the nested logit model.

Although the nested logit model is an alternative to the MNL model (the MNL model is a special case of nested logit in which the coefficient(s) of inclusive price are equal to one), it can be estimated with existing or modified logit software. Therefore, it is a practical approach for addressing possible correlations among the random components of utility functions.

Correlations Between Observed and Unobserved Attributes

Violation of the third assumption is a source of bias common to almost any statistical modeling. For example, this type of correlation is a potential source of bias in standard regression analysis. Two strategies, which are discussed in more detail in Appendix E, can be used to overcome this problem. If the analyst has some idea of the variables that might be correlated with currently observed variables, he could measure them and add them to the model. For example, if comfort is correlated with travel time, the analyst may try to measure this variable by using attitudinal scales or physical proxies for comfort. Alternatively, possible correlations between observed and unobserved variables may be eliminated through the collection of data after a change in the transportation system. As noted above, if people with a favorable bias toward transit tend to locate near bus routes, a typical mode choice model is likely to be biased because of this pattern of correlation between observed level of service and unobserved transit bias. However, people located next to a newly established route may be less likely to have a transit bias, thus removing the correlation between observed and unobserved characteristics.

The Alternative Approach: Model Forms That Do Not Require the IIA Assumption

A number of alternative model forms have been proposed that do not require the IIA assumption. These alternatives include the multinomial probit model, the CRA hedonics model, the nested logit model, the generalized extreme value model, and the dogit model. Each of these modeling techniques is discussed below.

The Multinomial Probit Model

The multinomial probit (MNP) model has been proposed as a means of correcting for the IIA property of the multinomial logit (MNL) model. The fact that MNP possessed the potential for applications that do not require the IIA assumption has been recognized, but computational difficulties have been perceived as prohibitive. Since the MNP model is based on the normal distribution, the properties of this distribution are used in the development and extension of the model. Of special importance is the fact that the sums and differences of normally distributed variables are also normally distributed. This property implies that the MNP structure is quite flexible, i.e., various assumptions underlying a particular model can be modified in such a way that the resulting model is still MNP.

Daganzo (*17*) explores many such modifications of the MNP. In many cases, these modifications are analogous to applications of the MNL. They will be described when the parallel MNL applications are presented.

A major advance in treating violations of the IIA emerged with the development of the Conditional Probit model by Hausman and Wise (*78*). The Hausman and Wise approach developed a computationally feasible method for calibrating a probit model that allowed for both a correlation in the unobserved attributes of the alternatives and variation in tastes across individuals. The approach allows three alternatives. The authors also state that extension to four or five alternatives is feasible.

The research of Daganzo, Bouthelier, and Sheffi (*87*) produced a computationally more efficient algorithm for computing the MNP parameters, using the "Clark approximation" method. This breakthrough was significant because it permitted a method of computing the MNP model with more than three alternatives. Lerman and Manski (*88*) have reported on a computer program that was also designed to overcome the computational difficulties. This program also incorporated the Clark method.

Despite the early encouragement that the Clark approximation would provide a significant breakthrough in disaggregate methodology, some experts have expressed reservations. For example, Horowitz (*89*) presented examples in which the Clark approximation produced fairly large errors. Concern has been raised that the method has produced unexpectedly large errors in estimation and forecasting for large numbers of choice alternatives (see App. F).

Random coefficient models are not confined to the disaggregate, discrete choice framework. Johnson and Hensher (*90*) have applied the concept to a regression model of shopping trip frequency.

The principal advantage of the MNP model over the MNL is that it allows for both variations in tastes across individuals and correlation of unobserved attributes across alternatives, both of which lead to violations of the IIA property of the MNL (see App. E). In the course of calibrating the MNP model, the planner is provided with estimates of the variance in tastes and the extent to which tastes are correlated and estimates of the extent to which the IIA is violated as a result of correlation in unobserved attributes.

The CRA Hedonics Model

A version of the MNL model recently developed at CRA treats the individual utility function parameters being estimated as random variables. This model, the CRA hedonics model, assumes that the marginal utilities of the attributes of the alternatives vary across individuals. This treatment is similar to one originally used by Quandt and resembles current work on the multinomial probit model being performed by a number of researchers.

Appendix F of this report compares and contrasts the MNL, MNP, and CRA hedonics models. The MNL model is useful in analyzing a large number of transportation problems, but its applicability depends on a rather strong independence assumption. The CRA hedonics model effectively relaxes what may be an important aspect of this assumption, since it allows for taste variations across individuals. No independence assumptions are required for the MNP model. However, the CRA hedonics model, when its assumptions are satisfied, is more flexible than the probit model in that it does not require the assumption of a normal distribution of the taste parameters. On the other hand, the probit model has the advantage of an explicit treatment of dependence among alternatives.

Nested Logit or Structured Logit

As mentioned earlier, Williams and Ortuzar (*80*) concluded that the nested logit model appears to be an effective and practical approach to mitigating potential IIA problems.

The "nested logit" model is the "hierarchical," "sequential," or "separable" logit model, which was first analyzed in depth by Ben-Akiva (*91*), Domencich and McFadden (*92*), and CRA (*13*). Ben-Akiva first pointed out that the "inclusive value" term, or "log of the denominator," would have a coefficient of 1.0 if the mode and destination choice were a joint multinomial logit model. A coefficient of between 0 and 1 results in the nested logit model. The importance of this finding was made more widely known at the Second International Conference on Behavioral Modeling (*93*) in the workshop report on quantitative methods. Readers interested in the technique should consult Sobel (*86*) for a discussion directed to the nontechnically oriented readers.

Intuitively speaking, the nested logit model is analogous to the use of the "inclusive value" term in specifying a joint mode and destination choice logit model. The concept can best be understood by reference to a destination choice model in which the representative utility of a destination contains a term representing the "inclusive value" of all modes to that destination:

$$V_d = \ldots + \theta I_d + \ldots \qquad (53)$$

in which

V_d = expected utility of d^{th} destination;
θ = coefficient of "inclusive value" of all modes to that destination; and
$I_d = \ln \sum_{m=1}^{M} V_{md}$; and
V_{md} = expected utility of the m^{th} mode to the d^{th} destination.

It can be seen that the inclusive value of modes to a given destination is in fact the "log sum" or "log of the denominator" from the previously estimated mode choice (given destination) equation. As noted above, Ben-Akiva (*91*) demonstrated the very important finding that constraining θ to 1.0 produced the equivalent of a joint logit model of mode and destination choice. It can be demonstrated that θ must be constrained to $0 < \theta \leq 1$, or irrational behavior will result. For a proof of this assertion, see McFadden's paper in Hensher and Stopher (*94*).

For example, if $\theta < 0$, it would imply that improved service to a destination would increase the inclusive value of that destination but reduce the likelihood of choosing that destination. If $\theta = 0$, improved service to a destination would have no effect on the destination's choice. If $\theta = 1$, as noted above, the general MNL model applies to a joint destination/mode model. If $\theta > 1$, the model overpredicts the effects of improved service by one mode to a destination (relative to the IIA assumption). This may imply diversion (away from other destinations) to modes to that same destination that were unaffected by the service improvement (i.e., $\theta = 1$ would cause equal percentage diversions from all modes and destinations not affected). In the language of economics, $\theta > 1$ may imply that two modes to the same destination were complements rather than substitutes (the cross elasticities may have the wrong sign).

In an earlier version of the McFadden paper it was stated that in a nested structure containing more than two layers, e.g., a frequency, destination, mode choice structure, the coefficients of inclusive price should not decrease as one moves up to higher layers. McFadden has since shown that the condition is not necessary. The only requirement for inclusive price coefficients is that they lie in the 0 to 1 range. The final version of the paper has been modified to report this condition. On page 49 of the Phase II report (*2*), the erroneous version of the condition on inclusive price was reported, which is now corrected.

The nested logit model (first proposed by Brand (*3*) and described in detail by Sobel (*86*)) is not merely a device for structuring a separable model such as destination and mode choice. Sobel proposes that the method be used to break up the IIA problem by first applying MNL to the modes thought to have dependent unobserved attributes, e.g., bus and train. This model is then used to develop an estimate of the "inclusive value," or expected utility of the "nest," which is then used as an alternative-specific variable in calibrating the choice between the remaining alternatives. Sobel also reminds the modeler to include attributes that are common to the nest but vary among the nest and other alternatives in the higher level estimation.

The nested logit model clearly has appeal in addressing structural choice situations, e.g., mode/destination/frequency decisions. As a technique for addressing a correlation in the unobserved attributes of alternatives, the technique may be less attractive than a model that does not require a priori specification of the structure of dependence, such as the MNP.

Generalized Extreme Value Model

McFadden's (*83*) GEV model is an important generalization of the nested logit model. The GEV model has not, as yet, been applied to an actual transportation planning study.

The Dogit Model

The dogit model, specified by Gaudry and Dagenais (*95*), incorporates a format somewhat similar to logit:

$$P_i = \frac{e^{V_i} + \theta_i e^{V_j}}{(1 + \sum_j \theta_j) \sum_j e^{V_j}} \quad (54)$$

in which

P_i = probability of the i^{th} alternative;
V_i = i^{th} alternative's systematic utility; and
$\theta_i \geq 0$.

It can be shown that the dogit model is a special case of logit when a number of fairly restrictive assumptions on choice set formation hold. Under certain circumstances (when the respective θ_i's are 0), the model collapses to the MNL for a subset of alternatives. The model has an apparent intuitive appeal in that the total share of an alternative can be shown to contain a fixed or captive share, irrespective of attributes allocated according to the logit model. These elements are represented by the first and second terms in the numerator of Eq. 54. This interpretation is only suggestive because a single choice for an individual is discrete and does not represent an allocation process. However, if repeated choices are made, it may be reasonable to consider an allocation process.

Although the dogit model may be appealing as a behavioral model in certain very special circumstances, it is not offered as a general remedy for violations of the IIA. Indeed, the most common causes of the violation of the IIA are likely to be the blue bus/red bus problem and the presence of taste variations in the population, neither of which is explicitly addressed by the dogit model. (In fact, dogit assumes a logit model to explain the discretionary share.) Unless the violation is diagnosed to be a result of the "captive mode" theory, the practitioner is advised to deal with models such as the MNP that explicitly address the likely sources of the violation.

Empirical Evidence on the Validity of IIA

It will be recalled that for a particular data set and model specification, two types of approaches have been used in testing the validity of IIA. McFadden et al. (*19*) used the diagnostic tests of McFadden, Tye, and Train to examine the mode choice models estimated with data in the BART service area. Most tests indicated that the IIA property appeared to yield reasonable results.

The second approach is to estimate a more complex model such as the MNP model and to compare it with an MNL model (note that this is one of the diagnostic tests suggested by Horowitz). Hausman and Wise (*78*) and Albright et al. (*81*) performed this test on worktrip mode choice models estimated with different samples from the 1968 Washington, D.C., data base. Both studies showed that the MNP model did not fit the data substantially better than did the MNL model. These findings are consistent with those of McFadden et al. (*19*) on the reasonableness of the IIA property for worktrip mode choice models. However, Hausman and Wise (*78*) showed that although the alternative models may fit the calibration data set equally well, they can yield substantially different forecasts of modal shares resulting from the introduction of a new mode. This finding suggests that a data set other than the calibration data set may be necessary in selecting among models and in completely examining the validity of the IIA. Development of tests involving independent data sets (which would be an extension of Horowitz's extrapolation test) might be a useful topic for future research.

Summary

Experience has shown that there are numerous errors that can substantially affect the accuracy of the disaggregate modeling approach. The preceding discussion has focused on specification errors and alternative modeling forms that do not require the MNL assumptions. These alternative model forms require increasingly sophisticated understanding on the part of the user and additional computational and data requirements for calibration and forecasting in most situations. Therefore, it is recommended that the diagnostic tests such as those specified in Appendix D and in McFadden, Tye, and Train (*76*) and Horowitz (*73*) be applied to ensure that these costs are warranted before departing from the more familiar logit format.

AGGREGATION OF DISAGGREGATE MODEL FORECASTS

There is presently a dispute regarding the most effective method of aggregating the results of disaggregate models. The following principal methods have been proposed:

1. Random sample enumeration—a random sample of decision makers is used or may be produced by Monte Carlo computer generation of a synthetic sample or pseudosample, based on knowledge of the distribution of underlying variables.
2. Mathematical integration of the predictions over the distribution of independent variables.
3. Classification, or market segmentation, where the population is classified into cells, a forecast is made for each cell, and the cell forecasts are aggregated.
4. The "naive" or "direct" aggregation approach, where disaggregate model results are applied directly to aggregate data.

Each of these approaches has been employed in actual transportation planning and research. The Chicago Area Transportation Study (*96*) has employed the "pseudosample" approach. The most obvious difficulty in employing this approach is, of course, the interdependence among the variables (for example, the socioeconomic variables are correlated and not independently distributed from the LOS variables, e.g., higher income workers may take longer trips).

The "random sample enumeration" technique is the preferred technique where the researcher has access to a disaggregate sample. The method simply makes forecasts of logit shares using a random sample and aggregates across the sample, effectively eliminating aggregation error. Atherton et al. (*97*), for example, used this technique for evaluating the impacts of carpooling incentives, and the Berkeley Travel Demand Forecasting Project (*6*) used it for estimating BART mode shares.

The paper by Bouthelier and Daganzo (*82*) is an important advancement in aggregation methodology. The authors have been creative in expanding the initial efforts of McFadden and Reid (*32*) and Westin (*98*). They have developed practical computer approaches to performing aggregation, or disaggregate models, to produce aggregate forecasts and to calibrate the MNP model dependence across alternatives (the red bus/blue bus problem). Users should note, however, that its value depends on the reasonableness of assuming a normal distribution for most of the explanatory variables and assuming a significant error in using the classification (or market segmentation) approach, with which transportation planners already have experience.

Appendix B addresses in greater detail this review of aggregation methodology, and Appendix C details the CRA method for implementing the market segmentation technique. The greatest attention has been given to market segmentation because it resembles the technique most familiar to planners—cross-classification.

FINDINGS ON TRANSFERABILITY OF LOGIT TRAVEL DEMAND MODELS

The analysis of urban transportation policies would be greatly simplified if a single disaggregate model, calibrated at one place and time, could be used for the evaluation of the current and future impacts of policies in many different places. Results presented in Appendix B indicate, however, that careless attempts to transfer and aggregate logit travel demand models could lead to sizable errors in the predictions of policy effects. This potential for error has stimulated research into quantifying the magnitude of the error introduced in applying logit models and devising procedures for reducing the errors. This section reviews the findings of the research and compares and evaluates the methods that have been proposed for transferring disaggregate demand models.

Previous Research on Transferability of Mode Choice Models

Previous works dealing with the transferability problem have sought to answer the questions: Can logit travel demand models be transferred from one city to another without modification? If not, are modifications short of complete recalibration of the models feasible? The answer to the first question appears to be: "No, at least not in general," while the answer to the second question appears to be "Yes, under some circumstances."

The first step in understanding the transferability problem is to identify the reasons why a model would not be transferable, i.e., reasons why a transferred model is a poor predictor in a new forecasting environment or why two models calibrated on different data sets produce entirely different estimates of behavioral parameters. These reasons include the following:

1. *Model specification differences which may or may not reflect true behavioral differences*—Two models may be specified differently even though they purport to forecast consistent behavior. For example, one model may include income; another may include the wage rate; others may include income as a separate additive term; and another may divide it into the cost term. Variables may not be defined consistently. For example, one model may use network (aggregate) level-of-service data; another uses perceived data; a third uses individually measured portal-to-portal values. Model coefficients may also vary with changes in the cost of living.

2. *Differences in sampling procedures*—Differences in sampling procedures can affect the model coefficients. For example, the corridor sampling in the 1972 CRA study in Pittsburgh produced a mode-specific constant with a "transit bias." That is, the model predicted that more persons would choose transit than auto when the independent (explanatory) variables were identical for both modes. Thus, it clearly was not transferable in the short run. This bias may have resulted from the calibration sample. Corridors with good transit service were chosen for the sample, increasing the share of persons in the sample who chose that residential location for reasons related to the availability of transit and were "biased toward transit." This underlying "taste for transit" in the calibration sample may have reduced the models' applicability to other "unbiased" populations. This, of course, can be a problem with any model calibrated on cross-sectional data.

3. *Differences in estimation techniques and sample size, etc.*—"Outliers" whose behavior cannot be explained in terms of behavioral relationships calibrated for the rest of the population can nevertheless have a large effect on the estimated coefficients. In effect, the estimation procedure strains to make as much sense as possible of this apparently irrational behavior. This properly raises questions as to whether such observations should be included in the calibration sample and what weight should be given to failure to predict the behavior of outliers in judging whether a model passes a test of transferability.

4. *True behavioral differences*—it may also be that two cities or groups may have different social and economic values influencing their choice behavior. For instance, New York City dwellers may have different values with respect to transportation than rural Midwesterners. These differences in taste may limit transferability from one cultural milieu to another. Tests of variations in tastes by Hausman and Wise (*78*) have demonstrated that they can significantly affect model results.

The fact that the various models cannot be transferred naively without consideration of at least some of these four factors is well known. Failure to account for these known differences between the two circumstances (calibration and forecasting) will lead to poor forecasts except in fortuitous circumstances. Apart from the fact that transferability requires a substantial burden of sophisticated understanding of data collection, model properties, and calibration methods on the part of the practitioner, the relative contribution of true behavioral differences is unknown. These considerations suggest strongly that assurances that the disaggregate models are transferable may be premature.

Atherton and Ben-Akiva (*68*) tested the ability of a work-trip mode choice model calibrated on Washington, D.C., data to explain travel behavior in Los Angeles and New Bedford, Massachusetts. Their model predicts the probability of choosing to drive alone, share a ride, or use transit for a worktrip. The independent variables include mode-specific constants, in-vehicle times, out-of-vehicle times, and out-of-pocket costs for the three modes, income, auto availability, and a dummy variable indicating if the tripmaker is the household head. The test they performed was to use the specification of the original (Washington) model in calibrating new models with Los Angeles and New Bedford data and then to compare the coefficients of the new models with those of the old model. The comparison of the coefficients consisted first of statistical tests of the null hypothesis that the individual coefficients of the Los Angeles and New Bedford models are equal to their Washington model counterparts. For both the Los Angeles and New Bedford models, only the coefficients on the auto availability variables were significantly different from their Washington counterparts. In order to determine whether the differences among coefficients that they found, whether significant or not, would cause much discrepancy among travel forecasts, Atherton and Ben-Akiva also compared the level of service elasticities implied by the Los Angeles and New Bedford models with those implied by the Washington model. As with the coefficients themselves, the elasticities by and large were similar, with large differences only for the elasticities with respect to the out-of-pocket cost variables in the Washington-New Bedford comparison.

From their results, the authors concluded that the evidence on the transferability of logit travel demand models is "encouraging," but that it is apparent that no model will be perfectly transferable and that procedures for "updating" (or adjusting) the model coefficients are required. They then describe and empirically test five update procedures.

The first procedure, the "do nothing" alternative, is simply to apply the model without adjusting any coefficients. This alternative is the only one possible if no disaggregate data set of observed choices is available for the city to which the model is to be applied.

The second procedure may be used if data on the aggregate population shares of the various modes are available along with data on the population averages of the independent variables. This procedure consists of using the population modal shares to adjust the "mode-specific constants" so that the original model accurately predicts these modal shares. None of the other coefficients are adjusted.

The rationale for this approach is that since mode-specific constants capture the mean effects of the unobserved factors and these factors cannot be measured and controlled for, they are the model coefficients most likely to vary from area to area. Differences between the values of the observed attributes for the two areas are presumably accounted for by including them explicitly in the forecasting equation. Note, however, that using this procedure with areawide averages of the independent variables will also be correcting for forecast errors due to aggregation bias in the direct aggregation method. To correct only for transferability, the adjustment should be made so that the transferred model has a good fit for a disaggregate sample for the forecast situation.

The third, fourth, and fifth updating procedures described by Atherton and Ben-Akiva all assume the presence of a small, disaggregate sample. *The third procedure* consists of simply using the disaggregate data set to calibrate a new model (as specified in the old model). Thus, all of the model coefficients would be recalibrated. The drawback of this procedure is that, for small samples, the coefficient estimates may be too unreliable.

The fourth procedure is to use the disaggregate sample to recalibrate the mode-specific constants and to calibrate a scalar that is used to scale all of the other coefficients (so as to keep the ratios between them unchanged). The rationale for recalibrating the mode-specific constants, as in the second procedure, is that it is assumed that the mean effects of unobserved factors are likely to vary from city to city, and these are reflected in the adjustment of the mode-specific constants. The rationale for adjusting the scale of the other coefficients is that travelers in different cities may differ in the level of importance they attach to the variables in the mode choice equation, but it is also likely that the relative weights attached to these variables will be the same for all cities. Maintaining the scale of the coefficients assumes constant tradeoffs between attributes, e.g., a constant "value of time" is preserved. Examination of the ratio of the cost and time coefficients suggests that the ratio demonstrates significantly more stability than does the absolute value of the coefficients.

Even if travelers in different cities do not differ in the importance they attach to the variables, a scale factor may still be appropriate if the variances of unobserved variables differ widely between cities. For example, comfort may be less variable in a city with all new buses than in a second city with a mix of old and new buses. If comfort is an unobserved variable, Tardiff's *(30)* finding suggests that the coefficients of a model estimated in the first city should be higher in absolute value by a scale factor than the coefficients of a model estimated in the second city.

The fifth procedure consists of using Bayesian techniques to combine the coefficients obtained by the first and third procedures. Essentially, this procedure combines the information contained in the original and new disaggregate samples by computing the updated coefficient on any one variable as the weighted average of the coefficient of that variable as calibrated in the original model and the coefficient as calibrated with the small disaggregate data set for the new area. The weights used in combining these coefficients are the inverses of the variances of the coefficient estimates.

Atherton and Ben-Akiva's criteria for goodness of fit were the ability of the procedure to replicate existing modal shares, to do so separately for low and high income groups, and to accurately predict the effects of various policy scenarios, where the "true" policy effects were those predicted by the model calibrated on the full New Bedford sample. (Subsets of this sample were used to simulate the procedures using a small disaggregate data set.)

In empirically evaluating the five procedures that they described, Atherton and Ben-Akiva found that the unadjusted Washington model (the first procedure) fit New Bedford data extremely well, much better than a model calibrated on a small New Bedford sample (the third procedure), and so well that attempts to adjust the mode-specific constants (the second procedure) slightly worsened the fit. Using a small disaggregate sample to adjust the mode-specific constants and scale the other variables (the fourth procedure) performed poorly, but better than an entirely recalibrated model (the third procedure). The Bayesian updating technique (the fifth procedure) provided a very slight improvement over the original model.

Research in England that is similar in spirit to the study done by Atherton and Ben-Akiva has been reported by Daly (*99*). He reported on an investigation of the applicability of a binary logit mode choice model, containing only a mode-specific constant and level-of-service variables, to mode choice problems in several English towns. His research indicated that the models for different towns were sufficiently different as to prevent a model for one town from successfully being used in another. Also, the differences among the towns were such that a "global" model, i.e., a single model calibrated on data pooled across all towns, did not adequately explain mode choice. He did find, however, that a satisfactory global model could be constructed by calibrating common coefficients for the level-of-service variables and then, for each town, adjusting the mode-specific constant to capture factors peculiar to that town (Atherton and Ben-Akiva's second procedure).

Watson and Westin (*100*) studied the transferability of logit mode choice models among different subareas within a single urban area. Their data were for the Edinburgh-Glasgow area of Scotland, and their model, like Daly's, was a binary model containing level-of-service variables and a mode-specific constant but no socioeconomic variables. They grouped their data into six categories of the location of trip origins and destinations. The six categories were: (1) both trip ends in the central city; (2) both trip ends in the suburbs; (3) both trip ends in the area peripheral to the urban area; (4) one trip end in the central city and one in the suburbs; (5) one trip end in the central city and one in the peripheral area; and (6) one trip end in the suburbs and one in the peripheral area. Identically specified models were calibraed separately on each of the six subsamples. Each of the six models was then used to predict the mode splits of the other five subsamples. Referring to the three categories that contain at least one trip end in the central city as the central group, the authors found that the models within the central group predicted well among themselves. The noncentral group (the remaining three categories), on the other hand, performed very poorly in their predictions for each other. Prediction between the central and noncentral groups (in either direction) showed mixed results but on balance were within reasonable bounds.

In order to pinpoint the reasons for the poor predictions among the noncentral group, Watson and Westin carried out tests for significant differences of all coefficients taken together in pairwise comparisons of the six models. In line with their results on predictive ability, they found that the coefficients on the model for any noncentral group were significantly different from the coefficients in the model for any other noncentral group, but no significant differences were found when the central group categories were compared with one another. In the comparisons with one central and one noncentral category, half of the differences were significant and half were not. Watson and Westin concluded from their research that the ability of the model of the central group to cross-predict accurately was favorable to the within-urban area transferability of mode choice models but that results for the noncentral groups indicated the need to refine the models to take account of locational differences.

One early application of disaggregate logit models for mode split, and for travel to the CBD and to non-CBD destinations, also produced significant differences in estimated coefficient values (*101*). Antti Talvitie and Daniel Kirshner (*102*) have recently completed another test of transferability which is substantially less optimistic than previous researchers. Their research, based on the use of four data sets, indicated that:

1. Outliers can have substantial impacts on the point estimates of some of the coefficients in logit models.
2. Model coefficients are highly sensitive to model specification.
3. Model coefficients do not appear transferable within region, between regions, or over time.

These results agree with some results of a less rigorous comparison of the estimated coefficients of time and cost and corresponding elasticities in the various disaggregate models of mode choice. These coefficients can differ significantly. To the extent these differences can be accounted for by known differences between the calibration and forecast data sets, they can be accounted for in the forecast. To the extent they are represented by biases in estimation or true behavioral differences, rather than variable definition, transferability will be doubtful.

An example of an extremely important difference between situations that can generally be accounted for is changes in the general price level and income (inflation). Unless the logit model specifically indexes the cost variable (by say dividing cost by income or wage rate), adjustments to the variable or the cost coefficient will be required. The requirement for this adjustment may be seen by referring to the log-odds formulation variable "auto cost minus bus cost." Clearly, if both of these variables increase by the same percentage due to inflation, their difference will increase, affecting predicted mode split. To adjust for the change, the coefficient or the variable must be deflated to ensure that choice is sensitive only to cost differences in real terms.

Recommended Approach

To summarize the available evidence on the transferability of disaggregate mode choice models, the works reviewed here seem to be in close agreement that not all travelers everywhere exhibit the uniformity in their trip-making behavior that one would have hoped for, at least with respect to worktrip choice of mode. However, from the work of Atherton and Ben-Akiva and Daly, it appears that the differences that do exist are sometimes amenable to reconciliation by an adjustment of model coefficients and that calibration of separate models for every traveler group is not always necessary, at least for mode split. Other research indicates that logit model estimation may not display robustness with regard to differences in traveler tastes, data collection, or model specification, and can be very unforgiving of errors on the part of practitioners. These difficulties suggest that for many applications the collection of a new data set and the calibration of new models may be required and is a far safer course than attempting to transfer a model, especially for those without advanced training in the use of disaggregate models. Several hundred new observations should be sufficient to test the transferability of the models.

In selecting one of the updating procedures to be recommended in the guidelines given in this report, the considerations are the ability of the procedure to perform adequately without an unduly heavy computational requirement. On this basis, Atherton and Ben-Akiva's fourth procedure, discussed earlier in this section, appears most suitable, despite their reservations, if the data are available. Their procedure uses a small disaggregate data set to adjust the mode-specific constants and scale the other coefficients of the original model so that the fit of the original model to the new data is improved. This procedure may be modified by substituting data on the aggregate choices of market segments for the small disaggregate data set. The principal arguments for the procedure are that it provides a better fit to the data on market segments while preserving the ratio of coefficients (which preserves such relationships such as the value of time).

Limitations of data may make the other approaches attractive. For example, if no data are available on the choices of the population prior to the policy change, the "do nothing" alternative must be considered. If data are available to update the model, the "do nothing" alternative should be rejected on the strength of the evidence in favor of the hypothesis that different cities have different representative utilities, Atherton and Ben-Akiva's results for Washington and New Bedford notwithstanding. Their second procedure, adjusting only the mode-specific constants, is attractive when data are available only for the aggregate choices and scaling of the other coefficients is unfeasible. Atherton and Ben-Akiva's third procedure, recalibrating the model, was rejected for the same reason that they found it unsuitable: the large standard errors of the coefficients in the recalibrated model make it unreliable. Finally, their fifth procedure, the Bayesian updating technique, was rejected because of the computational burden imposed by the requirement of variance-convariance matrices for the coefficients in the original and recalibrated models.

CHAPTER THREE

INTERPRETATION, APPRAISAL, AND APPLICATION

INTRODUCTION

This chapter focuses on the application of disaggregate approaches to travel demand analysis. In earlier chapters, the strengths and weaknesses of disaggregate models were discussed. The major conclusion emerging from reviewing the state of the art in disaggregate modeling is that disaggregate models are a valuable research tool for transportation planners. But they can be subject to significant errors (as can aggregate techniques) and require a relatively sophisticated understanding of the assumptions employed. The researcher must make a relatively heavy commitment to understanding what he is doing if these errors are to be avoided. Forewarned by this caveat, the user can be assured that disaggregate models will play a useful role in the transportation planner's repertoire of planning tools.

Disaggregate travel demand models can be (and have been) applied in several different ways. They can replace one or more individual components of the conventional transportation planning system. Alternatively, they can be used for problems that are not easily addressed by conventional planning tools, e.g., analysis of Transportation Systems Management (TSM) actions or the introduction of new transportation modes. They can either be computerized or used as sketch planning tools requiring only hand-held or programmable calculators.

EVALUATION OF SKEPTICISM REGARDING THE VALUE OF THE DISAGGREGATE APPROACH

As outlined in this report, disaggregate demand modeling techniques have made considerable progress in recent years. Yet, during the course of the present project there has been increasing skepticism regarding the value of disaggregate models. This skepticism is based on a number of concerns. The first concern regards the issue of accuracy. Talvitie and Kirshner (*102*) found that "outliers" caused by data entry errors or highly unusual behavior can have significant effects on the estimates of the coefficients. The authors tested the transferability of model coefficients within regions, between regions, and over time and rejected it for each case. The authors were also troubled by the sensitivity of the estimated coefficients to model specification. Finally, the authors were troubled by the fact that 60 to 80 percent of the explanatory power of the models is contained in the alternative-specific constants and only 20 to 40 percent in the LOS and socioeconomic variables. The authors glumly inquire whether disaggregate models have much to offer given the fact that service variables are only slightly affected by most policy transportation changes.

Typical of the expression of skepticism regarding the generality of the usefulness of disaggregate models is the attempt by Gomez-Ibanez et al. (*103*) to transfer disaggregate elasticities to evaluate auto restraint policies in the Boston area. The authors were dismayed to find that different studies produced such greatly differing elasticities, causing doubt about the transferability of models and concern about the sensitivity of the results to model specification. The authors review the various explanations for the differences in elasticities and caution users in applying transferred models because of the apparently large differences in estimated behavioral relationships.

Different researchers using different data have reported significantly different coefficients for time and cost variables. This lack of consistency has greatly troubled some researchers, such as Gomez-Ibanez et al., who consider the lack of uniformity a great shortcoming of the disaggregate approach. On the other hand, others, such as the Office of Technology Assessment, U.S. Congress (*104*), have found the relative consistency of value of time (after accounting for inflation) to be reassuring despite the differences in coefficient values, and have not hesitated to evaluate nationwide energy policy using the results of disaggregate models.

Skepticism has been especially keen regarding the value of large-scale attempts to substitute disaggregate models for the traditional four-step transportation planning process. Shunk and Kollo (*105*), representatives of the Metropolitan Transportation Commission (MTC) of San Francisco, have criticized the disaggregate model on the grounds that such sophisticated models are not appropriate for day-to-day use in the real world, despite their elegance or relative accuracy. The grounds for their complaints were the following:

1. Despite the claims that the models represent traveler decision-making, dramatic changes in the estimated constants are required to "validate" the models from the estimation subsample to the aggregate data set.
2. A "distance correction variable" for each of 30 districts was required for trip distribution, reminiscent of "friction factors" in aggregate models.
3. The distribution models required "unique adjustment factors" which had to be adjusted by hand to produce reasonable forecasts, rather than "responding independently."
4. The mode choice models required adjustment of the mode-specific constant which was specific to the interchange.
5. The model is unduly complex and costly to operate.

The range of error inherent in forecasting with disaggregate models must be considered a major disappointment. However, much of this error is not a result of the use of disaggregate models per se, but rather results from the error inherent in any forecasting process. It is useful in this regard to consider both errors in estimating the model (e.g., sampling error in estimating model coefficients) and errors in using the model to forecast the dependent variable (e.g., mode split). As discussed earlier, Horowitz (*31, 73, 74, 77, 89*) has reviewed the sources of errors in the logit model, diagnostic tests for detecting those errors, and the consequences of undetected errors. One may conclude that while the consequences of such errors are significant, good judgment and the proper use of proposed tests will reasonably safeguard against most serious errors.

APPLICATION OF DISAGGREGATE DEMAND MODELS

It is beyond the scope of this report to provide a comprehensive report on experiences in the use of disaggregate models. In addition to the voluminous literature, some of which is cited in the report, Spear (*56*) has conducted a survey of applications for dissemination of experience in use of individual choice models. Responding to a request of researchers to bridge the gap between research and practice, this report was written to provide transportation planners with a working reference on recent experience in applying disaggregate models and case studies. It is a basic reference document that should be on the shelf of every practicing transportation planner with an interest in disaggregate demand modeling. Hensher and Stopher (*106*) also present a comprehensive review of previous applications. The U.S. Department of Transportation has also included a logit calibration package as part of its battery of modeling programs.

Use in the Conventional Modeling System

Because mode choice has received considerable attention in previous research, it is not surprising that the integration of disaggregate mode choice models with conventional planning tools has been among the earliest applications. Spear (*56*) presents a comprehensive review of two efforts: the development of a multinomial logit work mode choice model in San Diego and the development of multinomial logit mode choice models for work and nonwork trips in the Twin Cities area.

Two major projects were designed to develop a battery of disaggregate models in actual transportation planning agencies that are comparable to the aggregate four-step process. One recent comprehensive effort is the "Sigmo Study" (*48*), which represented the state of the art in disaggregate demand modeling. The group of models specifies trip generation, modal split, and distribution in separate steps. These steps are linked by use of the "log sum of the denominator" from the prior step, e.g., as an accesssibility term (treated negatively as "resistances") in the distribution model (see Chapter 2). Separate models are derived for home-based work (HBW) and home-based other (HBO). The models were calculated using travel diary data (over 7 days) from all members of 3,000 households, peak and off-peak networks, and land-use data. The multinomial logit model was the primary functional form.

The overall results showed that model calibration produced estimated coefficients that generally satisfied a priori expectations. Validation of the model's forecasting properties was accomplished by grouping data by variables not used as forecasting variables (trip distance, destination sectors, etc.). The authors' conclusions were that for some forecasts the correspondence between the predicted and observed was "reasonable." For others, especially mode split for short trips, the results were "poor."

The second comprehensive model is illustrated by Figure 2, showing the Metropolitan Transportation Commission (San Francisco Bay area) model. The models are described in Ruiter and Ben-Akiva (*107*); in Ben-Akiva, Sherman, and Kullman (*49*); and in numerous working papers and manuals. The short-range generalized transportation policy analysis

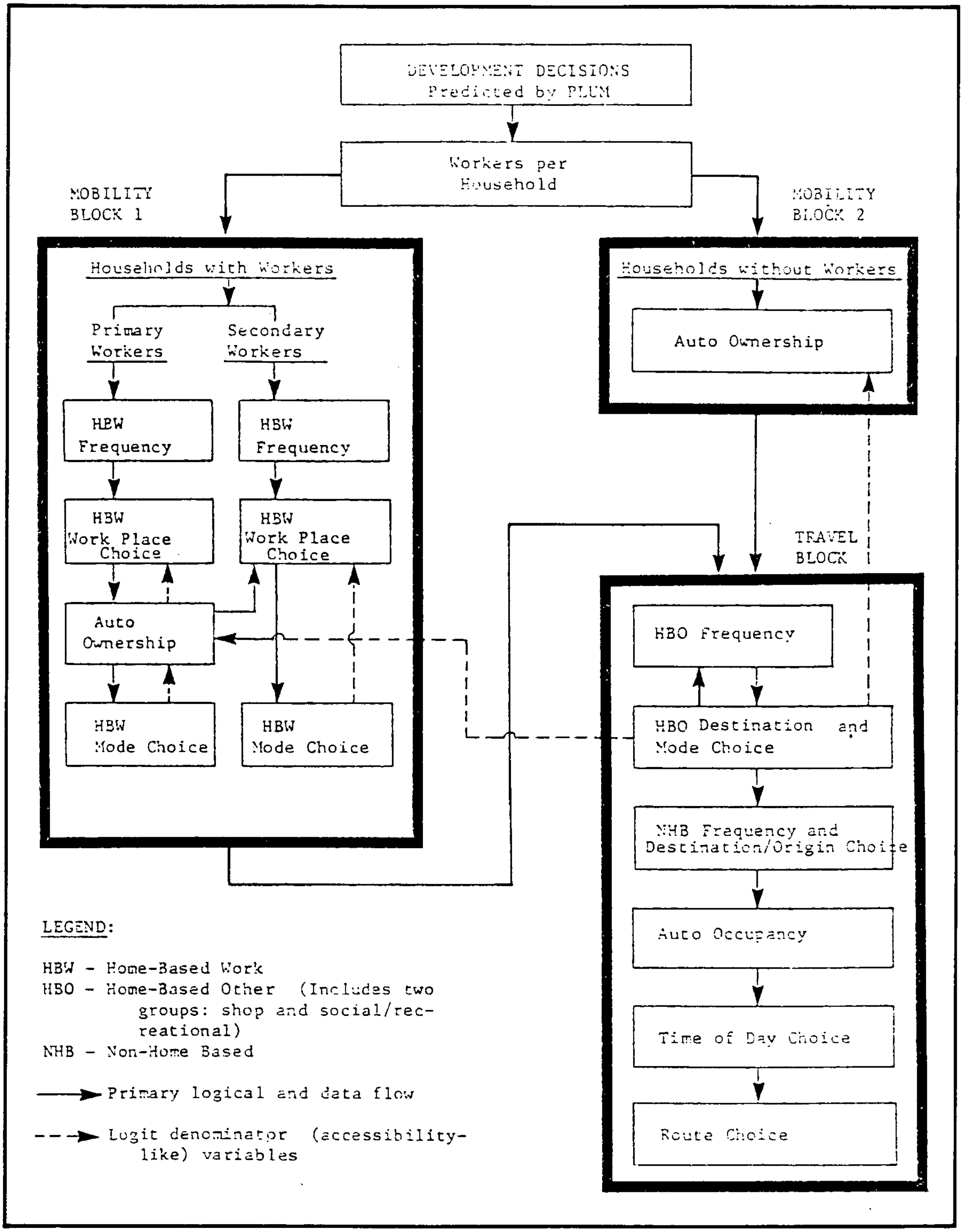


Figure 2. The MTC travel demand model system.

(SGRP) version of the modeling system has been adapted for energy conservation policy analyses (*108*).

Nonconventional Problem Areas

The greatest potential for application of disaggregate approaches may be in areas that are not currently being met by conventional planning methods—"short turn around" response to evaluation of "low-capital" alternatives. In an interim report as part of NCHRP Project 8-13, Charles River Associates (*2*) provided a lengthy example of such an application showing how to use the market segmentation technique as a means of aggregating the results of a disaggregate model forecasting process. The particular example chosen was the demand for "Park-and-Ride" in Baltimore. The results of the application indicated that: (1) improvements in time and cost were not responsible for all of the observed demand for park-and-ride, indicating that behaviorally relevant variables not included in the model can have a significant impact on the forecast; and (2) lack of knowledge about the appropriate mode-specific variables (especially when the new mode is a hybrid of two existing modes) is a difficulty that must be addressed in applying the model to demand for a new mode. In Appendix B, further analysis of the park-and-ride example is described. The results appear to be more encouraging.

One interesting example of the use of disaggregate models to analyze the relationship between auto ownership and travel demand was a study of carpooling incentives using the disaggregate modeling approach. Atherton, Suhrbier, and Jessiman (*97*) found that incentives for carpooling result in modest reductions in auto travel to work. However, the availability of the auto at home may stimulate additional nonwork travel, a behavioral response specifically accounted for by the model.

Readers interested in other examples of applications should consult Charles River Associates, *Policy Evaluation with Travel Behavior Models: Methodological Issues and Case Studies (109)*. This report contains analyses of the cost effectiveness of areawide integrated transit, transportation control plans, and detailed procedures for policy evaluation with small surveys and Census data.

Manual Methods

Manual methods for applying disaggregate techniques are generally incremental or pivot point approaches. In these approaches changes in modal shares are forecast based on changes in level of service or other explanatory variables. The existing modal shares are used as a base rather than forecast *de novo*. This may reduce the complexity of application and margin for error.

Several interesting projects have identified sketch planning tools using disaggregate models and hand-held calculators or programmable calculators. The first is incorporated in the "Work Sheets" in Cambridge Systematics (*110*). This approach relies on a formula for estimating a new logit share after a change in the transportation system, given only the change in the explanatory variables and the preexisting logit share. Letting Pi^{B} = initial share of i^{th} alternative, Pi^{N} = new share of the i^{th} alternative, ΔVi = charge in utility of i^{th} alternative, and using standard logit notation:

$$P^{B}_{i} = \frac{e^{V_i}}{\sum_j e^{V_j}} \tag{55}$$

$$P^{N}_{i} = \frac{e^{V_i + \Delta V_i}}{\sum_j e^{V_j + \Delta V_j}} = \frac{e^{V_i} e^{\Delta V_i}}{\sum_j e^{V_j} e^{\Delta V_j}}$$

$$= \frac{\left(e^{V_i} / \sum_k e^{V_k}\right) e^{\Delta V_i}}{\sum_j \left(\frac{e^{V_j}}{\sum_k e^{V_k}}\right) e^{\Delta V_j}} = \frac{P^{B}_{i} e^{\Delta V_i}}{\sum_j P^{B}_{j} e^{\Delta V_j}} \tag{56}$$

This formula provides the basis for a very straightforward method of revising modal shares: estimate the effect of the change of level of service on utility and insert the change in utility and original modal shares into the equation. An early application of this formula is found in a study of energy and mass transit by the Office of Technology Assessment of the U.S. Congress (*104*). A more recent application is Kumar's (*111*) analysis of changes in modal splits resulting from a proposed extension of a light rail line in Cleveland.

Incremental logit analysis estimates changes in demand resulting from changes in the transportation system. In this sense, it is conceptually similar to the use of elasticities in demand forecasting and other pivot point methodologies.

In addition to the differences in computational procedures, there is an important difference in the underlying assumptions made in methods that predict absolute levels of future demand versus methods that predict changes from a base demand. In the former case, it is implicitly assumed that the unobserved variables vary randomly over time. In the latter case, it is assumed that the unobserved variables have roughly the same values in the future time period as in the base period. Therefore, theoretical, as well as practical, issues are involved in choosing between the alternative approaches.

A study undertaken at the Massachusetts Institute of Technology by Manheim et al. (*112*) has attempted to apply the programmable calculator to sketch planning with disaggregate models and thereby extend the early Cambridge Systematics work. The object of this research is to combine the benefits of disaggregate modeling with the evolution of new, powerful, pocket programmable calculators. The MIT project is writing programs for such calculators, using the "incremental logit formula" (Eq. 56) and aggregation procedures discussed elsewhere in this report such as market segmentation. One particularly useful program uses the approach of generating a "synthetic sample" of households from published Census data. Examples are provided to show how the programs and models should be used.

REQUIREMENTS FOR DISAGGREGATE MODEL APPLICATION

The discussion in this and in the preceding chapter clearly indicates that the disaggregate modeling approach can involve different types of analytical approaches, which can be applied to different types of problems. These different uses of disaggregate models require somewhat different levels of resources and skills from the practitioner. In this section, the requirements for the various types of uses are discussed. To facilitate the presentation, a distinction is made among (1) model calibration; (2) model updating; and (3) model application. In discussing model application, a further distinction is made among (1) travel-based forecasts; (2) regional or subregional forecasts; and (3) manual methods.

Model Calibration

Much of the discussion in the earlier chapters was based on findings from previous model calibration studies. The practitioner who is considering specification of a new disaggregate model for his/her area would benefit from understanding the conceptual, theoretical, and operational issues discussed in these chapters.

Model calibration involves the following steps: (1) selection of the travel decision to be modeled, e.g., modal choice; (2) specification of the independent variables; (3) selection of the size of the calibration sample; (4) location of an existing data source and/or collection of a new data source that contains the relevant variables; (5) statistical estimation of the model(s); and (6) interpretation of model coefficients and other summary statistics.

The selection of the travel decision is clearly related to policy problems being addressed. For example, an analysis of high vehicle occupancy policies would require a modal choice model which contains high occupancy modes such as shared ride. In general, since modal choice models are the most advanced disaggregate travel demand models, they are likely to be applied most frequently in the near future.

The selection of independent variables should be guided by the nature of the policy under consideration, theoretical considerations, findings from previous studies, and data availability. Although there are some definite conclusions emerging on the specification of independent variables, there are differences in existing models. For example, the modal choice models estimated by CRA with the Pittsburgh data contain a relatively small number of independent variables, while the recommended mode choice model from the Urban Travel Demand Forecasting Project (*6*) contains a large number of explanatory variables. Thus there are no absolute guidelines for selecting explanatory variables. At a minimum, the analysts should include variables that are sensitive to the policies being considered. For example, preferential parking policies would affect out-of-vehicle or walking times. Familiarity with the previous disaggregate travel demand studies would be very useful in aiding the selection of independent variables.

As indicated in Chapter Two, disaggregate travel demand models can be estimated with small samples. For example, in this project, samples of slightly more than 100 individuals were used in the estimation of the models based on the Pittsburgh data set. In the Urban Travel Demand Forecasting Project, various versions of the pre-BART models were estimated with 161 and 771 individuals and the post-BART models were estimated with 635 individuals.

The estimation of disaggregate travel demand models requires a data set that contains the relevant dependent and independent variables. Although it is desirable to have data that were specifically collected for the purpose of disaggregate model estimation, the use of an existing transportation data set may be more cost-effective. For this reason, many existing models have been estimated with secondary data sources.

In general, information on the transportation choice (dependent variable) and socioeconomic characteristics is available from survey data. Characteristics of transportation alternatives, such as modal time and costs, are often obtained from secondary sources such as networks, land-use data, etc.

Collection of a new data set is a major undertaking. However, the data collection costs may be justified in light of the fact that most existing disaggregate models were estimated with data from the late 1960s or early 1970s. In addition to sample size and measurement of the variables in the model, which have already been discussed, the method of data collection is an important issue. Home interviews, which are a standard method of transportation data collection, are a reliable, but costly, approach.

Telephone interviews were collected in addition to home interviews for the Urban Travel Demand Forecasting Project (*6*). This approach appeared to yield very favorable results. It was concluded that response rates and data accuracy were reasonably comparable to that of home interview surveys. The telephone interviews required about one-third the costs of the home interviews.

Choice-based sampling theory implies that low-cost self-administered survey techniques can yield reasonable results. For example, on-board surveys can be used to collect information on transit users. Automobile users (single occupant and shared ride) can be contacted at particular check points and presented with mailback questionnaires. For example, Kavak and Demetsky (*113*) contacted auto users by means of license plate identification and CRA (*114*) distributed questionnaires at toll plazas.

Statistical estimation of disaggregate choice models requires a computerized estimation package. The QUAIL package was developed by the Urban Travel Demand Forecasting Project and the U.S. Department of Transportation UTPS package has the ULOGIT program. The National Bureau of Economic Research's TROLL package has the LOGIT routines. Estimation of more advanced models, such as multinomial probit, requires specialized programs that are not readily available and have not been thoroughly tested.

The basic skills required of the analyst include (1) data management skills so that the variables of the model can be entered correctly, and (2) understanding of the operation of the computer program in question. In general, although disaggregate choice model estimation packages are specialized and more complicated than standard statistical packages, working knowledge of standard packages would facilitate the use of disaggregate choice modeling packages.

Interpretation of model coefficients and summary statistics is facilitated by a thorough understanding of the issues discussed in the preceding chapters. In general, coefficients are checked for proper sign. For example, cost coefficients should be negative, since the utility of a travel alternative should decrease as cost increases. Similarly, the magnitudes of the coefficients should be reasonably similar to magnitudes found in previous studies, unless there are special circumstances that indicate otherwise. For example, the values of time derived from the model can be compared to values derived from previous studies and elasticity values can be checked.

Model Updating

Several transferring and updating procedures were identified in Chapter Two. Probably the most practical procedure involves the adjustment of cost (and linear) terms to reflect differences in the cost of living between areas and/or over time and adjustment in alternative specific constant terms so that predicted market shares match observed market shares.

The adjustment to the cost coefficient depends on the functional form of the cost variable in the model. This point can be illustrated with specific examples. Before presenting the examples, the variable INDEX is introduced, which converts cost variables for the application of the model into dollar amounts equivalent to those of the calibration data set.

The basic principle is that the coefficient is adjusted such that the representative utility does not change. That is, if C_1 is cost measured in current dollars (application data), C_0 is measured in dollars for the calibration data set ($C_1 = C_0$ $INDEX$), α_0 is the original coefficient, and α_1 the updated coefficient; it is desired that

$$\alpha_0 f(C_0) = \alpha_1 f(C_1) \tag{57}$$

where f is the functional form for the cost variable. An immediate result of this condition is that the model can be "updated" by deflating C_1 for each individual by $INDEX$ and setting $\alpha_1 = \alpha_0$. This procedure applies to all functional forms.

Case 1—Linear Cost Terms. In this case, the condition becomes

$$\alpha_1 (C_1) = \alpha_0 (C_0) \tag{58}$$

By expanding C_1 in terms of C_0

$$\alpha_1 (C_0 INDEX) = \alpha_0 (C_0) \tag{59}$$

which yields

$$\alpha_1 = \frac{\alpha_0}{INDEX} \tag{60}$$

That is, the coefficient is deflated by $INDEX$. This procedure was followed in Phase II of this study in which the cost coefficient of the model estimated with a Pittsburgh data set was deflated for application to a higher income Baltimore area.

Case 2—Cost Divided by Income. As mentioned in Chapter Two, cost is often divided by income in existing models. In this case, no adjustment in α_0 may be necessary. This follows from the fact that income is likely to change in the same way as cost and, therefore, the effect of $INDEX$ cancels out in the ratio of cost to income. The analyst should verify this for each case. If the cost/income ratio is not relatively constant, some adjustments may be necessary.

Case 3—Logarithmic Cost Term. The condition to be satisfied is

$$\alpha_1 \log C_1 = \alpha_0 \log C_0 \tag{61}$$

Using $C_1 = C_0 \cdot INDEX$,

$$\alpha_1 \log (C_0 \cdot INDEX) = \alpha_0 \log C_0 \tag{62}$$

This expression can be transformed to

$$\alpha_1 \log C_0 + \alpha_1 \log INDEX = \alpha_0 \log C_0 \tag{63}$$

That is, one sets $\alpha = \alpha_0$ and adds α_0 log $INDEX$ to the alternative specific constant term for each alternative. However, since the addition of the same constant to all alternative specific constants has no effect on the resulting choice probabilities, the α_0 log $INDEX$ term can be ignored. Hence, no adjustment is necessary. (It is possible to specify cost terms as alternative specific variables, i.e., the cost coefficients would vary by alternative. In this case, alternative specific constants would have to be adjusted. In practice, there are apparently no existing models with alternative specific cost variables, so no adjustment to alternative specific constants is necessary.)

Alternative specific constant terms are updated by adjusting the terms so that aggregate shares predicted by the model equals observed aggregate shares. The general procedure is represented by the following expression (for multinomial logit)

$$MS_i = \frac{1}{T} \sum_j N_j \frac{e^{\alpha_i + V_{ij}}}{e^{V_{Aj}} + \sum_{k \neq A} e^{\alpha_k + V_{kj}}} \tag{64}$$

in which

MS_i = the share for the i^{th} alternative;
T = the sample size;
N_j = the number of individuals in the j^{th} market segment;
V_{ij} = the average representative utility for alternative i in the j^{th} segment;
α_i = the alternative specific constant for the i^{th} alternative;
V_{Aj} = the representative utility for the "base" alternative (the one without a constant);
V_{kj} = the representative utility for the k^{th} alternative; and
α_k = the alternative specific constant for the k^{th} alternative.

If there are J alternatives, the expression yields J-1 equations with J-1 unknowns (the α_i).

Three special cases are of interest. First, if there is only one market segment (the whole sample), then the procedure involves use of the sample average representative utilities. Second, if $N_j = 1$ for all j, each individual is used directly in adjusting the constants. This procedure was used by CRA in updating the Pittsburgh work mode choice model for use in the Boston area (*109*). Third, the intermediate case in which $1 < N_j < T$ was used is the case study of Phase II of this project.

In general, solution of Eq. 64 requires an iterative procedure that is very similar to that used in model calibration. However, for the first special case in which there is only one market segment, there is a direct solution. It can be shown that

$$\alpha_i = \log \left(\frac{MS_i}{MS_A}\right) - (V_i - V_A) \tag{65}$$

Model Application

As noted earlier, a distinction is made among zonal based applications, regional or subregional applications, and manual methods.

Zonal Based Applications

This type of application is conceptually similar to standard aggregate forecasting. Zonal based forecasts are necessary when the travel flows between zones are of interest. For example, the traditional emphasis on providing new facilities requires fairly specific zone-to-zone travel forecasts. The following steps are necessary.

1. Calibrate or update an appropriate disaggregate choice model.
2. For each zone pair, develop average values for variables corresponding to the independent variables of the model. Those variables likely include zone-to-zone average travel times and costs, variables such as average income and auto ownership for the origin zone, and attraction variables such as retail employment for the destination zones.
3. An appropriate aggregation procedure must be used. Because of the errors inherent in simply using zonal averages in forecasting aggregate shares, the aggregation procedure involves additional information on the distributions of the variables such as the variances and covariances.

Because forecasts are made for each zone pair, and aggregation procedures are analytically complicated, computer assistance is usually necessary. To examine the effects of various policies, the zonal averages for the variables are changed to correspond to the policy change. For example, an increase in gasoline taxes would result in increases in the costs of automobile modes for each zone pair. An example of policy analyses using zonal based forecasts is given in CRA (*42*).

Regional and Subregional Applications

For some applications, specific zone-to-zone forecasts are not necessary. For example, the regionwide reduction in VMT and fuel consumption resulting from a reduction in transit fare may be the focus of a particular analysis. Two procedures have been developed for these types of problems. The market segmentation aggregation method presented in this report is one method. The other approach is the use of a disaggregte sample in forecasting (the random sample enumeration approach). This approach has also been called short-range generalized transportation policy analysis (SRGP).

The essentials of the random sample enumeration approach are as follows:

1. Calibrate or update a disaggregate choice model.
2. Delineate the region or subregion of interest.
3. Acquire a random sample of observations for the area. The observations must contain values for each of the independent variables in the model. Socioeconomic variables are likely to be available directly from the survey. Level-of-service variables may be derived from the survey data or from supplementary network or field measurements.

Alternately, a synthetic random sample can be generated using the SYNSAM or interative proportional fitting methods developed for the Urban Travel Demand Forecasting Project (*6*) and described in Appendix A.

4. For each observation of the sample, insert the values of the independent variables into the choice model and derive the choice probabilities.
5. Weight the choice probabilities by using the sampling fractions used to generate the sample.
6. Add the weighted probabilities to derive the market shares for the entire area.

The sample size necessary to implement this procedure is of the same order of magnitude as the size necessary for model estimation. However, if travel forecasts for subsamples are desired (e.g., income groups), a larger sample size may be necessary to ensure that subsample forecasts are reliable. The effects of alternative policies are forecast by changing the level-of-service variable for each individual observation in the sample. For example, an exclusive lane for high occupancy vehicles would change in-vehicle times for each individual.

The random sample enumeration procedure is described in detail in Appendix B of CRA (*2*) and in Cambridge Systematics' five-volume report on the use of disaggregate models in energy policy analyses (*108*).

Manual Methods

Manual methods are appropriate for problems in which choice probabilities are forecast for only a limited number of cases. For example, forecasting the changes in model shares resulting from improved transit level of service between a homogeneous residential area and the CBD may require only a single application of a choice model. Manual or pocket calculator methods for these problems have been developed by CSI (*108*) and Manheim et al. (*112*). The procedure can be illustrated by two examples.

Case 1—Forecasting the Effects of Auto Cost Changes Using a Binary Modal Choice Model. Suppose the data in Table 1 (and Eqs. 7 to 11) represent the travel environment from a homogeneous group of commuters. Because of increases in tolls and parking charges, the auto cost is increased to $1.25. The new modal shares are calculated as follows.

$$V_{1t} = 1[-5.72 + 1.38(1) + 4.07(1)] - 0.117(0) -0.0348(60) - 9.06(0.25) = -4.623 \quad (66)$$

$$V_{2t} = -0.117(7) - 0.0348(110) - 9.06(0.10) = -5.553 \quad (67)$$

$$P_t(1) = \frac{e^{-4.623}}{e^{-4.623} + e^{-5.553}} = 0.717 \quad (68)$$

$$P_t(2) = 1 - P_t(1) = 0.283 \quad (69)$$

Case 2—Incremental Logit Analysis of the Effects of an Exclusive Bus Lane. To perform incremental logit analysis, the modal shares before the policy change are necessary. In addition, the changes in the independent variables are needed.

Consider the same base case as in Case 1. Suppose the original modal shares are 0.639 and 0.361 for auto and transit, respectively, and that the exclusive lane reduces bus in-vehicle time by 10 min (to 100 min) and increases auto in-vehicle time by 5 min (to 65 min). The new modal shares are:

$$P^A{}_t(1) = \frac{P^B{}_t(1)\, e^{-0.0348\, \Delta INVTT_1}}{P^B{}_t(1)\, e^{-0.0348\, \Delta INVTT_1} + P^B{}_t(2)\, e^{-0.0348\, \Delta INVTT_2}} \quad (70)$$

$$P^N{}_t(2) = 1 - P^N{}_t(1) \tag{71}$$

in which

$P^A{}_t$ = the "after" modal share; and
$P^B{}_t$ = the "before" modal share.

Inserting the actual values for $\Delta INVTT_i$ and $P^B{}_t(i)$ into Eq. 70 yields

$$P^A{}_t(1) = \frac{0.799\,e^{-0.0348\,(5)}}{0.799\,e^{-0.0348\,(5)} + 0.201\,e^{-0.0348\,(-10)}} \tag{72}$$
$$= 0.702$$

$$P^A{}_t(2) = 1 - P^N{}_t(1) \tag{73}$$
$$= 0.298$$

It should be noted that although the model contains several independent variables only *INVTT* was used in the forecasting because none of the other variables changed.

FLEXIBILITY OF THE DISAGGREGATE APPROACH

Travel demand models are expected to conform to two conflicting objectives. First, they are expected to be flexible to respond to the data availability, time frame, degree of accuracy required, and skill level of the user. On the other hand they are expected to be routine, low-cost, and not demanding on scarce high-talent resources. Clearly, the models must meet a test of balance between these objectives. Neither "canned models" nor tailor-made approaches to every planning problem are the answer.

If disaggregate demand modeling is to be flexible enough to meet the wide range of demands, there can be no one disaggregate model. The appropriate tool for analysis of travel demand must be selected from a range of analytic alternatives depending on the circumstances. The transportation analyst must approach the issue at hand with a flexible framework that is based on the behavioral motivations of the individual as his starting point. The framework must take into account the individual's motivations, the nature of the individuals in the relevant population, the information from the analysis that is needed for decision-making, and the evidence from previous research. Naturally the forecast will require that the analysis of the individual be extended to forecast group behavior. This conceptual framework is the essence of disaggregate behavioral demand modeling.

Certain demand analysis will be required routinely, and transportation analysis can be standardized by identifying a common methodology and data relevant to the case in point. But even in these routine cases, the transportation analyst cannot abandon analysis and blindly follow formulas. Rather he must be constantly combining his standardized approaches with his common sense understanding and experience concerning the reasonableness of his analysis.

For these reasons, this study has emphasized empirical and conceptual model development. The approach has been to analyze the fundamental impediments to understanding and modeling travel behavior at the disaggregate level that are common to all applications: data requirements, desired variables, specification of the model (e.g., how to consider mode choice, destination choice, and trip frequency in the same model), the desirability of certain properties of the logit model (e.g., the Independence of Irrelevant Alternatives). The results have considerably advanced the state of knowledge required to implement the disaggregate conceptual framework to a given practical planning issue. However, the results should not be considered "a model" in the sense of a set of formulas to be routinely applied to whatever issue may arise.

OVERVIEW OF TECHNICAL APPENDIXES

This report contains six appendixes which document in greater detail the findings described in Chapter Two and also provide guidelines for implementing particular research procedures. This material is directed to the researcher and the advanced practitioner as background for future research and sophisticated application.

Appendix A summarizes research on the approaches that have been proposed for dealing with the aggregation problem. Based upon this review, the market segmentation approach is recommended, as described in Chapter Two.

Appendix B is a detailed analysis of two major issues described in Chapter Two: transferability and aggregation. The potential forecasting biases that might arise from not handling these problems appropriately are illustrated by means of several examples.

Appendix C provides detailed guidelines for using the recommended market segmentation aggregation procedure with Census data. This appendix also describes new and more favorable findings on the application of a disaggregate work mode choice model to forecast travel demand following the implementation of a new park-and-ride alternative in Baltimore. The original analysis of this case study was presented in the Phase II report (*2*).

Appendix D describes the development of the disaggregate travel demand models presented in Chapter Two. The presentation describes preparation of the three data sets used in model development (Pittsburgh, Twin Cities, and Baltimore); model specifications; and hypothesis tests involving alternative model specifications. The experience with the Baltimore data set may be of special interest because this data set was designed for development of advanced travel demand models.

Appendix E provides an in-depth discussion of the IIA property. The implications of IIA and the consequences of violating the IIA assumptions are discussed and illustrated with several examples.

Appendix F describes a modification to the multinomial logit model that allows the model coefficient to be random rather than fixed. This modification addresses one of the major potential violations of IIA.

CHAPTER FOUR

CONCLUSIONS AND SUGGESTED RESEARCH

CONCLUSIONS

The research in NCHRP Project 8-13 indicates that disaggregate models of travel demand deserve wide application to transportation planning problems. They are especially relevant to the analysis of policy issues not satisfactorily considered with conventional approaches, such as determining the effect of alternative air quality control and energy conservation policies, evaluating the impact of traffic engineering improvements and toll policies onthe use of roads, and other "low capital" policy issues.

Disaggregate demand approaches offer considerable advantages over conventional approaches in many applications because of substantially reduced data costs; the ability to predict the effects of public policy on travel demand; flexibility to meet different problems, accuracy requirements, and response times; and improved transferability of model estimation results from one geographic area to another.

This study has developed several new disaggregate demand models for worktrip mode choice and shopping choice of mode. The estimated models have been evaluated and found to be highly useful in analyzing policy issues.

An improved understanding of the Independence of Irrelevant Alternatives (IIA) property is an important contribution of the study. The IIA property was found not to be an inherent drawback to application of disaggregate modeling. Procedures have been developed to identify and account for violations of the property should they occur.

As disaggregate approaches have gained wider acceptance, they have also been subjected to closer scrutiny and criticism. This process has resulted in several significant developments in recent years:

1. The issue of separable (or sequential) versus joint specification of model structures (*94*): substantial refinement in understanding the differences and similarities of the two model forms has occurred, particularly in using the "inclusive price" concept (or more properly, "inclusive value" when expressed as the "log of the denominator").
2. Computational efficiency in estimating the multinomial probit (MNP) model form (*87*): recent developments in applying the Clark Method to model estimation have made the MNP specification a practical alternative to MNL. Although MNP offers a more general model form, it is not clear, however, whether the increased complexity is either necessary or even desirable for many practical applications.
3. Aggregation approaches (*82*): sophisticated approaches to apply the Clark method, a random sample (perhaps synthesized), market segmentation, and Monte Carlo methods for integration are now practical in many applications.
4. Choice-based samples (*67*): means are now available to use "on-board" surveys rather than home-interview surveys and nevertheless produce valid results. This development has important implications for efficient surveying techniques.

SUGGESTED RESEARCH

However, there still is a long way to go in solving problems that are impediments to more widespread applciation of disaggregate models to transportation demand forecasting problems. High on the list for priority in future research are the following:

1. Specification of independent variables—previous research has not always resulted in clear-cut findings on how to specify explanatory variables. For example, there are conflicting findings on whether level-of-service variables should be generic versus alternative-specific. Further research would be useful in resolving such issues.
2. Individual vs. household level of analysis—previous models have explained either household or individual travel choices. Research on the dynamics of decision-making within households, identified but not pursued in-depth in NCHRP Project 8-14 (*11*), would be useful in determining the appropriate level of analysis and how to incorporate the effects of family dynamics in travel choice models.
3. Transferability—researchers have differed on the feasibility of transferring a model estimated from data in one geographic area to a forecast in another area.
4. Practical approaches to implementation—practicing planners are confronted with a bewildering array of complex modeling issues with little guidance regarding what they need to know. Planners need guidelines for priorities in applying the models, more direction from those who have developed the models as to what they must know before applying them, and better information on the likely pitfalls they will face. Chapter Three is a starting point for such guidelines. Also useful in this regard are syntheses of the findings from this and related projects, such as Koppelman's (*115*) discussion at the 1982 Transportation Research Board meeting.
5. Model development for destination choice and trip generation—modelers are confronted with many difficult issues when the model is extended beyond mode choice: in particular, increased likelihood of a violation of the Independence from Irrelevant Alternatives property.

The greatest need for future research appears to be in the area of trip generation. The violation of the IIA assumption, as discussed in detail in Chapter Two, causes the assumptions of the logit model to be violated. A number of alternative modeling approaches have been suggested, such as the Markov process. More research should be conducted in this area. The failure of disaggregate modeling to achieve a significant breakthrough in trip generation must be considered a major disappointment.

6. Research on longer run mobility decisions—although there has been some research on automobile ownership and residential location, and the effects of these long-run decisions on household travel, further research would be useful in developing practical policy-sensitive models for analyzing the impacts of automobile and land-use policies.

Improved methods of collecting and processing disaggregate data are also a concern. Delay in receipt of the final data set from Baltimore was a major disappointment in NCHRP Project 8-13. Clearly one important direction for future research is to explore disaggregate model development using this data set, which was expressly designed for research on disaggregate modeling.

A May 1978 conference sponsored by the U.S. Department of Transportation, Federal Highway Administration, "Directions to Improve Urban Travel Demand Forecasting" (*116*), contains a detailed agenda of research needs for disaggregate modeling which researchers also are encouraged to consider.

DISSEMINATION

A formal, coordinated dissemination program should be initiated to increase awareness of the advantages of disaggregate approaches. This dissemination program should be directed to the potential user in state and local transportation agencies and not the research community. The program should produce manuals and training programs, with emphasis on the use of the results of case studies of actual experience. To the extent that standardized procedures for analyzing travel demand can be developed, they should be emphasized in the dissemination program. The formal widespread dissemination program directed to the ultimate user of the models should not begin until considerable progress has been made in developing new planning tools based on disaggregate demand analysis, and after experience demonstrating the new techniques has been achieved.

REFERENCES

1. CHARLES RIVER ASSOCIATES, "Disaggregate Travel Demand Models, Project 8-13: Phase I Report." Two volumes (Feb. 1976).
2. CHARLES RIVER ASSOCIATES, "Disaggregate Travel Demand Models, Project 8-13: Phase II Report." Cambridge, Mass. (May 1978).
3. BRAND, D., "Travel Demand Forecasting: Some Foundations and a Review." *HRB Special Report 143* (1973) pp. 239-282.
4. GENSCH, D. H., AND SVESTKA, J. A., "An Exact Hierarchical Algorithm for Determining Aggregate Statistics From Individual Choice Data." *Management Science*, Vol. 25 (Oct. 1979) pp. 939-952.
5. U.S. DEPARTMENT OF TRANSPORTATION, Federal Highway Administration. "Urban Transportation Planning." Washington, D.C. (1972).
6. THE INSTITUTE OF TRANSPORTATION STUDIES, "Urban Travel Demand Forecasting Project: Phase I Final Report Series." The University of California, Berkeley (1977).
7. MCFADDEN, D., "The Theory and Practice of Disaggregate Demand Forecasting for Various Modes of Urban Transportation." Paper presented at Seminar on Emerging Transportation Planning Methods, Daytona Beach, December 5-9, 1976.
8. SOSSLAU, A. B., HASSAM, A. B., CARTER, M. M., AND WICKSTROM, G. V., "Travel Estimation Procedures for Quick Response to Urban Policy Issues." *NCHRP Report 186* (1978) 70 pp.
9. SOSSLAU, A. B., HASSAM, A. B., CARTER, M. M., AND WICKSTROM, G. V., "Quick Response Urban Travel Estimation Techniques and Transferable Parameters: User's Guide." *NCHRP Report 187* (1978) 229 pp.
10. FRIED, M., HAVEN, J., AND THALL, M., "Travel Behavior—A Synthesized Theory: Final Report." NCHRP Project 8-14A final report, Laboratory of Psychosocial Studies, Boston College (1977).
11. ALLAMAN, P. M., TARDIFF, T. J., AND DUNBAR, F. C., "New Approaches to Understanding Travel Behavior." *NCHRP Report 250* (Sept. 1982) 142 pp.
12. TALVITIE, A., DEHGHANI, Y., SUPERNAK, J., MORRIS, M., ANDERSON, M., DEJOHN, A., KOUSHESHI, B., AND FEUERSTEIN, S., "Refinement and Application of Individual Choice Models in Travel Demand Forecasting" (2 volumes). Prepared for the Federal Highway Administration. State University of New York, Buffalo (1981).
13. CHARLES RIVER ASSOCIATES, "A Disaggregated Behavioral Model of Urban Travel Demand." Prepared for the Federal Highway Administration. Cambridge, Mass. (1972).
14. MCFADDEN, D., "Conditional Logit Analysis of Qualitative Choice Behavior." *In* P. Zaremka, Ed., *Frontiers in Econometrics*, Academic Press (1974).
15. LIOU, P. S., AND HARTGEN, D. T., "Issues for Implementing Disaggregate Travel-Demand Models." *In* P. R. Stopher and A. H. Meyburg, Eds., *Behavioral Travel-Demand Models*, D. C. Heath (1976).
16. BEN-AKIVA, M., AND LERMAN S., "Disaggregate Travel and Mobility Choice Models and Measures of Accessibility." *In* D. A. Hensher and P. R. Stopher, Eds., *Behavioral Travel Modelling*, Croom Helm, London (1979).
17. DAGANZO, C., *Multinomial Probit: The Theory and Its Application to Demand Forecasting*. Academic Press (1979).
18. PARODY, T., "Technique for Determining Travel Choices for a Model of Nonwork Travel." *Transportation Research Record 673* (1978) pp. 47-53.
19. MCFADDEN, D., TALVITIE, A. P. AND ASSOCIATES, "Demand Model Estimation and Validation." *The Urban Travel Demand Forecasting Project, Phase I Final*

Report Series, Vol. V., National Technical Information Service (June 1977).

20. TRAIN, K., AND MCFADDEN, D., "The Goods/Leisure Tradeoff and Disaggregate Work Trip Mode Choice Models." *Transportation Research,* Vol. 12 (1978) pp. 349-353.
21. BERECHMAN, J., AND PAASWELL, R. E., "The Impacts of Car Availability on Urban Transportation Behavior." Urban Studies Group Working Paper, Department of Civil Engineering, State University of New York, Buffalo (Aug. 1975).
22. BEN-AKIVA, M. E., AND ATHERTON, T. J., "Methodology for Short-range Travel Demand Prediction: Analysis of Carpooling Incentives." *Journal of Transport Economics and Policy,* Vol. XI, No. 3 (1977) pp. 224-261.
23. LANCASTER, K. J., "A New Approach to Consumer Theory." *Journal of Political Economy,* Vol. 74 (1966) pp. 132-157.
24. QUANDT, R., AND BAUMOL, W. J., "The Demand for Abstract Modes: Theory and Measurement." *Journal of Regional Science,* Vol. 6 (1966) pp. 13-26.
25. ADLER, T. J., AND BEN-AKIVA, M., "Joint-Choice Model for Frequency, Destination and Travel Mode for Shopping Trips." *Transportation Research Record 569,* (1976) pp. 136-150.
26. MCFADDEN, D., "The Measurement of Urban Travel Demand." *Journal of Public Economics,* Vol. 3 (1974) pp. 303-328.
27. CAMBRIDGE SYSTEMATICS, INC., "The Use of Disaggregate Travel Demand Models to Analyze the Energy Conservation Potential of Parking Restrictions Within the Metropolitan Area." Prepared for the Federal Energy Administration, Office of Contingency Planning. Cambridge, Mass. (1976).
28. TRAIN, K. E., "The Sensitivity of Parameter Estimates to Data Specification in Mode Choice Models." *Transportation,* Vol. 7. (1978) pp. 301-309.
29. TALVITIE, A., AND DEHGHANI, Y., "Comparison of Observed and Coded Network Travel Time and Cost Measurements." *Transportation Research Record* 723 (1979) pp. 46-51.
30. TARDIFF, T. J., "Specification Analysis for Quantal Choice Models." *Transportation Science,* Vol. 13, (1979) pp. 179-190.
31. HOROWITZ, J., "Sources of Error and Uncertainty in Behavioral Travel Demand Models." *In* P. R. Stopher, A. H. Meyburg, and W. Brog, Eds., *New Horizons in Travel-Behavior Research,* D. C. Heath (1981).
32. MCFADDEN, D. AND REID, F., "Aggregate Travel Demand Forecasting From Disaggregated Behavioral Models." *Transportation Research Record 534* (1975) pp. 24-37.
33. DOBSON, R., AND TISCHER, M. L., "Comparative Analysis of Determinants of Modal Choices by Central Business District Workers." *Transportation Research Record 649,* (1977) pp. 7-14.
34. GILBERT, G., AND FOERSTER, J. F., "The Importance of Attitudes in the Decision to Use Mass Transit." *Transportation,* Vol. 6 (1977) pp. 321-332.
35. DOBSON, R., DUNBAR, F., SMITH, C. J., REIBSTEIN, D., AND LOVELOCK, C., "Structural Models in the Analysis of Traveler Attitude—Behavior Relationships." *Transportation,* Vol. 7 (1978) pp. 351-363.
36. TARDIFF, T. J., "Causal Inferences Involving Transportation Attitudes and Behavior." *Transportation Research,* Vol. 11 (1977) pp. 397-404.
37. STOPHER, P. R., AND LAVENDER, J. O., "Disaggregate Behavioral Travel Demand Models: Empirical Tests of Three Hypotheses." *Proceedings of the Transportation Research Forum,* Vol. 13 (1972) pp. 321-335.
38. WATSON, P. L., "A Composite Model of Inter-city Mode Choice." International Conference on Transportation Research of the Transportation Research Forum, Bruges, Belgium, June 1973.
39. OUM, T. H., "A Warning on the Use of Linear Logit Models in Transport Mode Choices Studies." *The Bell Journal of Economics,* Vol. 10 (1979) pp. 374-388.
40. KOPPELMAN, F. S., "Non-Linear Utility Functions in Models of Travel Choice Behavior." *Transportation,* Vol. 10 (1981) pp. 127-146.
41. FEDERAL HIGHWAY ADMINISTRATION, "Nationwide Personal Transportation Study." U.S. Department of Transportation, Washington, D.C. (1972-1974).
42. CHARLES RIVER ASSOCIATES, "Estimating the Effects of Urban Travel Policies." Prepared for the Office of the Secretary, U.S. Department of Transportation. Cambridge, Mass. (1976).
43. THEIL, H., *Principles of Econometrics*. John Wiley and Sons, Inc. (1971) pp. 396-397.
44. MCFADDEN, D., "The Revealed Preferences of a Public Bureaucracy: Theory." *The Bell Journal of Economics,* Vol. 6 (1975) pp. 401-416.
45. TARDIFF, T. J., "A Note on Goodness-of-Fit Statistics for Probit and Logit Models." *Transportation,* Vol. 5 (1976) pp. 377-388.
46. MCFADDEN, D., "Quantal Choice Analysis: A Survey." *Annals of Economic and Social Measurement,* Vol. 5 (1976) pp. 363-390.
47. WATANATADA, T., AND BEN-AKIVA, M. E., "Spatial Aggregation of Discrete Choice Models: Areawide Urban Travel Sketch-Planning Model." *Transportation Research Record 673* (1978) pp. 93-99.
48. MINISTRY OF TRANSPORT, "Sigmo Study." Four Volumes. Den Hagg, The Netherlands (Sept. 1977).
49. BEN-AKIVA, M. E., SHERMAN, L., AND KULLMAN, B., "Disaggregate Travel Demand Models for the San Francisco Bay Area: Non-Home-Based Models." *Transportation Research Record 673* (1978) pp. 128-133.
50. DUNBAR, F. C., Discussion of "Disaggregate Travel Demand Models for the San Francisco Bay Area." *Transportation Research Record 673* (1978) pp. 133-135.
51. JONES, P. M., "Identifying the Determinants of Destination Choice." *In* D. A. Hensher, Ed., *Identification and Evaluation of the Determinants of Travel Choice*. Saxon House, London (1977).
52. LERMAN, S. R., "The Use of Disaggregate Choice Models in Semi-Markov Process Models of Trip Chaining Behavior." *Transportation Science,* Vol. 13 (1979) pp. 273-291.
53. LERMAN, S. R., FLUSBERG, M., PECKNOLD, W. M., NES-

TLE, R. E., AND WILSON, N. H. M., "A Model System for Forecasting Patronage on Demand Responsive Transportation Systems." *Transportation Research,* Vol. 14A (1980) pp. 13-23.
54. ADLER, T., AND BEN-AKIVA, M., "A Theoretical and Empirical Model of Trip Chaining Behavior." *Transportation Research,* Vol. 13B (1979) pp. 243-257.
55. HOROWITZ, J., "Disaggregate Demand Model for Nonwork Travel." *Transportation Research Record 673* (1978) pp. 65-71.
56. SPEAR, B. D., "Applications of New Travel Demand Forecasting Techniques to Transportation Planning—A Study of Individual Choice Models." U.S. Department of Transportation, Federal Highway Administration, Urban Planning Division (Mar. 1977).
57. TOBIN, J., "Estimation of Relationships for Limited Dependent Variables." *Econometrica,* Vol. 26 (1958) pp. 24-36.
58. MCKELVEY, R. D., AND ZAVONIA, W., "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *Journal of Mathematical Sociology,* Vol. 4 (1975) pp. 103-120.
59. RUYGROK, C. J., AND VAN ESSEN, P. G., "Development and Application of a Disaggregate Poisson Model for Trip Generation." Presented at the Fifty-Ninth Annual Meeting of the Transportation Research Board, Washington, D.C., January 1980.
60. JOHNSON, T., "Qualitative and Limited Dependent Variables in Economic Relationships." *Econometrica,* Vol. 40, No. 3 (1972) pp. 455-461.
61. BEALS, R. E., *Statistics for Economists.* Rand McNally (1972).
62. SHEFFI, Y., "Estimating Choice Probabilities Among Nested Alternatives." *Transportation Research,* Vol. 13B (1979) pp. 189-205.
63. HOROWITZ, J., "Confidence Intervals for the Choice Probabilities of the Multinomial Logit Model." *Transportation Research Record 728* (1979) pp. 23-29.
64. DAGANZO, C. F., "The Statistical Interpretation of Predictions with Disaggregate Demand Models." *Transportation Science,* Vol. 13 (1979) pp. 1-12.
65. RICHARDS, M. G., AND BEN-AKIVA, M. E., *A Disaggregated Travel Demand Model.* Lexington Books (1975).
66. U.S. DEPARTMENT OF TRANSPORTATION, Federal Highway Administration, Urban Planning Division. "Design of Procedures to Evaluate Traveler Responses to Changes in Transportation System Supply: Conference Summary and White Papers." Washington, D.C. (Sept. 1974).
67. LERMAN, S. R., MANSKI, C., AND ATHERTON, T. J., "Non-Random Sampling in the Calibration of Disaggregate Choice Models." Prepared for U.S. Department of Transportation, Washington, D.C. (1976).
68. ATHERTON, T. J., AND BEN-AKIVA, M. E., "Transferability and Updating of Disaggregate Travel Demand Models." *Transportation Research Record 610* (1976) pp. 12-18.
67. CHARLES RIVER ASSOCIATES, "Collection of a Disaggregate Data Set: Interim Technical Report." Boston, Mass. (Jan. 1978) 201 pp.
70. DAGANZO, C. F., "Calibration and Prediction with Random Utility Models: Some Recent Advances and Unresolved Questions." *In* P. R. Stopher, A. H. Meyburg, and W. Brog, Eds., *New Horizons in Travel-Behavior Research,* D. C. Heath (1981).
71. DAGANZO, C. F., AND SHEFFI, Y., "Estimation of Choice Models from Panel Data." Presented at the Fifth-ninth Annual Meeting of the Transportation Research Board, Washington, D.C., 1980.
72. BEGGS, S., CARDELL, S., AND HAUSMAN, J., "Assessing the Potential Demand for Electric Cars." *Journal of Econometrics,* Vol. 16 (1981) pp. 1-19.
73. HOROWITZ, J., "Identification and Diagnosis of Specification Errors in the Multinomial Logit Model." *Transportation Research,* Vol. 15B (1981) pp. 345-360.
74. HOROWITZ, J. L., "Sources of Error and Uncertainty in Behavioral Travel Demand Models." *In* Meyburg, A., Stopher, P., and Brog, W., Eds., *New Horizons in Behavioral Travel Research,* D. C. Heath and Company. (1981).
75. CHAN, Y., *Review and Compilation of Demand Forecasting Experiences: An Aggregation of Estimation Procedures.* Prepared for the U.S. Department of Transportation by the Pennsylvania Transportation Institute (June 1977).
76. MCFADDEN, D. L., TYE, W., AND TRAIN, K., "Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model." Working Paper No. 7616, Travel Demand Forecasting Project. Institute of Transportation Studies, University of California, Berkeley (1976).
77. HOROWITZ, J., "The Accuracy of the Multinomial Logit Model as an Approximation to the Multinomial Probit Model of Travel Demand." *Transportation Research* Vol. 14B (1980) pp. 331-341.
78. HAUSMAN, J. A., AND WISE, D. A., "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." *Econometrica,* Vol. 46 (1978) pp. 403-426.
79. HAUSMAN, J. A., AND MCFADDEN, D., "Specification Tests for the Multinomial Logit Model." *Econometrica,* Vol. 49 (1983) (forthcoming).
80. WILLIAMS, H. C. W. L., AND ORTUZAR, J. D., "Behavioural Theories of Dispersion and the Mis-Specification of Travel Demand Models." *Transportation Research,* Vol. 16B (1982) pp. 167-219.
81. ALBRIGHT, R. L., LERMAN, S. R., AND MANSKI, C. F., "Report on the Development of an Estimation Program for the Multinomial Probit Model." Cambridge Systematics (1977).
82. BOUTHELIER, F., AND DAGANZO, C. F., "Aggregation with Multinomial Probit and Estimation of Disaggregate Models with Aggregate Data: A New Methodological Approach." *Transportation Research,* Vol. 13B (1979) pp. 133-146.
83. MCFADDEN, D., "Modeling the Choice of Residential Location." *Transportation Research Record 673,* (1978) pp. 72-77.
84. RECKER, W. W., AND GOLOB, T. F., "An Attitudinal Modal Choice Model." *Transportation Research,* Vol. 10 (1976) pp. 299-310.
85. KITAMURA, R., "Urban Travel Demand Forecasting by Stratified Choice Models." Ph.D. Dissertation, Depart-

ment of Civil Engineering, University of Michigan (1978).

86. SOBEL, K. L., "Travel Demand Forecasting by Using the Nested Multinomial Logit Model." *Transportation Research Record 775* (1980) pp. 48-55.
87. DAGANZO, C. F., BOUTHELIER, F., AND SHEFFI, Y., "Multinomial Probit and Qualitative Choice: A Computationally Efficient Algorithm." *Transportation Science,* Vol. 11 (1977) pp. 338-358.
88. LERMAN, S. R., AND MANSKI, C. F., "An Estimator for the Generalized Multinomial Probit Choice Model." Paper presented at the Fifty-sixth Annual Meeting of the Transportation Research Board, Washington, D.C., January 1977.
89. HOROWITZ, J., "A Note on the Accuracy of the Clark Approximation for the Multinomial Probit Model." U.S. Environmental Protection Agency, Washington, D.C. (1978).
90. JOHNSON, L. W., AND HENSHER, D. A., "A Random Coefficient Model of the Determinants of Frequency of Shopping Trips." *Australian Economic Papers* (1979).
91. BEN-AKIVA, M. E., "Structure of Passenger Travel Demand Models." Ph.D. Dissertation, Department of Civil Engineering, Massachusetts Institute of Technology (1973).
92. DOMENCICH, T., AND MCFADDEN, D., *Urban Travel Demand: A Behavioral Analysis.* North-Holland Press, Amsterdam (1975).
93. MCFADDEN, D. L., "The Mathematical Theory of Demand Models." *In* P. R. Stopher and A. H. Meyburg, Eds. *Behavioral Travel-Demand Models*, D. C. Heath (1976).
94. MCFADDEN, D., "Quantitative Methods for Analyzing Travel Behavior of Individuals: Some Recent Developments." *In* David A. Hensher and Peter R. Stopher, Eds., *Behavioural Travel Modeling*, Croom Helm, London (1979).
95. GAUDRY, M. J. I., AND DAGINAIS, M. G., "The Dogit Model." *Transportation Research,* Vol. 13B (1979) pp. 105-111.
96. Chicago Area Transportation Study. "Travel Forecasting Process." Chicago, Ill. (Apr. 1979).
97. ATHERTON, T. J., SUHRBIER, J. H., AND JESSIMAN, W. A., "Use of Disaggregate Travel Demand Models to Analyze Carpooling Policy Incentives." *Transportation Research Record 599* (1976) pp. 35-40.
98. WESTIN, R. B., "Predictions from Binary Choice Models." *Journal of Econometrics,* Vol. 2 (1974) pp. 1-16.
99. DALY, A. J., "Research on Transport Demand Models." Proceedings of the Transportation Research Forum, Vol. 15 (1974) pp. 354-362.
100. WATSON, P. L., AND WESTON, R. B., "Transferability of Disaggregate Mode Choice Models." *Regional Science and Urban Economics,* Vol. 5 (1975) pp. 229-249.
101. PEAT, MARWICK, MITCHELL AND CO., "Implementation of the N-Dimensional Logit Model." Washington, D.C. (1972).
102. TALVITIE, A. AND KIRSHNER, D., "Specification, Transferability, and the Effect of Data Outliers in Modeling the Choice of Mode in Urban Travel." *Transportation,* Vol. 7 (1978) pp. 311-331.
103. GOMEZ-IBANEZ, J. A., FAUTH G. R., AND FERRIS, S., "Using Demand Elasticities from Disaggregate Mode Choice Models." Discussion Paper D78-79. Department of City and Regional Planning, Harvard University (Dec. 1978).
104. U.S. CONGRESS, Office of Technology Assessment. "Energy, the Economy, and Mass Transit." Washington, D.C. (Dec. 1975).
105. SHUNK, G. A., AND KOLLO, H. P. H., "Disaggregate Travel Demand Models for the San Francisco Bay Area: Discussions." *Transportation Research Record 673* (1978) pp. 135-136.
106. HENSHER, D. A., AND STOPHER, P. R., "Behavioral Travel Modelling." *In* Hensher, David A. and Stopher, Peter R., Eds., Behavioral Travel Modelling, Croom Helm, London (1979).
107. RUITER, E. R., AND BEN-AKIVA, M. E., "Disaggregate Travel Models for the San Francisco Bay Area: System Structure, Component Models, and Application Procedures." *Transportation Research Record 673* (1978) pp. 121-128.
108. CAMBRIDGE SYSTEMATICS, INC., "Analytic Procedures for Urban Transportation Energy Conservation." Five Volumes. Prepared for the Assistant Secretary for Policy and Evaluation, U.S. Department of Energy. Cambridge, Mass. (Oct. 1979).
109. CHARLES RIVER ASSOCIATES, Policy Evaluation with Travel Behavior Models. U.S. Department of Transportation and the U.S. Environmental Protection Agency. Cambridge, Mass. (Jan. 1979).
110. CAMBRIDGE SYSTEMATICS, INC., "Work Sheets from a Manual of Sketch Planning Demand Analysis for a Preferential Bus/Carpool Lane Example." Prepared for presentation at the Seminar on Emerging Transportation Planning Methods. Jointly sponsored by the Technology Sharing Program and the University Research Program, and the Office of the Secretary, U.S. Department of Transportation (Dec. 6, 1976).
111. KUMAR, A., "Use of Incremental Form of Logit Models in Demand Analysis." *Transportation Research Record 775* (1980) pp. 21-27.
112. MANHEIM, M. L., FURTH, P., AND SALOMON, I., "Responsive Transportation Analysis: Pocket Calculator Methods. Vol. 1: Introduction to Calculator Methods; Vol. 2: Examples." Working Paper CTS-Ramp-78-1 (Draft). Responsive Analysis Methods Project, Center for Transportation Studies, Massachusetts Institute of Technology (Apr. 11, 1978).
113. KAVAK, F. C., AND DEMETSKY, M. J., "Behavioral Modeling of Express Bus-Fringe Parking Decisions." *Transportation Research Record 534* (1975) pp.10-23.
114. CHARLES RIVER ASSOCIATES, "Ridesharing Market Study." Prepared for the Port Authority of New York and New Jersey. Cambridge, Mass (July 1980).
115. KOPPELMAN, F. S., "Summary of Recently Completed Projects Under Sponsorship of National Cooperative Highway Research Program and Federal Highway Administration." Presented at the meeting of the Transportation Research Board, Washington, D.C., January 1982.
116. U.S. DEPARTMENT OF TRANSPORTATION, Federal Highway Administration and Urban Mass Transportation Administration. "Directions to Improve Urban Travel Demand Forecasting: Conference Summary and White Papers." National Technical Information Service (1978).

APPENDIX A

PREVIOUS RESEARCH ON THE AGGREGATION OF DISAGGREGATE DEMAND MODELS

When confronted with a policy proposal whose impacts are to be evaluated at an aggregate level (i.e., the effect on more than one individual) and a model that relates the behavior of interest to the policy variable (and, possibly, to other variables also), the obvious approach to performing the policy evaluation is to replace the independent variables in the model with their "before" and "after" averages for the impacted population to observe the before and after aggregate behavior, from which the aggregate policy effect can be inferred. Unfortunately, if the model being used is a nonlinear one, such as the logit model, this approach will generally produce incorrect predictions of both the before and after behavior, as well as incorrect predictions of the difference between before and after behavior. The correct procedure with nonlinear disaggregate models is: 1) to predict the before behavior of each individual in the population and take the average of these predictions; 2) to predict the after behavior for each individual and take the average; and 3) to take the differences between these average predictions to obtain the aggregate policy effect. This procedure has been called the complete

enumeration method and we will refer to the obvious approach of using averages of the independent variables as the direct aggregation (sometimes "naive") method. The predictions obtained by the two methods will differ, and this difference is the aggregation bias that is described in Appendix B.

Included in our discussion of aggregation bias in Appendix B is a list of four properties of aggregation bias. While they can rigorously be shown to hold only under certain conditions, intuitively they can be expected to apply in many, if not most cases. Of these four properties, it is the third of these which has received the most attention in the search for methods to reduce aggregation bias. This property stated that the magnitude of the aggregation bias increases as the dispersion in the values of representative utility difference, V, increases. Obviously, if there were no dispersion in the value of V, there would be no aggregation bias. Consequently, it is this property that has motivated the various approaches that have been suggested to reduce aggregation bias without completely abandoning the computational simplicity afforded by the use of the direct aggregation method. The following sections present several proposed methods designed to reduce the level of aggregation bias in forecasting with disaggregate demand models.

TALVITIE'S METHOD: THE TAYLOR'S SERIES EXPANSION

The first approximation method to be discussed is one proposed by Talvitie (A1). Focusing on the binary logit model:

$$P_t(A) = \frac{1}{1 + e^{-V_t}} \tag{A-1}$$

where $P(A)$ is the probability that individual t will choose alternative A and V_t the value of V for individual t. Taking a Taylor's series expansion about $\bar{V}$, the mean representative utility for the population, gives

$$P_t(A) = \widehat{P(A)} + (V_t-\bar{V})\,\widehat{\frac{dP(a)}{dV}} + \frac{1}{2}(V-\bar{V})^2\,\widehat{\frac{d^2P(a)}{dV^2}} + \ldots \tag{A-2}$$

where the symbol "$\widehat{\ }$" denotes the function evaluated at $\bar{V}$. Making note of the fact that

$$\frac{d^2P(a)}{dV^2} = 2\cdot P(a)\cdot[2-P(a)]\cdot[1/2 - P(a)], \tag{A-3}$$

truncating Equation A-2 after the third term, and taking averages of both sides yields the approximation:

$$\overline{P(a)} = \widehat{P(a)} + Var(V)\cdot\widehat{P(a)}\cdot[1-\widehat{P(a)}]\cdot[1/2-\widehat{P(a)}] \tag{A-4}$$

where $\overline{P(a)}$ is the average of the individual probabilities and $Var(V)$ is the sample variance of the values of the representative utility. That is:

$$Var(V) = \frac{1}{T}\sum_{t=1}^{T}(V_t-\bar{V}^2). \tag{A-5}$$

Equation A-4 states that the share predicted by the enumeration method is approximately equal to the share pre-

dicted by the direct aggregation method plus a correction term which depends jointly on the sample variance of V and the value of the share predicted by the direct aggregation method. This approximation takes account of the four properties stated in Appendix B.

Talvitie's method might be extremely valuable if its level of computational simplicity were about the same as the simplicity of the direct aggregation method. Unfortunately, this is not the case.

Application of Talvitie's method requires a value for $Var(V)$, and this normally cannot be derived from aggregate data. One might compute $Var(V)$ directly from a sample, but doing so is no easier than estimating $\overline{P(A)}$ by applying the the logit equation to each data point and aggregating (the "random sample enumeration method"). Alternately, one can attempt to make use of the fact that, since $V = \underline{\alpha}'\underline{X}$, where $\underline{X}$ is the vector of (differences in) the independent variables and $\underline{\alpha}$ is the vector of coefficients, then $Var(V) = \underline{\alpha}'Z\underline{\alpha}$, where Z is the matrix of sample variances and covariances of (differences in) the independent variables. However, the analyst is unlikely to know the elements of Z, and to estimate is no easier than computing $\overline{P(A)}$.

Even if the analyst had estimates of the variance and covariance terms supplied from another source, it is unclear as to how these might be adjusted intuitively to reflect the particular aspects of the population being studied. For example, suppose one had a model, together with the variance and covariance terms, estimated on a New York City sample and one wished to use this model in an application for Los Angeles. Even if the coefficients can be assumed to apply equally well to the two cities, it is likely that the variances of, say, transit in-vehicle time are different for the two cities. Furthermore, it is unclear how the variance estimated on New York data should be adjusted to reflect conditions in Los Angeles. We know that it should be increased, but there is no way to know how much. This difficulty severely hinders the ready application of Talvitie's method by persons not knowledgeable in statistical theory. Furthermore, Reid (A2) has shown that this procedure can result in substantial error (even larger than the error resulting from the direct aggregation method).

WESTIN'S METHOD

A somewhat different aggregation method has been proposed by Westin (A3). Viewing the (binary) logit equation

$$P_t(A) = \frac{1}{1 + e^{-V_t}} \qquad \text{(A-6)}$$

as a transformation of V_t to $P_t(A)$, Westin reasons that if the distribution in the population of the values of V_t were known, then the (population) distribution of the values of $P_t(A)$ could be derived from the transformation given by Equation A-2. Once the distribution of $P(A)$ is known, the

mean of that distribution is the exact aggregate share of alternative A. In order to obtain a distribution for V, Westin makes the assumption that the vector of independent variables, $\underline{X}$, is multivariate normal with a mean vector $\underline{\mu}_X$ and covariance matrix Σ_X. Then $V = \underline{\alpha}'\underline{X}$ is univariate normal with mean $\mu_V = \underline{\alpha}'\underline{\mu}_X$ and variance $\sigma_V^2 = \underline{\alpha}'\Sigma\underline{\alpha}$. The distribution of V being known, the distribution of $P(A)$ can be derived, the mean of the $P(A)$ distribution being the predicted choice share.

The advantage of Westin's procedure is that, if the assumption that the independent variables are multivariate normal is true, the choice shares predicted by his method are exact, i.e., aggregation bias is completely eliminated. The disadvantages of his approach are many, however. First, the independent variables may well be non-normal, especially truncated variables such as income. Second, the analytical expression for the mean of the distribution of $P(A)$ is intractable and the mean must be computed by numerical integration techniques, a nontrivial computational exercise. Third, and perhaps most important, the results have been worked out only for the binary case. Extensions to the multinomial case were not presented, nor is it clear that they may be readily derived.

AGGREGATION WITH PROBIT MODELS

McFadden and Reid (A4) have applied Westin's procedure to the binary probit model. For an individual, the probit model yields the following probability

$$P_i = \Phi(V_i) \tag{A-7}$$

where P_i is the probability of individual i selecting the first alternative, V_i is the difference in the utility functions for the two alternatives, and Φ is the standard normal distribution function. If, as both Westin and McFadden and Reid assume, V_i is normally distributed, then the population proportion is given by

$$\overline{P} = \Phi\left(\frac{\mu_V}{1 + \sigma_V^2}\right) \tag{A-8}$$

where μ_V and σ_V^2 are defined the same as in Westin's method and $\overline{P}$ is the population proportion.

Since the procedure is very similar to Westin's method, it shares many of the same properties. In particular, information on the variances of the independent variables is needed in estimating σ_V^2 . (The probit example clearly illustrates how heterogeneity among individuals biases prediction. If individuals were the same, $\sigma_V^2 = 0$ and the direct method would yield the correct predictions.) Since information of this nature may be difficult to obtain, the practical usefulness of these methods has not been completely established.

The probit method produces aggregate shares from a normal distribution. Since this distribution is more common than that resulting from Westin's method (the S_B distribution), the probit approach might be preferable on practical grounds. Since the binary probit and logit models yield

almost identical probabilities, practical considerations are especially important in selecting the model.

Bouthelier and Daganzo (A5) have extended McFadden and Reid's method to the case of multinomial probit. They have also demonstrated how existing or slightly modified probit software can be used in estimating probit models from aggregate data, e.g., zonal data. Like the binary case, information on the variance structure of the independent variables is necessary in applying this method.

THE RANDOM SAMPLE ENUMERATION AND MARKET SEGMENTATION APPROACHES

The last two aggregation procedures to be described represent straightforward compromises between the complete enumeration method and the direct aggregation method. These procedures are the random sample enumeration method and the market segmentation approach. Although he makes no claim to being their originator, these methods have been researched most thoroughly by Frank Koppelman (A6, A7, A8). The random sample enumeration method is identical to the complete enumeration method except that, as its name implies, predicted choice probabilities are computed for only a subset of the population being studied. Although the complete enumeration method was defined as taking the average of the predicted probabilities for all individuals in the population, in practice this is, of course, impossible and the average choice probability of all the individuals in the calibration sample is calculated. In the random sample enumeration method, only a subsample of the calibration sample is used, or, if the model is being transferred, an entirely different sample is used. The rationale for the random sample enumeration method is that not all of the precision required for model calibration is necessary for prediction, and thus only a portion of the observations need be retained when applying a previously calibrated model. The drawback of this method is that it requires a disaggregate data set, albeit a comparatively small one, which may not always be available.

A variation on the random sample enumeration method has been proposed as part of the Urban Travel Demand Forecasting Project at the University of California (A9). This approach involves the use of a computer model to generate a synthetic representative sample of households (SYNSAM) based on Census data and projections of population and economic conditions. Once the synthetic sample is provided, it is used for forecasting in a manner similar to an ordinary random sample.

The last aggregation method to be discussed is the market segmentation or "classification" approach. This approach seeks to group the data into market segments so that the variance of V is minimized within any group and maximized across groups. Reducing the variance of V within any segment should reduce the aggregation bias that results

when the model is applied to the means of the independent variables for that market segment. Aggregate choice share predictions are obtained by taking the weighted average of the probabilities of, say, choosing auto for the various market segments, where the weights are the segments' shares of all trips. For policy analysis, aggregate predictions are calculated for the baseline choice shares and then again after the policy variables have been changed.

The motivation for using the market segmentation approach can be seen by considering what happens if this approach is followed to its logical extreme, i.e., if a market segment is created for each individual. At this extreme, the market segmentation approach and the complete enumeration method become identical. The objective in the market segmentation method is to approximate this extreme case with a relatively small number of segments, the assumption then being made that the trips and tripmakers within any one segment are not significantly different enough to affect the results.

There is no a priori constraint on how the segments are to be defined other than the desire to create relatively homogeneous groups. Any grouping which successfully achieves a reduction in the variability of V would be a satisfactory market segmentation. Koppelman (A10) has shown that a successful first step in defining the market segments is to group the observations into segments which are defined on the "relevance" of the various choices. Thus, in a binary (auto/transit) mode split example, an obvious initial segmentation would be to establish one segment for trips taken by individuals whose households do not have an automobile available, another segment for trips taken by individuals who do not have transit available (i.e., there is no transit stop within a reasonable walking distance), and a third segment containing all other trips. The first segment would be assumed to choose only transit, the second only auto, and the model could be applied to the third segment to determine the mode split.

To achieve further reduction in the variability of V within the segment whose members have more than one choice available to them, the best approach appears to be to define the segments grouping the observations so that those observations with approximately the same values of the independent variables are grouped together. Recall from the discussion of Talvitie's approach that the variance of V could be expressed as the (weighted) sum of the variances and covariances of the independent variables. By grouping similar values of the independent variables together, we obviously reduce their variances and covariances within each group to levels less than the entire population.

Reid (A2) developed a procedure for reducing the variance in V which involves a stratification based upon values on V. The advantage of this procedure is that it involves only the single composite variable V rather than several variables as is typical of segmentation based upon independent variables. However, since standard sources of aggregate data such as Census data do not correspond closely to the utility functions of travel demand models, this procedure may not be usable with many existing data sources.

Koppelman's results (A6) indicate the market segmentation approach performed about the same or slightly better than other methods. Selection of an aggregation method must also take into account ease of implementation. On this basis, the market segmentation approach again appears to be superior. It relies only on the relative size of the various market segments and the average values of the independent variables for each segment. These are figures for which the analyst is likely to have some intuitive feel (or hard data) with which to improve the transferability of the model from one application to another. Furthermore, the analyst may have some special interest in some of the market segments per se and the market segmentation approach allows him to observe the effect of various policies on each of the market segments. Thus, in addition to knowing the total predicted effect of a proposed policy, the analyst can ascertain the extent to which various members of the population are likely to be affected.

For the reasons outlined in the preceding paragraph, we have elected to design a market segmentation method for applying travel demand models. Our recommendations for the particular market segmentation scheme to be constructed are described in Appendix C.

REFERENCES

A1. Talvitie, Antti. "Aggregate Travel Demand Analysis with Disaggregate or Aggregate Travel Demand Models." Proceedings of the Transportation Research Forum, Vol. 14, pp. 583-603, 1973.

A2. Reid, Fred A. "Minimizing Error in Aggregate Predictions from Disaggregate Models." <u>Transportation Research Record</u>, No. 673, pp. 59-65, 1978.

A3. Westin, Richard B. "Predictions from Binary Choice Models." <u>Journal of Econometrics</u>, Vol. 2, pp. 1-16, May 1974.

A4. McFadden, Daniel and Reid, Fred. "Aggregate Travel Demand Forecasting from Disaggregated Behavioral Models." <u>Transportation Research Record</u>, No. 534, pp. 24-37, 1975.

A5. Bouthelier, F. and Daganzo, C. F. "Aggregation with Multinomial Probit and Estimation of Disaggregate Models with Aggregate Data: A New Methodological Approach." <u>Transportation Research</u>, Vol. 13B, pp. 133-146, 1979.

A6. Koppelman, Frank S. "Travel Predictions with Models of Individual Choice Behavior." Unpublished dissertation, M.I.T., Department of Civil Engineering, May 1975.

A7. Koppelman, F. S. "Prediction with Disaggregate Models: The Aggregation Issue." Transportation Research Record, No. 527, pp. 73-80, 1974.

A8. Koppelman, F. S. "Methodology for Analyzing Errors in Prediction with Disaggregate Choice Models." Transportation Research Record, No. 592, pp. 17-23, 1976.

A9. McFadden, Daniel, et al. Demographic Data For Policy Analysis, Volume VIII. The Urban Travel Demand Forecasting Project, Phase I Final Report Series. National Technical Information Service (1977).

A10. Koppelman, Frank. "Guidelines for Aggregate Travel Prediction Using Disaggregate Choice Models." Transportation Research Record, No. 610, 1976, pp. 19-24.

APPENDIX B

FINDINGS: TRANSFERABILITY AND AGGREGATION ISSUES IN APPLYING DISAGGREGATE DEMAND MODELS

FINDINGS ON MARKET SEGMENTATION

Despite their advantages disaggregate models of travel demand, logit or otherwise, have not been widely used by transportation planners. Aside from the relative novelty of the technique, there are probably two reasons. First, calibration of a disaggregate model requires the existence of a data set containing observations on individual trips. While the amount of data necessary is considerably less than that collected by the massive household interview surveys conducted in many cities over the past two decades, the task of collecting such a data set is a barrier to implementation if a new data set is required for each application. Researchers have hoped that regularities in the behavior of individuals would allow a single model calibrated in one place at one time to be used in applications in other places and at other times, thereby making the calibration of new models unnecessary. However, evidence on the transferability of disaggregate models, while still somewhat inconclusive, precludes the universal application of a single model.

The second obstacle to the widespread application of disaggregate behavioral models is the difficulty encountered in using these models to make aggregate forecasts. Policy makers are usually interested in the total (i.e., aggregate) impact of their policies. For example, suppose a policy analyst wishes to use a logit model which related the probabilities of choosing auto and bus to the costs of traveling by these two modes to predict the effect on bus ridership of a systemwide bus fare increase. The obvious and simplest approach would be to use the urban area average values of auto cost and bus fare to predict the "before" and "after" aggregate modal splits. Unfortunately, this approach will, in general, lead to erroneous forecasts, even if the model being used is the correct one for the population being studied. The most reliable approach is to use the values of cost for each individual in the affected population to predict the individual probabilities and then to sum these probabilities. The forecasts obtained by the disaggregate method will be more accurate than those obtained by the aggregate approach because, in technical terms, "the average of a nonlinear function is not equal to the function evaluated at the averages of the independent variable." The difference between the forecasts obtained by the two approaches is known as the "aggregation bias." Unfortunately, the disaggregate approach requires an explicit aggregation procedure, which substantially increases the complexity of the analysis.

It is important to recognize that the aggregation problem is inherent in the prediction problem and is not a problem that arises because of the disaggregate approach. Aggregate approaches merely "sweep the problem under the rug" by ignoring the aggregation error.

Appendix C develops procedures that enhance the applicability of disaggregate techniques by reducing the impediments to their widespread application imposed by the problems of transferability and aggregation. In developing such procedures, a choice must be made between designing procedures that can be applied in a uniform fashion, regardless of the policy under consideration or the environment into which the policy is to be introduced, and procedures that are customized to fit the situation at hand. The advantage of standardization is in ease of implementation while the advantage of customization is in the accuracy of the fit obtained in the individual cases. Since our primary interest is encouraging the application of disaggregate travel demand models, we have chosen a standardized approach using market segmentation. Thus, we will propose procedures that will enable analysts to use market segments with disaggregate models to evaluate a wide range of policies with a minimum of input data and computations. To the extent possible, these procedures will be invariant to the particular policy application.

The findings presented in Appendix C use local transportation data and Census and other nationally available data to construct "market segments." In some respects, the market segmentation approach to applying disaggregate models is very similar to the familiar "cross-classification" method of trip generation forecasting. (See U.S. Department of Transportation, Federal Highway Administration, (B1)). Both methods involve the constructing of tables (or cross-tabulation) where values of the variables thought to determine the relevant travel behavior are divided into categories. The intersection of a single category from one variable with a category from a second variable (and possibly with a category from a third variable, a fourth variable, etc.) defines a "cell" or market segment. The purpose of the market segmentation approach is different than that of the cross-classification method, however. In the latter method, the number of trips per household per day, for example, is assumed to be correlated with the variables defining the cells in some unspecified fashion. That is, households falling in a given cell are predicted to make a certain number of trips, with no explicit model which relates tripmaking behavior to the "explanatory" variables. In the market segmentation approach, on the other hand, the behavioral model is quite explicit and the cells are nothing more than a convenient way of disaggregating the population so that the model may be applied.

The discussion of market segmentation is divided into three sections. The rest of this appendix describes in greater detail the transferability and aggregation problems encountered in applying logit models. Chapter Two reviews previous research into the transferability problems and details what can be learned from these studies. Appendix C outlines procedures for constructing market segments from local data, Census, and other national data sources. Appendix C also describes how market segments may be employed for policy evaluation with disaggregate models. For readers interested in pursuing an example, the Phase II report (B2) illustrates the procedures by employing them in an actual case study of the introduction of a park-and-ride facility in Baltimore, Maryland. Further analysis of this example is presented in Appendix C.

TRANSFERABILITY AND AGGREGATION PROBLEMS IN APPLYING DISAGGREGATE DEMAND MODELS

In Chapter Two, the mathematical formulation of the MNL model was derived. A crucial assumption is embodied in this formulation of the systematic portion of the utility function, namely, the parameters β_k, $k=1,\ldots,k$ and $\alpha_l, l=1,\ldots L$, do not vary across individuals. We are thus assuming that all individuals have the same systematic utility functions, or equivalently, that all individuals attach the same weights to the various observed choice attributes and socioeconomic characteristics. Thus, we would predict that individuals with identical observed socioeconomic characteristics who face identical choice alternatives having identical observed attributes (a "homogeneous market segment") would all have the same probabilities of choosing the various alternatives. All members of the homogeneous market segment will not make the same choice, however, because the actual choices are determined by the random, unobserved factors (the ε_{it} component).

Transferability Problems Due to Different "Representative Utilities"

Potential violations of the assumption that all individuals share a common "representative utility" function (V_i) are one possible source of the transferability problem. At one extreme, it can be assumed that there is just one representative utility for all of the trips made by all tripmakers. At the other extreme is the assumption that representative utilities vary from individual to individual, and, for a single individual, that they vary from trip to trip. Existing logit travel demand models have made, explicitly or implicitly, assumptions which fall between these two extremes. The most common assumption is that, for any given urban area, all trips made for a common purpose are made by individuals having identical representative utilities. Thus we have Pittsburgh worktrip models, San Francisco

shopping trip models, etc. There is fairly strong evidence that representative utilities are different for trips with different purposes. It is not clear, however, that the representative utilities for trips with the same purpose vary across urban areas. Indeed, the thrust of the transferability issue is provided by a desire to use models calibrated in one city as the basis for policy evaluation in other cities.

We will use the following example to illustrate the potential for error when an attempt is made to transfer a model from one city to another when, in fact, the representative utilities for the two cities are different.

Suppose the City X Transit Authority wishes to know the decline in bus ridership for worktrips to be expected following a proposed fare increase. Currently, 27 percent of worktrips are made by bus. We intend to use a worktrip modal split model calibrated on City Y to forecast the decrease in bus ridership. The City Y model gives the probabilities of choosing the auto drive-alone mode (as an alternative to the bus passenger mode) as follows:

$$P(A) = \frac{e^{1.0-0.2T_A-10.0C_A}}{e^{1.0-0.2T_A-10.0C_A} + e^{-0.2T_B-10.0C_B}} \qquad \text{(B-1)}$$

where $P(A)$ is the probability of choosing auto and T_A, T_B, C_A and C_B are the travel times and costs of the auto and bus modes. Equation B-1 can be, and usually is rewritten in the equivalent form:

$$\ln\left(\frac{P(A)}{P(B)}\right) = 1.0 - 0.2T - 10.0C \qquad \text{(B-2)}$$

where $P(B)$ is the probability of choosing bus (and, in this binary example, is equal to $1-P(A)$), and $T=T_A-T_B$ and $C=C_A-C_B$.

Suppose, however, that the true model for City X is

$$\ln\left(\frac{P(A)}{P(B)}\right) = 1.0 - 0.04T - 2.0C \qquad \text{(B-3)}$$

To avoid problems of aggregation bias, to be discussed separately below, assume that all of the workers in City X have round trip travel times of 20 minutes by auto and 30 minutes by bus. Also assume that all round trip auto costs are \$0.70. The current systemwide bus fare in City X is \$0.25 (or \$0.50 for a round trip) and the transit authority proposes to raise it to \$0.40 (\$0.80 round trip). Since all tripmakers have identical values for all of the independent variables, the individual probabilities given by Equations B-2 and B-3 are the same as the aggregate shares of the auto and bus modes.

Table B-1 gives the results of this example. The example was chosen so that both models would accurately predict the "before" modal split, namely 73 percent for the auto drive-alone mode and 27 percent for the bus passenger mode. The "incorrect" model predicts a decline in bus ridership to 2 percent of worktrips, however, whereas the "true" model forecasts a fall to 17 percent.

Table B-1

THE TRANSFERABILITY PROBLEM: MODAL SHARES BEFORE AND AFTER BUS FARE INCREASE AS PREDICTED BY TWO MODELS

(Hypothetical Data for Round Trips)

Model Variables	Before Bus Fare Increase	After Bus Fare Increase
Auto Time	20 minutes	20 minutes
Bus Time	30 minutes	30 minutes
Difference in Times (Auto minus Bus)	-10 minutes	-10 minutes
Auto Cost	\$0.70	\$0.70
Bus Cost (fare)	0.50	0.80
Difference in Costs (Auto minus Bus)	0.20	-0.10
Modal Shares Predicted by Two Models (percent of worktrips)		
"Transferred" Model (Eq. B-2)		
Auto	73	98
Bus	27	2
"True" (Eq. B-3)		
Auto	73	83
Bus	27	17

Of course, it is not surprising to find that applying an incorrect model will lead to erroneous forecasts. Admittedly, the foregoing example was exaggerated somewhat to highlight the potential severity of the problem. Nevertheless, differences in the ratio of the parameters of the models in this example are of a plausible order of magnitude in terms of the models which have been calibrated. Thus, this example indicates that the errors associated with applying an incorrect model could be substantial.

In the foregoing example, all workers in City X were assumed to share a common representative utility, and this representative utility was different from the common representative utility shared by workers in City Y. The difference between the two representative utilities resulted in the forecast error. Forecast errors will also result, however, if the assumption that all workers in City X share a common representative utility is itself invalid. It may be, for example, that workers with suburban residences have different representative utilities from workers living in the city. A reason for this might be that the choice of suburban living is indicative of a certain type of lifestyle that has embodied in it a valuation of the attributes of travel different from valuation associated with urban living. Talvitie and Kirshner (B3), in fact, concluded that work mode choice models do not appear to be transferable within regions or between cities. This finding is in contrast to Atherton

and Ben-Akiva's (B4) earlier favorable finding on the transferability of models between cities.

Whatever the differences or the reasons for them, the possibility that not all residents of an urban area have identical representative utilities is certainly plausible. We can illustrate the potential error that can result from incorrectly assuming a common representative utility with the worktrip mode split example that was used above. Assume now, however, that Equation B-2 gives the representative utility of half of the workers in City X (Group 1, say), the representative utility for the other half (Group 2) being given by Equation B-3. Suppose that we have incorrectly assumed a common representative utility and calibrated it to be an average of the two groups:

$$\ln\left(\frac{P(A)}{P(B)}\right) = 1.0 - 0.12T - 6.0C \qquad \text{(B-4)}$$

Table B-2 gives the results of this example. Again, both the incorrect model and the true model correctly estimate the baseline modal shares prior to the policy change. However, the model which incorrectly assumes that all workers share a common representative utility predicts that the bus fare increase will result in bus ridership falling to 6 percent of worktrips while the correct prediction is a decline to 9 percent of worktrips. While the incorrect model does yield an erroneous forecast, this example indicates that the error associated with incorrectly assuming

Table B-2

MODEL MISSPECIFICATION: MODAL SHARES BEFORE AND AFTER BUS FARE INCREASE AS PREDICTED BY TWO MODELS
(Hypothetical Data for Round Trips)

Model Variables	Before Bus Fare Increase	After Bus Fare Increase
Auto Time	20 minutes	20 minutes
Bus Time	30 minutes	30 minutes
Difference in Times (Auto minus Bus)	-10 minutes	-10 minutes
Auto Cost	\$0.70	\$0.70
Bus Cost (Fare)	0.50	0.80
Difference in Costs (Auto minus Bus)	0.20	-0.10

Modal Shares Predicted by Two Models (Percent of Worktrips)

	Before Bus Fare Increase	After Bus Fare Increase
Common Utility Model (Eq. B-4)		
Auto	73	94
Bus	27	6
Group 1 Model (Eq. B-2)		
Auto	73	98
Bus	27	2
Group 2 Model (Eq. B-3)		
Auto	73	83
Bus	27	17
"True Model" (one half of Group 1's share plus one half of Group 2's share)		
Auto	73	91
Bus	27	9

a common representative utility for all members of a group is substantially less than the error that occurs when a model calibrated on one group is applied to another group which has a representative utility different from the first group (under the conditions assumed in the example).

The two types of forecasting errors we have described were both caused by a violation of the assumption that all individuals in the sample share a common representative utility. In the first instance, the error arises because individuals in the prediction group do not share the same representative utility as the individuals in the calibration sample. Therefore the model cannot be transferred from the calibration sample to the prediction group without modification. In the second instance, the error arises because all individuals within the calibration sample itself do not share a common representative utility. The second type of error, while very similar in origin to the transferability problem, is actually a type of aggregation bias which we will refer to as "aggregation misspecification." It would not arise if we calibrated separate models for properly delineated subgroups of the sample or if the model form itself were specified to account for "taste variations" in the population.

"Aggregation Bias"

A second kind of aggregation error is due to the fact that all individuals do not share common personal attributes or attributes of the choice alternatives. Unlike aggregate misspecification, this type of aggregation bias can arise even if we have correctly specified our model. It will occur if we attempt to use the model to make predictions for a group by simply inserting into the behavioral model the group means of the independent variables. The term "aggregation bias" will be used specifically to refer to this type of error.

As before, the potential for error can be described most readily with a hypothetical example. Assume that our (correctly specified) model for worktrip mode choice is the one given in Equation B-4, which we repeat here for convenience.

$$\ln\left(\frac{P(A)}{P(B)}\right) = 1.0 - 0.12T - 6.0C \qquad \text{(B-5)}$$

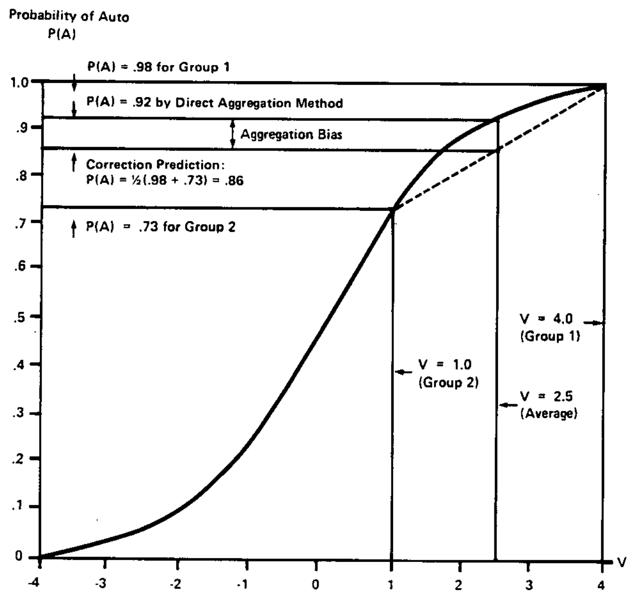
As in the previous examples, assume that all individuals face an auto time of 20 minutes, a bus time of 30 minutes, and an auto cost of \$0.70, all for a round trip. Now suppose, however, that half of the workers live in an "outer zone" where the bus fare is \$0.50 (\$1.00 for a round trip) and the other half live in an "inner zone" with a \$0.25 bus fare (\$0.50 round trip).

If, for those individuals living in the outer zone, we replace the variable T in Equation B-4 with the value -10 (auto time minus bus time for all individuals) and we replace C with the value -0.30 (auto cost minus bus cost for individuals residing in the outer zone) we get a mode split of 98 percent auto and 2 percent bus. Similarly, for individuals living in the inner zone, we replace T with -10 and C with 0.20 for a mode split of 73 percent auto and 27 percent bus. Thus, the true mode split for the entire worktrip population is 86 percent auto $(\frac{1}{2}(93 + 73) = 86)$ and 14 percent bus $(\frac{1}{2}(2 + 27) = 14)$. However, if we attempt to replace C with its population mean value of -0.05, again replacing T with -10, we get a predicted mode split of 92 percent auto and 8 percent bus. Thus, replacing the independent variables with their average values results in the baseline bus ridership being underestimated by upwards of 50 percent in this example. A graphical illustration of the aggregation bias is given in Figure B-1.

At this juncture, a few points regarding aggregation bias should be noted. First, although we posited the existence of different values for one of the independent variables, namely, bus fare, to illustrate the aggregation bias, the aggregation bias was due to "within-group variation" in the values of the difference of the variables

Figure B–1

ILLUSTRATION OF AGGREGATION BIAS
WITH HYPOTHETICAL EXAMPLE



Assume:

(1) $V = 1.0 - 0.12T - 6.0C$;

(2) $P(A) = \frac{1}{1+e^{-V}}$

between the alternatives (auto minus bus, in our example). Thus, had the individuals living in the outer zone had a round trip auto cost of \$0.80 while the individuals living in the inner zone faced a \$0.30 round trip auto cost, there would have been no aggregation bias for the model we have assumed since all individuals would have had a value for the difference in costs of -0.20.

Secondly, the heterogeneity in the differences of the independent variables causes the aggregation bias by creating heterogeneity in the difference between the representative utilities of the alternatives. If we suppress the t subscript from Equation B-5 and define V_{ij} to be the difference between the representative utility of the ith alternative and the representative utility of the jth alternative ($V_{ij} = V_i - V_j$), then Equation B-6 can be rewritten:

$$P(i) = \frac{1}{\sum_{j=1}^{J} e^{-V_j}}$$

$$= \frac{1}{1 + \sum_{\substack{j=1 \\ j\neq i}}^{J} e^{-V_{ij}}} \qquad \text{(B-6)}$$

In the binary case, with alternatives A and B, say, we have

$$P(A) = \frac{1}{1+e^{-V}} \tag{B-7}$$

where we have dropped the subscript on V since, in the binary case, there is only one difference between distinct alternatives. It is the nonlinearity in the relationship between the probability of choice and the representative utility difference, as given in Equation B-6 (or B-7), that causes the aggregation bias. Thus, in our last example, had the residents of the inner and outer zones had different values for travel time differences that exactly offset the heterogeneity in cost differences, so that all the heterogeneity in V had been eliminated, there would have been no aggregation bias. The likelihood of such an event, of course, would be remote.

Lastly, it should be emphasized that aggregation bias arises only because of a desire to use a shortcut method to make aggregate predictions. In no way does aggregation

bias indicate an error in specifying or calibrating the model. Rather, it occurs because the model is nonlinear and thus the population probability of choice (i.e., the choice share) cannot be obtained by inserting into the prediction equation the population means of the independent variables.

The shortcut approach to obtaining the aggregate probability by replacing the independent variables with their average values will be referred to as the "direct aggregation" method, sometimes called the "naive method." The correct approach of computing the probabilities of each individual in the sample and taking the average of these probabilities to obtain the aggregate probability will be referred to as the "complete enumeration method."

Not only will aggregation bias usually lead to errors in estimating baseline choice shares, but it can also result in incorrect forecasts of policy effects. To continue with our current example, in which all workers have an auto time of 20 minutes, a bus time of 30 minutes, and an auto cost of $0.70, suppose the transit authority proposes to increase all fares by $0.05 (or $0.10 for a round trip). The effects of the bus fare increase as predicted by both the direct aggregation method and the enumeration method are given in Table B-3. The direct aggregation method predicts that bus ridership will fall by 43 percent of

Table B-3

AGGREGATION BIAS: MODAL SHARES BEFORE AND AFTER BUS FARE INCREASE AS PREDICTED BY TWO METHODS
(Hypothetical Data for Round Trip)

Model Variables	Before Bus Fare Increase	After Bus Fare Increase
Group 1 ("Outer Zone")		
Auto Time	20 minutes	20 minutes
Bus Time	30 minutes	30 minutes
Difference in Times (Auto minus Bus)	-10 minutes	-10 minutes
Auto Cost	$0.70	$0.70
Bus Cost (Fare)	1.00	1.10
Difference in Costs (Auto minus Bus)	-0.30	-0.40
Group 2 ("Inner Zone")		
Auto Time	20 minutes	20 minutes
Bus Time	30 minutes	30 minutes
Difference in Times (Auto minus Bus)	-10 minutes	-10 minutes
Auto Cost	$0.70	$0.70
Bus Cost (Fare)	0.50	0.60
Difference in Costs (Auto minus Bus)	-0.20	0.10

Table continued on following page.

Table B-3 (Continued)

AGGREGATION BIAS: MODAL SHARES BEFORE AND AFTER BUS FARE INCREASE AS PREDICTED BY TWO METHODS
(Hypothetical Data for Round Trip)

Modal Shares Predicted by Two Methods (Percent of Worktrips)

	Before Bus Fare Increase	After Bus Fare Increase
Direct Aggregation Method (Eq. B-4)		
Auto	92	96
Bus	8	4
Group 1 (Eq. B-4)		
Auto	98	99
Bus	2	1
Group 2 (Eq. B-4)		
Auto	73	83
Bus	27	17
Enumeration Method		
(1/2 Group 1 & 1/2 Group 2)		
Auto	86	91
Bus	14	9

present ("before") ridership (from 8 percent of all worktrips to 4 percent to worktrips) whereas the complete enumeration method predicts a decline of 38 percent (from 14 percent of worktrips to 9 percent).

Some general properties of aggregation bias have been shown by Frank S. Koppelman (B5) to hold under certain conditions:

1) For any pairwise comparison of alternatives, the aggregation bias will result in the share of the dominant mode being overpredicted. Thus, in our worktrip mode split example, since the average value of V (i.e., $\overline{V}$) is greater than zero, the direct aggregation method predicts a higher auto share (and lower transit share) than that predicted by the enumeration method. If $\overline{V}$ had been less than zero, so that transit had been the dominant mode, the naive method would have predicted a higher bus share (and lower auto share) than the enumeration method.

2) The aggregation bias will be greater, all other things equal, when the average representative utility difference falls in the more highly curved portions of the logit function, roughly the range from $\overline{V} = -2$ to $\overline{V} = -1$ and $\overline{V} = 1$ to $\overline{V} = 2$.

3) The aggregation bias increases, all other things equal, with an increase in the variability of V. Thus, in our worktrip mode split example, if the different bus fares

in the "inner" and "outer" zones had been further apart (say, $0.30 per round trip fare in the inner zone and $1.20 per round trip in the outer zone) the errors in the forecasts caused by aggregation bias would have been even larger.

Using "before" fares of $1.20 and $0.30 for the outer and inner zones, respectively, and fares of $1.30 and $0.40 as the "after" fares, the "before" and "after" modal shares calculated by the direct aggregation method are the same as in our original example since the average "before" and "after" fares are still $0.75 and $0.85, respectively. However, the "before" modal shares as predicted by the enumeration method are now 72 percent auto and 28 percent bus, and the "after" shares are 80 percent auto and 20 percent bus. Thus, the decline in ridership is 27 percent of the prefare increase ridership (versus a decline of 43 percent predicted by the direct aggregation method). As can be seen, the error due to aggregation bias has increased with increased variation in V.

4) The direct aggregation method is not a "consistent" estimator of the choice shares. That is, the aggregation bias will, in general, remain even if the sample size becomes infinitely large. Thus, the result in the example above is the same if the number of trips originating in each of the two zones is 100 or 1,000.

Having described the nature of the transferability problem and of aggregation bias, and having illustrated their potential magnitudes with very hypothetical examples, it should be noted that there is some evidence from Frank S. Koppelman (B6) that, for many practical applications, the magnitude of the aggregation error is not large when compared to other sources or error in model estimation and application (such as errors in data and model misspecification). This finding suggests that a simple technique for eliminating much of the aggregate bias would have much to offer. In Appendix C, we present such an approach.

REFERENCES

B1. U.S. Department of Transportation. "Guidelines for Trip Generation Analysis." Washington, D.C., Federal Highway Administration, June 1967.

B2. Charles River Associates. "Disaggregate Travel Demand Models, Project 8-13: Phase II Report." NCHRP, National Research Council, Cambridge, MA, May 1978.

B3. Talvitie, Antti and Kirshner, Daniel. "Specification, Transferability, and the Effect of Data Outliers in Modeling the Choice of Mode in Urban Travel." Transportation, Vol. 7, pp. 311-331, 1978.

B4. Atherton, T. J. and Ben-Akiva, M. E. "Transferability and Updating of Disaggregate Travel Demand Models." Transportation Research Record, No. 610, pp. 12-18, 1976.

B5. Koppelman, Frank S. "Travel Predictions with Models of Individual Choice Behavior." Ph.D. Dissertation, Department of Civil Engineering, Massachusetts Institute of Technology, May 1975.

B6. Koppelman, Frank. "Guidelines for Aggregate Travel Prediction Using Disaggregate Choice Models." Transportation Research Record, No. 610, 1976, pp. 19-24.

APPENDIX C

FINDINGS: GUIDELINES FOR USING THE MARKET SEGMENTATION TECHNIQUE WITH CENSUS DATA

PROCEDURES FOR CONSTRUCTING MARKET SEGMENTS WITH CENSUS DATA

In Appendix A, various proposed methods of aggregating logit travel demand models are compared. Based upon this analysis, it appears that the market segmentation approach is a promising one for developing standardized procedures to be used by local planners in applying logit models. In this appendix, general methods for constructing market segments are presented. The reader is referred also to Charles River Associates (C1) and Dunbar (C2).

The discussion in this appendix is designed to be general. Some readers may be confused by the sometimes complex notation required for a general discussion. Those readers should refer to the Phase II interim report (C3) for an example of market segmentation to forecast park-and-ride demand. Further analysis of the park-and-ride example is presented at the end of this appendix.

The market segmentation approach consists of constructing a cross-tabulation, where the variables defining the cells in the cross-tabulation are the independent variables (or proxies for them) in the model being applied. Each of the cells (or market segments) must specify that segment's share of all trips being modeled and that segment's average values of the independent variables in the model (for

all of the alternatives). The effect of a policy is evaluated in two steps. First, for each segment, the independent variables in the model are replaced with their average values to obtain the baseline choice shares for the segment. The aggregate baseline choice shares are obtained by computing a weighted average of these choice shares, where the weights are the segment shares. The second step is simply a repeat of the first step except that the independent variables are adjusted to reflect implementation of the policy. The effect of the policy is given by the difference between the aggregate choice shares calculated in the second step and those calculated in the first step. In this appendix we describe methods by which the data required for constructing market segments may be assembled.

In the situation where the policy induces only a small change in the choice shares, this procedure (of taking the difference of "before" and "after" choice shares) may not sufficiently eliminate the aggregation bias. In these situations, it is better to use the market segments to calculate an aggregate slope of the logit function with respect to the policy variable and to use this slope to calculate the (small) change in the aggregate choice shares. The aggregate slope is calculated in the same manner as the aggregate baseline choice shares, except that, in each segment, the choice shares are replaced by the slope of the logit function for that segment. The method by which logit slopes are calculated and used to predict small changes in choice shares will be described below.

The market segmentation scheme will employ the following definitions. A "homogeneous market segment" is a group of individuals with identical observed explanatory variables and parameters, i.e., possessing identical "representative utilities." A "market scheme" is a cell defined by a range of the explanatory variables chosen so that the variability of the representative utility (and thus aggregation error) is minimized (see Appendix A). The "cell frequency" is the probability that a random individual falls in a particular market segment. The "market share" is the predicted probability that a given alternative is chosen, contingent on the particular market segment.

Models to be Applied with the Market Segmentation Approach

It is, of course, impossible to describe a particular market segmentation scheme to be followed in evaluating travel demand policies without first specifying the dimensions of travel demand behavior that are the subject of analysis and the model to be used in performing the analysis.

There are five choice dimensions in urban travel demand which might, at one time or another, be of concern to local planners. These five dimensions are 1) the number of trips; 2) the time of day at which they occur; 3) their origin

and destination, 4) their mode, and 5) their route. The first, third, fourth, and fifth of these dimensions correspond approximately to the conventional urban transportation planning package trip generation, trip distribution, modal split, and trip assignment, respectively. The second dimension, that of the time of day of tripmaking, is obviously of concern for the design of peak capacity of facilities and thus should be added to the dimensions of travel demand in the conventional planning package. Of these five dimensions, it is anticipated that disaggregate models capable of predicting tripmaking responses with respect to the first four will be desired by local planners. While there are many urban transportation problems which have as their primary, if not sole, concern the selection of routes, the route choice problem is beyond the scope of the present analysis.

Furthermore, since the models most likely will be short-run models, in which the locations of residences, workplaces, and business establishments are taken as given, the trip distribution dimension of travel demand choices will be limited to the selection of a destination for discretionary trips (i.e., trips in which the tripmaker can make short-run choices from among alternative destinations, such as the choice of which store to patronize on a shopping trip). In particular, changes in land-use or other long-run factors affecting origin-destination patterns will be ignored.

Destination Choice Models. The problems associated with destination choice deserve further comment. A desirable feature of a method of applying logit models is that the method be equally appropriate for policy analysis at the level of the entire urban area or at any subarea (e.g., corridor) level. Unfortunately, this feature is not present in the market segmentation approach with respect to destination choice. In general it is impossible to specify the destination alternatives for all trips regardless of where they originate in the urban area. However, if the area level of analysis is confined to only a small portion of the urban area, such as a traffic zone or a corridor, then the analysis of destination choice becomes much more tractable. For this reason, the market segmentation procedures to be described will be primarily applicable to policies that affect only part of an urban area, although those procedures not related to destination choice can also be applied to the entire urban area.

Model Specification. Because the independent variables in the model are used as the classifying variables in constructing market segments, it is necessary to specify the model to be used before a market segmentation scheme can be outlined. The analysis will assume the general form the models are

likely to take and the important variables they probably will contain.

The primary issue to be resolved before the design of a market segmentation scheme can begin is the decision whether to use a joint or separable model for nonwork trips. A joint model gives the probability that an individual will choose a combination of travel demand alternatives as a function of the level-of-service attributes of the various combinations of alternatives and socioeconomic variables. For instance, a joint model for discretionary trips might give the probability that an individual will choose a particular mode/time of day/destination combination as a function of socioeconomic variables and the times and costs of traveling by various modes at various times of day to various destinations.

In a separable model, the alternatives entering any equation are limited to a single dimension of travel demand. This is done by forming a sequence of conditional equations. For example, one of the equations in a separable model might give the probability of choosing a particular mode conditional on a trip being taken and on the choice of a particular time of day and destination. The choice of mode would be a function of only the times and costs of taking the various modes at the given time of day to the given destination (plus socioeconomic variables). Another equation in the same separable model might explain the choice of a particular time of day conditional on a trip being taken and on the choice of a particular destination but not conditional on mode. In such an equation the time of day choice would be a function of the times and costs of traveling at the various times of day to the given destination by all of the available modes. Still another equation in the same separable model would give the probability of choosing a particular destination conditional on a trip being taken, but at any time of day by any of the available modes. The final equation in this separable model would give the probability of taking a trip to any destination, at any time of day and by any mode.

There has been considerable controversy over which of the two types of models, joint or separable, is preferable. It is not our intention to enter into the controversy here. On an empirical basis, both types appear to perform adequately. Both may be derived from the same behavioral assumptions. Thus, for the purpose of constructing market segments, selection between them hinges on the determination of which lends itself more readily to applications using market segmentation. On this basis, the separable approach apparently is preferable. For applications involving the full range of travel demand choice dimensions, the two

types of models would impose equivalent computational burdens on a market segmentation scheme. However, if the policy analyst is only concerned with one or two of the choice dimensions, then the separable model offers the possibility of substantially reducing the computations involved by reducing the number of market segments required to minimize heterogeneity in the independent variables.

For example, suppose an analyst is interested in making revenue projections for a proposed transit fare change. He or she is only interested in the effect on ridership of the fare change (and thus does not care about time of day or destination effects). Furthermore, the analyst believes the trip generation effects of the change will be trivial. With an appropriately structured separable model, an analyst could obtain the projection by constructing market segments for modal alternatives only. With a joint model, on the other hand, the analyst would still have to perform all of the computations required for an analysis of all of the choice dimensions and then sum across the irrelevant dimensions to get the modal effects. Thus, the separable approach offers the potential, in certain circumstances, of sizable savings in computational effort while being no less general in its range of possible applications.

Of course it should be recognized that a forecast based on a model which only considers one dimension of travel choice will not capture the "secondary" effects of a proposed policy change or system alternative. For instance, these secondary effects could be a change of destination in response to a fare change. Here, a fare change would not only stimulate a mode shift for some individuals, but also would affect the choice of destination. Forecasting the effect of the policy change with only mode choice conditional on destination (*P(M/D)*) would not reflect the change in destination. However, such secondary effects will be trivial in many applications. Therefore, the market segmentation scheme described here will be based upon the assumption that a separable model for nonwork travel will be used. In particular, it will be assumed that the separable model will follow the sequence: 1) mode given time of day, destination, and frequency; 2) time of day given destination and frequency; 3) destination given frequency; and 4) frequency.

For the sake of brevity, however, this appendix addresses only the segmentation scheme for mode choice. Readers interested in an example of a separable model of choice which includes destination, should see the Phase II interim report for Project 8-13 (C3).

The last issue to be resolved before outlining the market segmentation procedure is the specification of which variables are to be included in the models and which are to be used in defining the market segments. The

market segmentation scheme described here will be based on variables included in models which have already been calibrated. Any substantial modification to these models, or the use of entirely different models, can be expected to require corresponding modifications to the market segmentation scheme.

In the discussion that follows, all trips are expressed in terms of "round trips." Furthermore, variables are expressed in current monetary units. These must be inflated or discrepancies in the value of money over time must otherwise be accounted for when applying the models.

Market Segments for Mode Choice

Modal split equations for both work- and nonwork trips generally include, as level-of-service variables, cost, walk time, and measures of wait and linehaul time for each of the modal alternatives. The most likely candidate for socioeconomic variables in the mode split equations are household income and a measure of household automobile availability, e.g., the number of automobiles per licensed driver.

Assuming that the variables just listed are the variables included in the mode split equations, a 36-cell segmentation has been developed for the application of the mode split models. These 36 cells are defined by a four-way cross tabulation. The four classifying variables are trip distance (two categories), distance from home to public transportation (three categories), income (two categories), and auto availability (three categories, for a total of 2 X 3 X 2 X 3 = 36 cells).

Trip distance was chosen as a classifying variable because it is highly correlated with auto operating costs and linehaul times and is also correlated, although not so closely, with transit times and costs. The two categories of trip distance would be, not surprisingly, short trips and long trips, where any convenient measure of central tendency is used to separate the categories. The most convenient measure is probably average trip distance, although median trip distance probably is slightly more appropriate.

Distance from home to transit was selected as the proxy for walk time that is most likely to be available. Clearly, modes other than transit with walk access have walk times associated with them, and walking is likely to be required at the nonhome end of a transit trip. However, it is felt that walk times are far more important for transit trips than other modes and that data on the distribution of trips by distance from transit stop to workplace or shopping location would rarely, if ever, be available. A reasonable segmentation on distance from home to

to transit would be: 1) less than three blocks; 2) three to six blocks; and 3) more than six blocks. A block is assumed to be one twelfth of a mile.

Income and auto availability were selected as classifying variables because research to date indicates that they are significant in explaining mode choice behavior. Income would be divided into high-income and low-income categories. Although some measure of central tendency could be used to define the two income categories, the analyst may want to break the categories at a relatively low value of income in order to observe the impact of transportation policies of the poor.

The three categories of automobile availability are as follows:

1. Zero automobiles per driver, indicating a restricted choice set for mode;

2. Between zero and one automobiles per driver, indicating competition within the household for the use of automobiles; and

3. One (or more, although few households have more than one) automobile per driver, indicating no competition for the use of the automobiles.

The proposed market segmentation scheme is illustrated in Figure C-1. The market segmentation illustrated in Figure C-1 will be referred to as the mode split (market) segmentation. This segmentation is intended to provide reasonable detail necessary for most applications. Indeed, for many applications substantially less detail may be sufficient. Circumstances in which the full 36-cell segmentation is likely to provide greater accuracy than necessary are given below.

The entries required for each of the cells in the mode split market segmentation scheme given in Figure C-1 are the percent of all trips that fall within the cell and the cell's average value (for each mode) of the independent variables. If such data are available directly from a survey conducted for the analysis, clearly these data should be used. If not, estimates can be derived from other sources. At every step where data on local conditions are needed, we propose alternative reasonable values which may be assumed for forecasting purposes if necessary. We now describe a procedure by which data on cell frequencies may be obtained from data sources likely to be accessible to planners in all urban areas. The procedure consists of the following four steps:

1. Use Census data to determine the percent of households falling in each income and auto ownership market segment;

Figure C-1

PROPOSED MARKET SEGMENTATION FOR MODE SPLIT ANALYSIS

Less Than Three Blocks from
Home to Public Transportation

Automobile Availability
(Automobiles per Driver)

	Zero		Between Zero and One		One (or More)	
	Short Trips	Long Trips	Short Trips	Long Trips	Short Trips	Long Trips
Income Less Than $X						
Income More Than $X						

Three to Six Blocks from
Home to Public Transportation

Automobile Availability
(Automobiles per Driver)

	Zero		Between Zero and One		One (or More)	
	Short Trips	Long Trips	Short Trips	Long Trips	Short Trips	Long Trips
Income Less Than $X						
Income More Than $X						

Figure C-1 (Continued)

PROPOSED MARKET SEGMENTATION FOR MODE SPLIT ANALYSIS

More Than Six Blocks from
Home to Public Transportation

Automobile Availability
(Automobiles per Driver)

	Zero		Between Zero and One		One (or More)	
	Short Trips	Long Trips	Short Trips	Long Trips	Short Trips	Long Trips
Income Less Than $X						
Income More Than $X						

2. For each income group convert data on household auto ownership to data on autos/driver using the Nationwide Personal Transportation Study (NPTS) or equivalent local data;

3. Using NPTS and local data, determine percent of households in each category of transit access for each income and autos/driver market segment, and

4. Using NPTS or local data, determine the percent of trips (both short trips and long trips) for each transit access -- income -- autos/driver market segment.

More specifically, the steps are as follows.

1. Distribution of Households by Income and Number of Autos Owned

If an areawide policy is being evaluated, the joint distribution of households by income and auto ownership for the entire area can be obtained directly from Census data. In particular, Table H050 from the Sixth Housing Summary Tape of the 1970 Census of Population and Housing is a four-way cross-tabulation of number of households by 11 categories of income by four categories of number of automobiles available (zero, one, two, and three or more) by four categories of race of household head by two categories of housing tenure. (As the 1980 Census data became available, they would be used in place of 1970 Census data.)

Table H050 is available for each SMSA and for each county within an SMSA having a population of 50,000 or more persons. The selection of geographic area for which the table should be obtained depends on the locale of the policy action being contemplated.

A table equivalent to Table H050 is not published for Census tracts and must be estimated for the subarea. The Census Bureau has tabulated both the marginal (one-way) distribution of households by income category and the marginal distribution of households by number of autos owned for every Census tract. The marginal distribution of households by number of autos owned is published in the "Census Tracts" publications of the Census Bureau (C4). This marginal distribution also is available from Table H17 of the Fourth Count Housing Summary Tapes.

The marginal distribution of households by income is available from Table H117 of the Fourth Count Tapes. The "Census Tracts" volumes contain an income distribution of families, not households, which can serve as a proxy if the Fourth Count Tapes are not available. The biggest discrepancy between the income distribution of families and that of households is that the latter includes data on single person households (and households containing only unrelated individuals), while the former does not. Income data on unrelated individuals, combined with data on the

number of single person households, may be used to reduce the discrepancy.

The marginal income and auto ownership distributions are used as control totals to adjust the cell frequencies in an income by auto ownership cross-tabulation. The adjusted cell frequencies serve as the estimates of the joint income-auto ownership distribution. The cell frequencies from Table H050 or equivalent data on a similar population can be used to account for the fact that auto ownership and income are dependent. The adjustment process assumes that the dependence between auto ownership and income for the Census tract is similar to that of the larger area.

The adjustment procedure suggested by Johnson (C5) is as follows. We start with the situation illustrated in Table C-1, which represents an overlay of two tables. One table, the combination of Tables H17 and H117, does not have the percent of households for the specific combinations of income and automobile availability. The other table, Table H050, contains the missing entries but these entries are for a geographic area different from the policy application area. (All cell frequencies have been converted to percents.) In general, we would not expect the X^R's in any one of the row total cells (or X^C's in any of the column total cells) in Table B-2 to be equal to the Y (or Z)

Table C-1

OUTPUT OF FIRST THREE STEPS IN PROCEDURE FOR CONSTRUCTING MARKET SEGMENTS FOR MODE SPLIT ANALYSIS

(Entries are Percent of Households)

Income	Number of Automobiles Available				
	Zero	One	Two	Three or more	Total[1]
Less than \$2,000	X	X	X	X	X^R,Y
\$2,000 to \$3,999	X	X	X	X	X^R,Y
\$4,000 to \$5,999	X	X	X	X	X^R,Y
\$6,000 to \$9,999	X	X	X	X	X^R,Y
\$10,000 to \$14,999	X	X	X	X	X^R,Y
\$15,000 to \$24,999	X	X	X	X	X^R,Y
\$25,000 or more	X	X	X	X	X^R,Y
Total	X^C,Z	X^C,Z	X^C,Z	X^C,Z	100.00

[1]The "X's" indicate entries available from Table H050 or some other source, the "Y's" indicate entries available from Table H117, and the "Z's" indicate entries available from Table H17. The superscript "R" on the "X's" indicates a row total and the superscript "C" indicates a column total.

in the same cell since these figures are for different geographic areas. However, it is the Y's (or Z's) that are the correct figures for the policy application area. In order to get the X's in the body of the table to sum to the Y row totals and Z column totals, one first multiplies each of these X's by the quotient Y/X^R for the row in which the X is contained. Let us label the result of this multiplication X_1 (i.e., $X_1 = (Y/X^R) \cdot X$). Although the sum of the X_1's in any row now will be equal to Y for that row, we still would not expect the sum of the X_1's in any column to be equal to the Z for that column. Thus, the X_1's next are multiplied by $Z/\sum_c X_1$, where $\sum_c$ denotes the sum of the cells in a column. Let X_2 be the product: $X_2 = (Z/\sum_c X_1) \cdot X_1$. Now, for the X_2's, the column sums agree but the row sums are off again, although not as far as they were initially. This process of multiplying the entries in each row by the quotient of the desired total for that row divided by the actual sum of that row, then using the resulting entries to do the same for each column, then using these results for each row again, will eventually lead to the sum of the entries converging to the desired row <u>and</u> column totals (within any specified tolerance range). In most cases, about six iterations should be sufficient. (An iteration is one series of multiplications to bring row sums into line followed by a series of multiplications to bring column sums into line.)

Simple BASIC computer programs to perform these calculations can be and have been written to facilitate computation.

2. Distribution of Households by Income and Autos/Driver

The preceding step results in a distribution of households by income and number of autos, not autos per driver. If a three-way cross-tabluation giving the count of households for each income-auto ownership-number of drivers category were available, the derivation of the income-autos per driver distribution would be straightforward. The percent of low (high) income-zero autos per driver households is simply the sum of the three-way cell frequencies for all low (high) income-zero autos cells. The percent of households with low (high) income and between zero and one autos per driver is the sum of the three-way cell frequencies for low (high) income households with one auto and two or more drivers, etc. Likewise, the percent of households in the low (high) income-one auto per driver segment is the sum of three-way cell frequencies for low (high) income households with one auto and one driver, two autos and two drivers, etc.

Unfortunately, three-way cross-tabulations giving cell frequencies of households for income by auto ownership by number of drivers categories are not often available for a particular subarea and the cell frequencies must be estimated. The cell frequencies are estimated by multiplying the joint distribution of households by income and auto ownership (the output of Step 1 above) by a conditional distribution of number of drivers,

given income and auto ownership, obtained for a different, but similar, population. Data from the Nationwide Personal Transportation Study (NPTS) (C6) may be used for this purpose. The NPTS was conducted by the Census Bureau for the Federal Highway Administration in 1969. This survey sampled 6,000 households across the country and obtained data on: 1) the number of trips taken on the day prior to the survey day; 2) the purpose, mode, travel time, and distance of each trip taken; 3) the distance from home to public transportation; and 4) household socioeconomic characteristics including income, number of automobiles available, and number of licensed drivers in the household.

The required conditional distribution is computed from a three-way (income by auto ownership by number of drivers) count of households from the NPTS data. This three-way cross-tabluation can be restricted to those living in an urban area of a particular size, living in a particular region of the country, or living in the central city (or suburbs) of the urban area. Thus, data for a similar population are used. When the estimated cell frequencies for the three-way table for the application area are obtained, the computation of the income-autos per driver segment shares proceeds as described above.

3. Distribution of Households by Transit Access, Income, and Autos per Driver

The marginal distribution of households by distance to public transportation is very city-specific, i.e., cities differ in the layout of their transit routes and in the spatial distribution of the population. Even within a single urban area, the different parts of the area vary in the accessibility of transit. Thus, the best approach to computing the percent of households living within each access to transit category is by reference to transit maps or an actual household survey. The contour lines for each transit access category around the transit routes are drawn on the maps. If the spatial distribution of the population is approximately uniform, then the percent of population in each transit access segment is simply the area between the relevant contour lines. If the spatial distribution of the population deviates severely from uniformity, then block population statistics, available from U.S. Census Publications (C7), may be superimposed on the map and the population in each transit access segment enumerated.

The joint distribution of transit access with the other segmentation variables could be derived by assuming independence. However, an improved approach would be the combination of local data on the marginal distribution of transit access with national data on the conditional distribution of households by access to transit, the joint transit access-income-autos per driver distribution may be estimated by multiplying the marginal transit access distribution by a conditional distribution of the percent of

households in each transit access category, given income and autos per driver. The conditional distribution should be for a similar population and can be obtained, for example, from the NTPS (C6) data by constructing a three-way cross-tabulation of the percent of households by transit access, income, and autos per driver.

4. Distribution of Trips by Trip Distance, Transit Access, Income, and Autos per Driver

The completion of Step 3 above results in a probability for a household falling in a given transit access, income, and autos per driver cell. The household distribution is converted to a trip distribution by use of trip generation rates in each cell. We multiply the percent of households in each cell by a pair of households to trips scalars (one for long trips, one for short trips) that are unique for each segment. Local data on the expected number of short and long trips per household for the appropriate purpose are desirable. Rates can be estimated from the NPTS data by counting the number of short work trips, for example, in each market segment (a cell in the three-way cross-tabulation using transit access, income, and autos per driver as the classifying variables) and dividing this number of trips by the number of households, and likewise for long work trips. When this has been done for all market segments, the result will be the required joint distribution of trips by trip distance, transit access, income, and autos per driver.

<u>Determination of Values of Explanatory Variables for Market Segments</u>

In the above four steps, a procedure for computing the percent of trips in each market segment was presented. Application of the market segmentation technique also requires data on the average values of the independent variables for each segment. The average income for all low income households can be computed from Census data. The problem is that we need the average income of tripmakers, not households. If necessary, the household income of low income families can be assumed to be the same as the trip weighted average income of tripmakers from low income families. Furthermore, the trip weighted average income for trips by low income persons can be assumed to be invariant across the different categories of transit access, autos per driver, and trip distance. Clearly, neither of these assumptions is strictly true, but the error they introduce is unlikely to justify the effort required to relax them.

Similar to the assumptions made about average income, the average value for transit access distance can be assumed to vary only across transit access categories -- not across income, trip distance, and autos per driver categories for a given transit access category -- and to be

the same on a trip-weighted basis as on a household-weighted basis. The average on a household-weighted basis can be computed from the contour lines described in Step 3 above.

There is probably little error in assuming the average values of autos per driver to be 0, 0.5, and 1 for the zero, between zero and one, and one autos per driver categories respectively. However, more precise estimates may be obtained by calculating the short and long trip-weighted average within each cell of the three-way (transit access, income, and autos per driver) cross-tabluation of NPTS data.

Methods for obtaining average trip distance depend on the policy being evaluated. If it is a policy that affects trips being taken to all destinations in the urban area, then average trip distances for trips of the appropriate purpose in cities of the appropriate size can be obtained from NPTS data. Again, it is of questionable value to obtain separate averages for the different categories of income, transit access, and autos per dirver. If the policy only affects some trips, then average trip distances must be computed on the basis of local knowledge about the trip patterns of the affected trips.

The averages just described are used to estimate the average values of the variables in the mode choice model (cost, walk time, wait plus linehaul time, income, and autos per driver) as follows. The averages of income and automobiles per driver are direct inputs to the model. The average walk time for the transit with walk access mode can be calculated by multiplying the average distance from home to public transportation (which is given in blocks) times 1/12 (twelve blocks to the mile) and multiplying this result times an assumed walking speed of 19 minutes per mile. This result is then doubled to give the walking time of the combined outbound and return trip. Average operating cost for the auto modes can be calculated by multiplying average trip distance (given in miles) by $0.035, which was calculated to be the national average of automobile operating costs in 1969 (see U.S. Census (C8)). Doubling this result gives the auto cost of a round trip. Local planners probably can provide precise data for their urban areas for round trip transit fare. For the auto modes, auto travel times depend on roadway trip distances, and trip distances are available in the local data. Transit travel times are not nearly so closely correlated with trip distance, however. It would probably be best, therefore, if data on average transit times were specified by the local planners, although these data could be related to trip distance if necessary. If transit travel times are obtained by relating the transit travel times reported in the NPTS (C6) data to the trip distances, then the estimated walk times must be subtracted from the reported travel times before the relationship is computed.

PROCEDURES FOR EMPLOYING THE MARKET SEGMENTATION APPROACH IN THE APPLICATION OF LOGIT TRAVEL DEMAND MODELS

The preceding section of this paper described the construction of market segments that can be used in making aggregate predictions of policy effects with disaggregate travel demand models. In this section, the procedures are given by which these market segments are employed in policy analysis.

Reducing the Number of Mode Split Market Segments

The first step in the application of logit travel demand models using the market segmentation described above is the determination of whether all of the 36 mode split market segments (as outlined in Figure C-2) are necessary. There are at least two conditions which, if present, would indicate that a less detailed segmentation would be sufficient. They are as follows:

1. Some tripmakers are "captive" to one of the modes. An example would be an application where the only two modes being modeled are auto (drive alone) and a single transit mode. In this situation, individuals without automobiles would be captive to the transit mode and, therefore, would be predicted to choose transit with a probability of one. Since the logit model assumes that individuals have a choice, it is superfluous to apply the model to these tripmakers. These "transit captive" tripmakers correspond to

the zero automobile per driver cells in Figure C-1, of which there are 12. Thus, the number of segments to which the model is to be applied can be reduced from 36 to 24 by combining these 12 "transit captive" cells into a single cell, the 25th, to which the model is not applied but rather a unit probability of taking transit is assumed. Similarly, if the only transit mode being modeled is transit with walk access, then individuals living more than six blocks from public transportation might be assumed to be auto captives, allowing for the combination of all the "more than six blocks from public transportation" cells into a single cell where auto is assumed to be the only mode chosen.

2. All tripmakers are homogenous with respect to one of the classifying variables. This condition is a very obvious one. The purpose of market segmentation is to eliminate heterogeneity in the independent variables. Clearly, if there is no heterogeneity, then segmentation is unnecessary. An example would be applications to areas with a very high incidence of poverty. In this situation, the segmentation on income is unnecessary. Before an application is made, the tripmakers to which the model is being applied should be analyzed with respect to each of the classifying variables in the mode split segmentation, even if only cursorily, to determine if there is so little

variation in one of the variables that segmentation with respect to that variable is unnecessary.

Reconciling the Modes Being Analyzed with Those Used in Model Calibration

It cannot be expected that the modes among which the planner wishes to analyze policy-induced shifts are exactly the same as those used to calibrate the (worktrip and non-worktrip) mode split models. For brevity, we will refer to the modes for which the planner wishes to analyze policy-induced modal shifts as the "prediction modes" and will refer to the modes used to calibrate the model as the "calibration modes." There may be some calibration modes in which the local planners are not interested, or the planner may be interested in special prediction modes that were not included as calibration modes. Mismatches of either variety pose problems in the use and interpretation of socioeconomic variables and mode-specific constants. In model calibration it is probably preferable, for statistical reasons, to include a mode specific constant for all but one of the calibration modes. Model application is easier, however, if the modes are grouped into, say, an auto-oriented group and a transit group. All the (calibration) modes within the auto-oriented group would be specified to have identical representative utility functions. In addition to the level-of-service variables, this identical representative utility function would contain a single mode-specific constant (common to all auto-oriented modes) and the socioeconomic variables (e.g., income and automobile availability). Likewise, all of the (calibration) modes within the transit group would be specified to have identical representative utilities, but their representative utility function would only contain the level-of-service variables.

If the models being applied have grouped modes according to type (auto versus transit), then the procedure required for the reconciliation of prediction modes with calibration modes is simply the classification of the various prediction modes as auto-oriented or transit. If, on the other hand, the models being applied have included a mode specific constant in the representative utility of all (but one) of the modes, with some corresponding arrangement of socioeconomic variables, then reconcilation of prediction modes with calibration modes requires the matching of each of the prediction modes with the calibration mode most similar to it with respect to unobserved factors. Such a matching process would be largely intuitive in many cases. In any event, before the mode split application of the models can proceed, all of the prediction modes must be assigned a calibration mode whose representative utility it will be assumed to share. Clearly, the forecast is sounder if there is a calibration mode corresponding to each prediction mode.

Transferring the Logit Model

For reasons given in Chapter Three, the particular updating procedure that appears to provide the best combination of accuracy and ease of implementation is to adjust the mode-specific constant(s) and scale the other independent variables. We first describe this procedure for the case where there are only two modal alternatives (and thus just a single mode-specific constant to be adjusted). Following this is a description of the extension of the procedures when more then two modes are being analyzed. Either case requires knowledge of the actual modal split in each market segment.

If a choice between just two modes (auto and bus, say) is being modeled, then the adjusted mode-specific constant and the coefficient scalar can be calculated by running a weighted linear regression using any standard regression package. The observations for this regression are the market segments constructed for the mode split analysis. The dependent variables in the regression are the quantities $\ln(SH_{at}/SH_{bt})$, where SH_{at} is the auto share in the tth market segment and SH_{bt} is the bus share for that segment. The independent variables are a constant and the quantities $V_t=\underline{\alpha}'\underline{X}_t$ where $\underline{\alpha}$ is the vector of (original) model coefficients (excluding the mode specific constant) and $\underline{X}_t$ is the vector of the average values of the differences between auto and transit in the independent variables for the tth market segment. The observations in the regression are weighted by the segment shares (which means that the dependent and independent variables, including the constant, are multiplied by the square root of the segment share before running the regression). If the standard regression package automatically adds a constant, this feature must be suppressed before the regression is run. The constant estimated by this regression will be the constant of the updated model; the coefficient of V will be the scalar to be multiplied by the coefficients to obtain the updated coefficients of these other variables.

The procedure to be followed when there are more than two modes (modes m_1, m_2, ... m_M, say) is similar to the two-mode case. One of the modes, e.g., auto drive alone, is arbitrarily specified as a base mode. There are M-1 observations for each market segment. For each observation, the dependent variable is $\ln(SH_{jt}/SH_{At})$ for $j \neq A$, where A denotes the base mode. The independent variables are $V_{jt} = \underline{\alpha}'\underline{X}_{jt}$ where α is the vector of original model coefficients excluding the constant(s), and $\underline{X}_{jt}$ is the vector of the average values of the differences between modes j and A on the independent variables for the tth segment. In addition, a constant corresponding to mode j is included

as an independent variable if the original model contained such a constant. The coefficient of V_{jt} is the scale factor and the updated constants are those estimated by this regression.

Forecasting Modal Shares

After completion of the model transferring procedure, all the ingredients necessary for forecasting the effects of transportation policies on modal shares are present. Baseline modal shares are computed for each market segment, by evaluating the logit equation for each mode using the prepolicy averages of the level-of-service variables. The modal shares thus predicted for each market segment are then aggregated by weighting them by the segment shares and summing the weighted modal shares across market segments. Next, the policy variable is changed to reflect implementation of the policy, and the predicted aggregate modal shares are recomputed. The effect of the policy is, of course, the difference between the pre- and post-policy modal shares.

If accuracy in the predicted change in the share of one of the modes is critical and if the change as predicted by this method is less than one tenth of the original share of that mode, then it is advisable to re-estimate these changes using the slope of the logit function. This is done by computing, for each market segment, the slope of logit function with respect to a policy variable, x_m as:

$$SLOPE_t = \frac{\partial P(m)}{\partial x_m} = \alpha[P_t(m)][1-P_t(m)] \tag{C-1}$$

or

$$SLOPE_t = \frac{\partial P(n)}{\partial x_m} = \alpha[P_t(m)][P_t(n)] \tag{C-2}$$

where α is the (updated) coefficient of the policy variable, and $P_t(m)$ and $P_t(n)$ are the predicted prepolicy shares of modes m and n in market segment t. Equation C-1 is used to predict direct effects, e.g., the effect of a transit fare change on the transit's share. Equation C-2 is used to predict cross effects, e.g., the effect of the transit fare increase on auto's share. Once the slopes for all the modes have been computed for each market segment, the aggregate slope for each mode may be derived in the same manner that aggregate modal shares are calculated from the modal shares of the individual segments. The aggregate slopes are then multiplied by the change in the policy variable to obtain the effect of the policy:

$$dP(n) = \frac{\partial P(n)}{\partial x_m} \cdot dx_m \tag{C-3}$$

where $dP(n)$ is the change in the aggregate share of mode n, dx_m is the change in the policy variable, and $\frac{\partial P(n)}{\partial x_m}$ is the aggregate slope of the logit function with respect to x_m.

NEW FINDINGS ON THE USE OF A DISAGGREGATE DEMAND MODEL TO FORECAST DEMAND FOR PARK-AND-RIDE

In the report for Phase II of this project (C3), a disaggregate worktrip model was applied to estimate modal shares

following the introduction of a new park-and-ride facility in Towson, Maryland. The performance of the model was apparently somewhat discouraging. In particular, when the estimated shares were compared to actual shares, park-and-ride demand was substantially underpredicted and auto demand was overpredicted.

Subsequent analysis of this park-and-ride application indicates that the performance of the disaggregate worktrip model may not be as discouraging as previously believed. Two conclusions from the earlier analysis were that: 1) the model substantially underpredicted the demand for the new park-and-ride service; and 2) the market segmentation techniques did not perform much better than the direct aggregation approach for that specific application.

The conclusion on the performance of the model was based upon the assumption that the new park-and-ride lot improved the existing round-trip in-vehicle time (including wait time) of the bus with auto access mode by 13.4 minutes because of the new express bus service. However, no other changes in level-of-service variables were assumed. In particular, no change in walking time for the bus with auto access mode at the new park-and-ride lot was assumed. The conclusion on the lack of difference between the results of the market segmentation and direct aggregation methods arose from an error in calculating the direct aggregation results.

The use of alternative level-of-service assumptions and the correct calculations for the direct aggregation method lead to substantially different conclusions. If it is assumed that the park-and-ride lot reduces walking time for the bus with auto access mode (park-and-ride), as well as the linehaul time, the forecast is much improved. The following discussion presents a reapplication of the park-and-ride example with the above adjustments. The original calculations can be found in the Phase II Report (C3).

Equation C-4 is the multinomial logit model used in the example.

$$P(AUTO) = \frac{e^{V_A}}{D}, \quad V_A = -13.20 - 0.06426\ IVT - 0.2834\ WT \qquad \text{(C-4)}$$

$$- 1.480\ COST + 5.644\ A/D + .3470\ INC.$$

$$P(BUS) = \frac{e^{V_B}}{D}, \quad V_B = -0.06426\ IVT - 0.2834\ WT - 1.480\ COST$$

$$P(Park'n\ Ride) = \frac{e^{V_{PR}}}{D}, \quad V_{PR} = -13.64 - 0.06426\ IVT$$

$$- 0.2834\ WT - 1.480\ COST + 5.644\ A/D + .3470\ INC.$$

$$D = e^{V_A} + e^{V_B} + e^{V_{PR}}$$

where:

- *IVT* is a combined measure of in-vehicle and wait times;
- *WT* is walk times in minutes with walk time assumed to be zero for the auto mode;
- *COST* is the cost of making a trip, measured in 1967 dollars;
- *A/D* is the number of automobiles per licensed driver in the household; and
- *INC* is family income in thousands of dollars, decoded as the midpoints of the ranges in a categorical income variable.

Table C-2 presents the relative sizes of the 12 market segments and Table C-3 gives the values of the independent variables for each of the 12 segments before the implementation of the park-and-ride lot and the express bus service. These tables are from the NCHRP Phase II report (C3). Details on their construction are reported there.

After the implementation of the new service, it is assumed that park-and-ride in-vehicle time (IVT) is reduced to 75 minutes and walking time (WT) is reduced to 6.8 minutes. A walking time reduction of this magnitude resulting from a new parking lot is a plausible alternative assumption to the original assumption of no advantage.

The market segmentation aggregation technique, which is described earlier, is then applied, using the data in Tables C-2 and C-3. These resulting aggregate modal shares are multiplied by the 781 worktrips in this example to produce modal ridership figures. In addition, modal shares from the direct aggregation method are estimated.

The results of the analyses are presented in Table C-4. Two conclusions are apparent. First, the performance of the model is substantially improved relative to the original assumptions. An implication of this result is that forecasts, especially of alternatives with small shares, can be very sensitive to the measurement of level-of-service changes. Second, the market segmentation approach clearly performs better than the direct aggregation method. In particular, the latter method assigns no trips to the bus with walk access mode. This occurs because of the averaging

Table C-2

MARKET SEGMENT SHARES FOR BALTIMORE PARK 'N RIDE APPLICATION

(Percent of All Worktrips)

Less than three blocks from home to public transportation:

	Between Zero and One Auto Per Driver	One Auto Per Driver
Income less than $15,000	0.010	0.025
Income more than $15,000	0.009	0.014

Between three and six blocks from home to public transportation:

	Between Zero and One Auto Per Driver	One Auto Per Driver
Income less than $15,000	0.011	0.015
Income more than $15,000	0.006	0.009

More than six blocks from home to public transportation:

	Between Zero and One Auto Per Driver	One Auto Per Driver
Income less than $15,000	0.190	0.248
Income more than $15,000	0.170	0.293

SOURCE: Charles River Associates (C3), Table 6.

Table C-3

VALUES OF INDEPENDENT VARIABLES
(Round Trips)

Market Segments			Independent Variables (Before No. 26 Bus Service)				
Distance to Transit	Income (Dollars)	Autos Per Driver	IVT (Auto) (Minutes)	COST (Auto) (Dollars)	Autos Per Driver	INC (Thousands)	IVT[1] (Park 'n Ride) (Minutes)
Less Than 3 Blocks	<15,000	0-1	58.0	3.22	0.50	14.7	88.4
		1+	58.0	3.22	1.02	14.7	88.4
	>15,000	0-1	58.0	3.22	0.56	42.6	88.4
		1+	58.0	3.22	1.08	42.6	88.4
3-6 Blocks	<15,000	0-1	58.0	3.22	0.48	14.7	88.4
		1+	58.0	3.22	1.12	14.7	88.4
	>15,000	0-1	58.0	3.22	0.57	42.6	88.4
		1+	58.0	3.22	1.10	42.6	88.4
Greater Than 6 Blocks	<15,000	0-1	58.0	3.22	0.50	14.7	88.4
		1+	58.0	3.22	1.02	14.7	88.4
	>15,000	0-1	58.0	3.22	0.56	42.6	88.4
		1+	58.0	3.22	1.07	42.6	88.4
Average for direct method:			58.0	3.22	0.8427	28.68	88.4

Market Segments			Independent Variables (Before No. 26 Bus Service)				
Distance to Transit	Income (Dollars)	Autos Per Driver	COST (Park 'n Ride) (Dollars)	WT[2] (Park 'n Ride) (Minutes)	IVT[1] (Bus) (Minutes)	COST (Bus) (Dollars)	WT (Bus) (Minutes)
Less Than 3 Blocks	<15,000	0-1	1.80	8.6	88.4	1.50	13.4
		1+	1.80	8.6	88.4	1.50	13.4
	>15,000	0-1	1.80	8.6	88.4	1.50	13.4
		1+	1.80	8.6	88.4	1.50	13.4
3-6 Blocks	<15,000	0-1	1.80	8.6	88.4	1.50	22.9
		1+	1.80	8.6	88.4	1.50	22.9
	>15,000	0-1	1.80	8.6	88.4	1.50	22.9
		1+	1.80	8.6	88.4	1.50	22.9
Greater than 6 Blocks	<15,000	0-1	1.80	8.6	88.4	1.50	60.8
		1+	1.80	8.6	88.4	1.50	60.8
	>15,000	0-1	1.80	8.6	88.4	1.50	60.8
		1+	1.80	8.6	88.4	1.50	60.8
Average for Direct Method			1.80	8.6	88.4	1.50	56.5

[1]After the implementation of park and ride service, the variable is reduced to 75 minutes.

[2]WT = Walking Time. After the implementation of park and ride service, the variable is reduced to 6.6 minutes.

SOURCE: Charles River Associates (C3), Table 12.

Table C-4

PERFORMANCE OF WORK MODE CHOICE MODEL UNDER ALTERNATIVE LEVEL-OF-SERVICE ASSUMPTIONS

Mode	Actual Number of Trips[1]	Predicted Number of Trips: Linehaul Time Improvement Only[2]			Predicted Number of Trips: Linehaul and Walk Time Improvement[3]	
		Market Segment[1]	Corrected Direct[4]	Original Direct[1]	Market Segment[4]	Direct[4]
Auto	564	662	676.5	659	602.5	613.9
Bus with Auto Access (Park-and-Ride)	200	102	104.5	106	164	167.1
Bus with Walk Access	17	17	0	16	14.5	0
Total	781	781	781	781	781	781

[1]From Table 13 of Charles River Associates (C3).

[2]It is assumed that the express bus service reduces linehaul time by 13.4 minutes round trip.

[3]In addition to the linehaul improvements, the park-and-ride lot is assumed to reduce walk time by 2 minutes.

[4]Calculated with the data in Tables C-2 and C-3.

SOURCE: Charles River Associates, 1980.

of the walk time variable. The overall average value of 56.50 minutes results in a miniscule modal share prediction, even though walking time is quite favorable in particular segments.

Therefore, we conclude that the model appears to be transferable from Pittsburgh, with appropriate adjustments, and that with careful analysis of the assumptions made about level-of-service changes, the market segmentation is an acceptable and accurate approach to forecasting with disaggregate demand models. The method is sensitive to assumptions about the predicted value of level-of-service variables; however, if these changes are predicted carefully, the market segmentation method should forecast the impact of change more accurately than does the direct aggregation method.

REFERENCES

C1. Charles River Associates. "Estimating the Effects of Urban Travel Policies." Prepared for the Office of the Secretary, U.S. Dept. of Transportation, 1976.

C2. Dunbar, Frederick C. "Policy Contingent Travel Forecasting with Market Segmentation." Transportation Research Record, No. 637, 1977, pp. 27-32.

C3. Charles River Associates. Disaggregate Travel Demand Models, Project 8-13: Phase II Report. Prepared for National Cooperative Highway Research Program, National Research Council. Cambridge, MA, May 1978.

C4. U.S. Bureau of the Census. "Census Tracts." 1970 Census of Population and Housing. Series PHC(1). Washington, D.C., 1972.

C5. Johnson, Richard M. "Marginal Weighting." Market Facts Inc., October 1972.

C6. Federal Highway Administration. Nationwide Personal Transportation Study. Washington, D.C., U.S. Department of Transportation, 1972-1974.

C7. U.S. Bureau of the Census. "Block Statistics." 1970 Census of Housing. Series HC(3). Washington, D.C., 1972.

C8. U.S. Bureau of the Census. 1971 Statistical Abstract of the United States. Washington, D.C., 1971, Table 854.

APPENDIX D

EMPIRICAL AND CONCEPTUAL MODEL DEVELOPMENT

INTRODUCTION

The results in this appendix are addressed to the research commmunity and are therefore highly technical. Readers not interested in the details of the model development are advised to limit their reading to Chapter 2 of this report.

This appendix provides a selected description of the empirical and conceptual disaggregate demand model development research carried out during this project. Using disaggregate data from three cities, several models have been estimated. Mode split models for work and shopping trips using Pittsburgh data were reported in Chapter 2. A more detailed report of the research results may be found in Appendix C of the Phase I report. A substantial amount of the research on conceptual model development in that report is not reported here.

Separate discussions of the following issues are presented to amplify the material presented in Chapter Two:

- Generic versus mode-specific level-of-service variables;
- Income segmentation;
- Automobile availability effects;
- Transformations of level-of-service variables;
- Empirical choice set formation; and
- Lessons on disaggregate data set collection.

The final three sections of this appendix are specific to each of the three data sets used in this project. Modeling efforts with each data set are described separately.

1. THE DISAGGREGATE DATA SETS USED IN THE EMPIRICAL RESEARCH

Introduction

Three data sets were used for the empirical research in this project. Data originally compiled from the 1967 Pittsburgh household interview survey and prepared for disaggregate analysis in line with an earlier CRA study (D1) were suitable for estimation of mode, destination, and frequency choice models.

A second data set, processed during this study from the 1970 Twin Cities household interview survey, was suitable for logit mode split model estimation. Preparation of these data sets for disaggregate model estimation required augmenting the household-level socioeconomic and trip record information, contained in home interview survey files, with network level-of-service data; in the case of the Pittsburgh data, this entailed considerable modification and improvement to existing transportation supply data.

The third data set was collected in Baltimore in 1977 for the expressed purpose of disaggregate travel modeling. This data set contains home interview survey data on vehicles, persons, and travelers for nearly 1,000 households. For each individual the data set contains a one-day travel diary and extremely detailed data, including alternatives, on a selected trip from each household.

Each of these data bases is described below.

The Pittsburgh Data Base

The data used for model estimation were derived from a household interview survey (HIS) conducted by the Southwestern Pennsylvania Regional Planning Commission (SRPC) in 1967 and from transportation network analysis data compiled by SRPC. The travel information in the HIS documents household questionnaire responses detailing travel activity of all household members over five years of age. In constructing the data for estimation, a random sample of households was drawn from two major travel corridors in the Pittsburgh area; one east of the Central Business District (CBD) extending to the city limits and the second to the suburbs in the south. Both corridors had relatively good transit service from suburban locations to the CBD.

Trips selected for the estimation sample had to meet two criteria: 1) the household location had to be located within the defined corridors; and 2) the terminus of a reported trip also had to be within the defined corridors.

A total of 115 trip records (observations) were included in the sample used for worktrip mode split. The shopping trip

estimation sample contained 169 observations, of which 143 households reported one shopping trip and 26 observations represented 0 (shopping) trip households.

Table D-1 summarizes the distribution of mode and purpose travel patterns in the Pittsburgh data used for model estimation.

The Twin Cities Data Base

The disaggregate data used for estimating the worktrip mode split models were constructed from the Established Person-Trip Analysis File (EPTAF) available through the Twin Cities Metropolitan Council (D2). The EPTAF consists of home interview survey data, augmented with highway and transit network level-of-service (LOS) data. The home interview survey (HIS) was conducted in 1970. In all, 5,700 households were surveyed (a 1 percent sample), yielding a total of 45,714 one-way, linked person-trips.

In order to prepare the data for disaggregate mode split model estimation, several processing steps were undertaken.

1. A subset of the trips for which transit was an available alternative was singled out for further analysis. This screening procedure effectively limited the sample to only those travel corridors within the Twin Cities region where transit service was provided. (Of the total 45,714 trips in the HIS, 3,640 trips were considered to have transit

Table D-1
MODE AND PURPOSE DISTRIBUTION IN
THE PITTSBURGH ESTIMATION SAMPLE

Worktrip Sample

Mode	Number of Trips	Percent of Total Trips
Auto	62	54
Bus	53	46
Total	115	100

Shopping-Trip Sample

Mode	Number of Trips	Percent of Total Trips
Auto	78	56
Bus	65	44
Total	143	100

Nontravel Households: 26

Total in Sample: 169

SOURCE: Charles River Associates, 1976.

available. Less than 10 percent of the trips for which transit was available were actually made by transit. This "choice file" was constructed by the Metropolitan Council, based on traveler responses to the question "Was transit available for this trip?".)

2. Several "screening" criteria were employed to eliminate trip records with inconsistent or missing data. For example, households who failed to report income and trip records where LOS data were missing or incomplete were eliminated from the sample. After screening, 3,006 one-way trips remained in the sample.

3. Complex tour trips were eliminated from further processing. One-way trips were combined into round trips. LOS data for outbound and inbound trips were added to represent round trip characteristics. A total of 721 round trips were left after this final processing step.

Table D-2 summarizes the distribution of mode and purpose travel patterns in the final estimation sample. Although travel choices were recorded for auto driver, auto passenger, and public bus modes, all estimations were performed with binary choice (auto driver/bus) models. It is clear from Table D-2 that for all the nonworktrip purposes, there were an

Table D-2

MODE AND PURPOSE DISTRIBUTION
IN THE TWIN CITIES ESTIMATION SAMPLE

	Auto Drive	Auto Passenger	Public Bus	Total
		(Trips)		
Work	325	56	25	406
Personal Business	95	14	2	111
Medical	7	1	-	8
Social/Recreation	43	16	-	59
Outdoor Recreation	5	4	-	9
Shopping	75	16	-	91
School	28	8	1	37
Total	578	115	28	721
		(Percent)		
Work	80.0	13.8	6.2	100
Personal Business	85.6	12.6	1.8	100
Medical	87.5	12.5	-	100
Social/Recreation	72.9	27.1	-	100
Outdoor Recreation	55.6	44.4	-	100
Shopping	82.4	17.6	-	100
School	75.7	43.2	2.7	100
Total	80.2	15.9	3.9	100.0

SOURCE: Charles River Associates, 1976.

insufficient number of transit users to allow for mode split model estimation.

As with the Pittsburgh data base, each observation in the Twin Cities estimation sample contained detailed information on household socioeconomic and life cycle descriptors, characteristics of individual tripmakers, and LOS attributes of the chosen and rejected modes.

The Baltimore Disaggregate Data Set

The Baltimore Disaggregate Data Set is composed of five linked but separate computer files, describing 966 households, the vehicles owned by each household, the members of the household, summary origin and destination data on all trips taken by household members during the preceding day, and a detailed report on a randomly selected trip broken down into links, with an enumeration of all alternative modes or destinations for each link used within the last six months. A sixth file containing land-use characteristics of all origin and destination zones has been prepared which can be used to develop choice sets for destination choice model calibration. In order to estimate mode and destination choice models, several of these files must be merged.

Two approaches to constructing a suitable modeling data set were identified and are discussed after a summary of the contents of each of the six data files comprising the Baltimore Data Set.

The household file contains information on distance to public transportation, expressway, and other facilities. It also contains information on the total number of household members, automobiles, and travellers as well as other household characteristics such as renter/owner occupied, income, type of dwelling unit, and number of rooms. (For more detail on the data files, see Section 10 of Appendix D).

The vehicle file describes each vehicle available to a household. The information includes type, make, model and year, passenger capacity, how far it is parked from the household, type of parking place, cost to park, and who usually uses it to go to school or work. Much of this detail would be useful for disaggregate auto ownership models.

The person file contains information on each household member. These data include age, sex, employment, working hours, driving status, personal income, education, marital status, and language spoken.

The trip file contains a one-day trip diary for every household member over age 12. It parallels what is usually obtained in an origin-destination survey. For each trip it contains origin, destination, mode, purpose, household members making the trip, auto occupancy, household vehicle used, time of start and arrival, and land use. Systems data describing the travel time from the centroid of the origin zone to the centroid of the destination zone have been appended to the end

of each record. Originally, only the travel time for the selected mode, auto or transit, was available, making this file unsuitable for choice model estimation. Network data for some of the modal alternatives became available recently. These data will enable researchers to estimate mode choice models using the 7,000 trips described in this file. However, the only original choice for model calibration was the detailed link file.

The detailed link file describes links of so-called "complete round trip," for instance, from home back to home or from work back to work. This trip was selected at random by choosing a one-way trip made by a random household member. It contains data on the round trip and some alternatives to that trip. Information for each link of the round trip was obtained for the highway route taken or bus route number, toll costs, minutes to park or wait for a bus, parking facility type and costs, carpool cost sharing, bus fare, type of bus shelter, whether a seat was available on the bus, and time and distance in travel.

For discretionary detailed trips, information was obtained on alternate modes, when the mode was last taken, possible alternative origins and/or destinations, alternate times of day, and alternate days that the trip may have been made. The alternatives for nondiscretionary (work/school) trips are more narrowly defined; no alternative times or destinations were gathered. Systems data were appended to each reported link and its alternatives.

It would seem apparent that the detailed link file would be the better choice for a source of data for disaggregate choice model estimation. However, there were some limitations to the use of the data. Owing to the technique of identifying alternatives and to the observed travel patterns of households, only a limited number of alternatives were reported. (An alternative had to have been actually used in the last six months in order to be reported.) With respect to the household travel patterns observed, 135 primary respondents reported no travel. Consequently, only 831 detailed trip reports were gathered. A total of 779 alternative trips are reported. Of these only 418 reported alternative destinations for discretionary purposes and were therefore suitable for destination choice. Also, complex tours (or multidestination trips) are often covered in the detailed link data. These trips are not amenable to simple logit modeling for reasons discussed elsewhere in this report.

Two worktrip mode choice data sets were excerpted from the detailed link file. One described the 30 detailed work- or school-trips for which the primary respondent had used an alternative in the previous six months. (Data attrition using empirical choice sets from the DETLINK file is severe.)

A second data set included both the 199 work-/school-trips for which no respondent-specific alternatives data were gathered and the 30 observations from the first data set. The 199 observations are cases where the primary respondent had used no alternative mode for his detailed work-/school-trip in the six previous months. Network-level details on LOS were, however, available for these detailed trips on the final release tape. This afforded a larger sample of work-/school-trips for calibration purposes. Once missing data points were excluded, 175 observations remained in the merged data base.

2. GENERIC VERSUS MODE-SPECIFIC LEVEL-OF-SERVICE VARIABLES

Introduction

It should be recalled from Chapter 2 that there are two principal advantages to using generic LOS data in disaggregate transport choice models:

1. Generic LOS variables are consistent with economic utility theory; and
2. Use of generic LOS facilitates forecasts of demand for new choice alternatives.

The use of abstract commodity attributes in utility theory was introduced by Lancaster (D3), and applied in practical applications to numerous aggregate (D4) and disaggregate (D1, D5, D6) travel demand model studies. In a mode choice modeling framework, generic LOS representation assumes that an additional minute (or dollar) spent traveling on a bus is valued equally to an additional minute (or dollar) spent traveling by auto. If this were not the case it would be due to the effects of other mode-specific attributes omitted from the model (e.g. comfort, privacy, or reliability.) Thus, in a well-specified model that explicitly accounts for all attributes that significantly affect choice, the use of generic representations of LOS is justified.

A Test of the Hypothesis

In practice, it is generally not possible to ascertain a priori whether choice models are sufficiently well specified to justify the use of generic LOS variables. To test the validity of generic LOS representation, three binary logit mode split models were estimated using the Pittsburgh work trip data base (as shown in Equation D-1). In the first model, generic variables were employed for both in-vehicle travel time and out-of-pocket cost. The second model employed mode-specific travel times and costs, while the third model included a generic travel time variable and mode-specific costs.

$$\ln\left[\frac{p(a)}{P(t)}\right] = \theta_0 + \theta_1 \text{ HINC} + \theta_2 \text{ DIRCOMP} + \theta_3 \text{ INDCOMP} + \alpha_4 \text{ OVTT} + \tag{D-1}$$

$$\text{model 1} \left\{ \beta_1 \text{ IVTT} + \beta_2 \text{ COST} \right.$$

$$\text{model 2} \left\{ \begin{array}{l} \beta_{1a} \text{ AUTIME} + \beta_{1t} \text{ TRANTIME} + \\ \beta_{2a} \text{ AUTCOST} + \beta_{2t} \text{ TRANCOST} \end{array} \right.$$

$$\text{model 3} \left\{ \beta_1 \text{ IVTT} + \beta_{2a} \text{ AUTCOST} + \beta_{2t} \text{ TRANCOST} \right.$$

where:

- p(a),p(t) = mode selection probabilities for auto, transit.
- HINC = an income-specific dummy variable defined as 1 for travelers from households with income greater than $7,000, 0 otherwise. A measure of income-dependent auto preference.
- DIRCOMP = 1 if traveler's household has at least as many autos as workers, 0 otherwise. A measure of competition among workers for the household's auto(s).
- INDCOMP = 1 if traveler's household has at least as many autos as licensed drivers, 0 otherwise. A measure of competition by nonworkers for the household's auto(s).
- IVTT = difference in auto and transit in-vehicle travel time (generic).
- COST = difference in auto and transit out-of-pocket costs (generic).

AUTIME = auto in-vehicle travel time.

TRANTIME = transit in-vehicle travel time.

AUTCOST = auto out-of-pocket cost.

TRANCOST = transit out-of-pocket cost.

θ, β = estimated model parameters.

Table D-3 summarizes the estimation results for the LOS variables included in the three models. Complete listings of all parameter estimates for the three models are displayed in Tables D-4 through D-6.

In model 1, the estimates of both the generic travel time and cost parameters are of the correct sign and significant at the 5 percent confidence level. In model 2, the estimates of the parameters of transit cost and auto travel time are statistically insignificant and incorrect in sign. The transit cost parameter was also found to be insignificant in the estimation of model 3.

These estimation results suggest that our mode choice models are not able to distinguish significantly different traveler valuations of travel times and costs between auto and transit. In fact, the hypothesis that travelers' valuations of the LOS varibles do <u>not</u> differ between modes can be tested statistically. A well known method for testing a null hypothesis H_0 (in our case, the hypothesis that the time and cost parameters are not different between modes) against an alternative H_1 is the likelihood ratio test introduced in

Table D-3

TESTS OF THE VALIDITY OF GENERIC LEVEL OF SERVICE REPRESENTATION

	Model 1	Model 2	Model 3
INVTT	-.044 (-2.19)		-.041 (-2.17)
COST	-2.24 (-4.44)		
TRANTIME		.057 (2.243)	
AUTTIME		.01 (1.58)	
TRANCOST		-7.95 (-1.13)	.63 (.149)
AUTCOST		-3.26 (-3.35)	-2.25 (-4.4)
lnL (at convergence)	-23.821	-22.06	-23.741
$-2 \ln \frac{L(\theta \text{ model } 1)}{L(\theta \text{ model } n)}$,		3.52	.16
		$\chi^2_{.05}(2)=5.99$	$\chi^2_{.05}(1)=3.84$

SOURCE: Charles River Associates, 1976.

Table D-4

WORKTRIP MODE SPLIT ESTIMATION

RESULTS FOR GENERIC LOS REPRESENTATION

$$\ln \frac{P\ (auto)}{P\ (transit)} = \underset{(-3.60)}{-5.67} + \underset{(2.39)}{2.39\ HINC} + \underset{(3.48)}{3.30\ DIRCOMP} + \underset{(2.12)}{2.02\ INDCOMP}$$

$$\underset{(-1.72)}{-.10\ OVTT} - \underset{(-2.19)}{.04\ INVTT} - \underset{(4.44)}{2.24\ COST}$$

NOBS = 115

ρ^2 = .70

$L(0) = -79.71$ $L(\theta) = -23.82$

Percent Estimated Correctly = 93.04

P(auto) = probability of choosing auto

P(transit) = probability of choosing transit

HINC = 1 if household income exceeds $7000/year; 0 otherwise

DIRCOMP = 1 if number of autos is $\geq$ number of workers; 0 otherwise

INDCOMP = 1 if number of autos = number of licensed drivers; 0 otherwise

OVTT = difference in (auto-transit) out-of-vehicle travel time (in minutes)

INVTT = difference in (auto-transit) in-vehicle travel time (in minutes)

COST = difference in (auto-transit) cost in dollars

SOURCE: Charles River Associates, 1976.

Table D-5

WORKTRIP MODE SPLIT ESTIMATION

RESULTS FOR MODE-SPECIFIC TIME AND COST REPRESENTATION

$$\ln \frac{P\ (auto)}{P\ (transit)} = \underset{(-0.41)}{-1.46} + \underset{(2.08)}{2.17\ HINC} + \underset{(3.26)}{3.55\ DIRCOMP} + \underset{(2.26)}{2.44\ INDCOMP}$$

$$\underset{(-1.58)}{-.10\ OVTT} + \underset{(0.35)}{.01\ AUTIME} + \underset{(-2.43)}{.057\ TRANTIME} - \underset{(-3.35)}{3.26\ AUTCOST}$$

$$- \underset{(1.13)}{7.95\ TRANCOST}$$

NOBS = 115

ρ^2 = .72

$L(0) = -79.71$ $L(\theta) = -22.06$

Percent Estimated Correctly = 91.30

P(auto) = probability of choosing auto

T(transit) = probability of choosing transit

HINC = 1 if household income exceeds $7000/year; 0 otherwise

DIRCOMP = 1 if number of autos is $\geq$ number of workers; 0 otherwise

INDCOMP = 1 if number of autos = number of licensed drivers; 0 otherwise

OVTT = difference in (auto-transit) out-of-vehicle travel time (in minutes)

AUTIME = auto in-vehicle travel time in minutes

TRANTIME = transit in-vechicle travel time in minutes

AUTCOST = auto cost in dollars

TRANCOST = transit cost in dollars

SOURCE: Charles River Associates, 1976.

Table D-6

WORKTRIP MODE SPLIT ESTIMATION RESULTS

FOR MODE-SPECIFIC COST REPRESENTATION

$$\ln \frac{P\,(auto)}{P\,(transit)} = \underset{(-1.63)}{-4.71} + \underset{(2.35)}{2.37\,HINC} + \underset{(3.44)}{3.37\,DIRCOMP} + \underset{(2.09)}{2.01\,INDCOMP}$$

$$\underset{(-1.73)}{-.10\,OVTT} \quad \underset{(-2.17)}{-.04\,IVTT} + \underset{(.149)}{.62\,TRANCOST} \quad \underset{(-4.41)}{-2.25\,AUTCOST}$$

NOBS = 115

ρ^2 = .70

L(0) = -79.71 *L(θ)* = -23.75

Percent Estimated Correctly = 93.04

P(auto) = *probability of choosing auto*

P(transit) = *probability of choosing transit*

HINC = 1 *if household income exceeds $7000/year; 0 otherwise*

DIRCOMP = 1 *if number of autos is* $\geq$ *number of workers; 0 otherwise*

INDCOMP = 1 *if number of autos = number of licensed drivers; 0 otherwise*

OVTT = *difference in (auto-transit) out-of vehicle travel time (in minutes)*

IVTT = *difference in (auto-transit) in-vehicle travel time (in minutes)*

TRANCOST = *transit cost in dollars*

AUTCOST = *auto cost in dollars*

SOURCE: Charles River Associates, 1976.

Chapter 2. Theil (D7) has shown that the statistic $-2 \ln \lambda$ (where λ is the ratio of the likelihood of the constrained parameter vector which falls under H_0 to the likelihood from H_1) is asymptotically distributed as $\chi 2$. (This $\chi 2$ distribution has q degrees of freedom, where q is the number of parameter constraints in H_0. See Chapter 2 for a discussion of the likelihood ratio test.) Table D-3 displays the value of the likelihood ratio test statistic derived from models 1 (constrained), 2, and 3 (unconstrained), as well as the critical value of the $\chi 2$ distribution at the 95 percent confidence level. It is clear that in our application we cannot reject the null hypothesis that the time and cost parameters do not differ between modes.

A similar finding on the statistical validity of generic LOS variable representation was found in estimation results on the Pittsburgh shopping mode split models. Table D-7 displays parameter estimates for a binary logit mode split shopping model with generic LOS representation. In Table D-8 mode-specific travel time and cost parameters were estimated. The likelihood ratio test statistic for this case was computed as $\chi 2(2) = 0.48$, well below the critical value of $\chi 2(2)$ at the 5 percent significance level (5.99).

These findings are very encouraging. We noted at the outset that in models that explicitly specify all significant influences on travel choice, one would expect time or cost to

Table D-7

SHOPPING-TRIP MODE SPLIT ESTIMATION

RESULTS FOR GENERIC LOS REPRESENTATION

$$\ln \frac{P\ (auto)}{P\ (transit)} = \begin{array}{l} -10.69 + 3.18\ HINC + 4.48\ APERDR - .55\ OVTT - 0.48\ INVTT \\ (-3.19)\quad (2.36)\qquad (2.07)\qquad\quad (-2.82)\quad (-1.09) \end{array}$$

$$\begin{array}{l} -6.78\ COST \\ (-3.37) \end{array}$$

NOBS = 140

ρ^2 = .86

L(0) = -97.04 L(θ) = -13.87

Percent Estimated Correctly = 94.29

P(auto) = probability of choosing auto

P(transit) = probability of choosing transit

HINC = 1 if household income exceeds $7000/year; 0 otherwise

APERDR = autos per licensed driver

OVTT = difference in (auto-transit) out-of-vehicle travel time (in minutes)

INVTT = difference in (auto-transit) in-vehicle travel time (in minutes)

COST = difference in (auto-transit) cost in dollars

SOURCE: Charles River Associates, 1976.

Table D-8

SHOPPING-TRIP MODE SPLIT ESTIMATION

RESULTS FOR MODE-SPECIFIC TIME AND COST LOS REPRESENTATION

$$\ln \frac{P\ (auto)}{P\ (transit)} = \begin{array}{l} -9.64 + 3.28\ HINC + 4.45\ APERDR \quad -.07\ AUTIME \\ (-2.22)\ (2.05)\qquad (1.96)\qquad\quad (-1.04) \end{array}$$

$$\begin{array}{l} -.04\ TRANTIME - 6.18\ AUTCOST \quad -5.83\ TRANCOST \\ (-0.62)\qquad\quad (-2.76)\qquad\quad (-0.94) \end{array}$$

NOBS = 140

ρ^2 = 0.86

L(0) = -97.04 L(θ) = -13.63

Percent Estimated Correctly = 94.29

P(auto) = probability of choosing auto

P(transit) = probability of choosing transit

HINC = 1 if household income exceeds $7000/year; 0 otherwise

APERDR = autos per licensed driver

AUTIME = auto in-vehicle travel time in minutes

TRANTIME = transit in-vehicle travel time in minutes

AUTCOST = auto cost in dollars

TRANCOST = transit cost in dollars

SOURCE: Charles River Associates, 1976.

be valued "abstractly." The models presented here are admittedly weak in differentiating the comfort, safety, privacy and other amenity characteristics between alternative modes. Nonetheless, we have found that use of generic LOS variables is statistically justified. See Chapter 2 for a discussion of other findings relevant to the use of generic LOS variables.

3. INCOME SEGMENTATION

Introduction

An important empirical question for disaggregate model estimation and application concerns the need for market segmentation. The issue here is that unless proper account is taken of identifiable and systematic variations in travel behavior between different types of travelers (market segments), the resulting models may have two (related) unfortunate properties:

- Some of the estimated coefficients may be biased; and
- The models will be of questionable validity in inferring traveler behavior for groups of individuals whose tastes differ from that of the sample used for estimation.

Proper account can be taken either by estimating and applying different models for each identified market segment or by including a sufficient number of explanatory variables in the models to account for market segment differentiation. Empirical research on this project has explored the importance

of accounting for differences in travel behavior between income-stratified market segments using the worktrip mode split Pittsburgh data base.

Household income may be entered in a binary mode split model either as a pure alternative-specific variable having the value of income in the utility function of one mode and zero in the other, or as a generic variable when income is combined with a variable that varies from one mode to another.

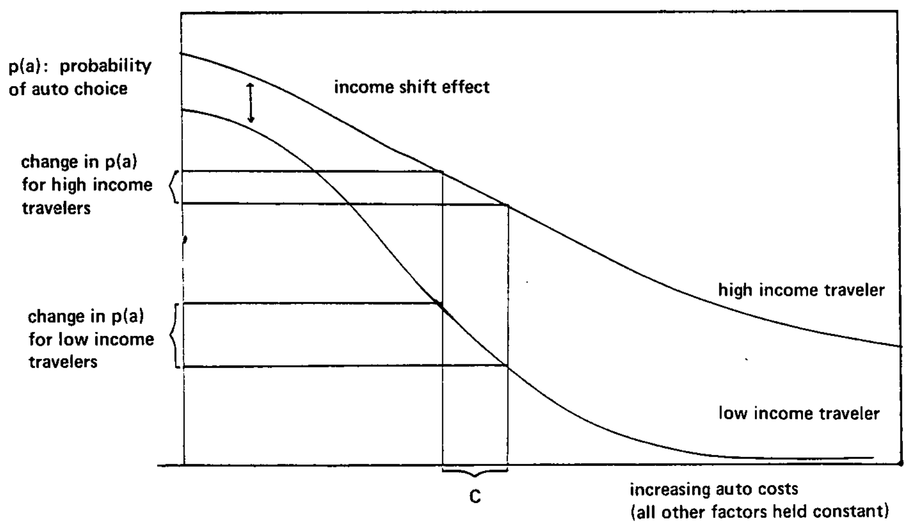
Representing income as an alternative-specific variable allows for an "income shift effect," e.g., a case where higher income travelers have a higher preference for auto choice than lower income travelers (all other factors being equal). Combining income with a generic variable, e.g., dividing cost by income, allows for a test of the hypothesis that travelers of different income circumstances value travel costs differently.

A variety of functional forms may be specified to test the hypothesis of income-dependent value of time variation. The technique employed here, dividing travel cost by income, represents one possible way of "weighting" the LOS variables by household income.

Figure D-1 displays the two different types of income effects on traveler choice behavior. Representing income by

Figure D–1

INCOME DIFFERENTIATED LOGIT RESPONSE CURVES



both an alternative specific variable and combined with a generic travel cost variable generates distinctly different income-differentiated logit response curves. In Figure D-1 two such curves are shown corresponding to a low- and high-income traveler. The pure income shift effect is illustrated by the differences in the auto selection probabilities for low- and high-income travelers when auto costs (and all other factors) are equal. As shown, higher income travelers have a higher auto preference than low income travelers.

To illustrate the income-differentiated valuation of travel cost, Figure D-1 displays the change in auto choice probabilities for low- and high-income travelers when auto costs change by the amount ΔC. As indicated in the figure, low-income travelers appear to be more sensitive to travel costs than high income travelers. This can be seen by comparing the respective changes in auto choice probabilities for the two income groups. For the given change in auto costs, there is a greater change in the probability of auto choice for low-income than for high-income travelers. An empirical test of these hypothesized income effects in travel choice behavior is presented in the following paragraphs.

The Pure Income Shift Effect

In order to test the hypothesis on the "pure shift" effect of income on travel choice behavior, two models of mode split for work trips were estimated using the Pittsburgh data set. In model 1 shown in Table D-9, no account was taken for the income status of the traveler. Model 2 illustrated in Table D-10 is identical to model 1 except for the addition of an income shift dummy variable, HINC defined as 0/1 for low-/high-income travelers (with $7,000 per year as the cut-off point). HINC was entered as the 0/1 dummy variable in the auto utility function and as 0 in the transit utility function.

The coefficient of the income term is significant at the 1 percent level and, as expected, has a positive sign. This indicates that higher income travelers express a higher preference for auto than lower income travelers. Moreover, comparing the predictive power of the models, model 2 (Table D-10) has a higher percentage of correctly predicted mode choices in the estimation sample (93.04 percent versus 91.30 percent). It is also important to note how the coefficients of the other variables differ between the two models.

In both models the intercept term is a measure of the "transit bias" of the sample used to calibrate the model (the

Table D-9

WORKTRIP MODE SPLIT MODEL WITH NO ACCOUNT OF INCOME EFFECTS

$$\ln \frac{P\ (auto)}{P\ (transit)} = \underset{(-3.62)}{-3.84} + \underset{(3.77)}{3.42}\ DIRCOMP + \underset{(1.47)}{1.15}\ INDCOMP$$

$$- \underset{(-2.22)}{.112}\ OVTT - \underset{(-2.21)}{.041}\ INVTT - \underset{(4.47)}{2.02}\ COST$$

NOBS = *115*

ρ^2 = *.66*

$L(0) = -79.71 \quad L(\theta) = -27.29$

Percent Estimated Correctly = 91.30

P(auto) = probability of choosing auto

P(transit) = probability of choosing transit

DIRCOMP = 1 if number of autos is $\geq$ number of workers; 0 otherwise

INDCOMP = 1 if number of autos = number of licensed drivers; 0 otherwise

OVTT = difference in (auto-transit) out-of-vehicle travel time (in minutes)

INVTT = difference in (auto-transit) in-vehicle travel time (in minutes)

COST = difference in (auto-transit) cost in dollars

SOURCE: Charles River Associates, 1976.

Table D-10

WORKTRIP MODE SPLIT MODEL WITH A PURE INCOME SHIFT EFFECT INCLUDED

$$\ln \frac{P\,(auto)}{P\,(transit)} = \underset{(-3.60)}{-5.67} + \underset{(2.39)}{2.39}\ HINC + \underset{(3.48)}{3.30}\ DIRCOMP + \underset{(2.12)}{2.02}\ INDCOMP$$

$$+ \underset{(-1.72)}{-.10}\ OVTT - \underset{(-2.19)}{.04}\ INVTT \underset{(4.44)}{-2.24}\ COST$$

NOBS = 115

ρ^2 = .70

L(0) = -79.71   L(θ) = -23.82

Percent Estimated Correctly = 93.04

P(auto) = probability of choosing auto

P(transit) = probability of choosing transit

HINC = 1 if household income exceeds $7000/year; 0 otherwise

DIRCOMP = 1 if number of autos is ≥ number of workers; 0 otherwise

INDCOMP = 1 if number of autos = number of licensed drivers; 0 otherwise

OVTT = difference in (auto-transit) out-of-vehicle travel time (in minutes)

INVTT = difference in (auto-transit) in-vehicle travel time (in minutes)

COST = difference in (auto-transit) cost in dollars

SOURCE: Charles River Associates, 1976.

sample was chosen from a travel corridor with a relatively high level of transit service). The "transit bias" estimated for the low-income population (-5.67 in Table D-10) is above that estimated for the aggregate population (-3.84 in Table D-9), and that of the high-income population (-5.67 + 2.24 = -3.43) is below that of the aggregate population. The two groups are clearly different market segments.

The importance of explicitly accounting for differences in traveler behavior among different income classes has a direct bearing on the validity of applying disaggregate models estimated on one group of data (e.g., from one city) to forecast travel behavior for another group of travelers. Ignoring income-specific travel behavior when it is important will result in parameter estimates that are dependent on the income distribution found in the sample used for estimation.

In order to show the importance of this problem, it is instructive to consider the following hypothetical example where the parameter estimates of the two models presented earlier are applied to predicting the mode choices of two homogeneous income groups. In particular, assume that the characteristics of these groups are as summarized in Table D-11.

Table D-11

ASSUMED CHARACTERISTICS OF TWO

INCOME-DIFFERENTIATED HOMOGENEOUS TRAVEL GROUPS

	Low-Income Group	High-Income Group
Number of autos in household	1	2
Number of workers in household	1	1
Number of licensed drivers in household	2	2
Transit walk time	10 min.	10 min.
Dif. in in-veh. time (auto-transit)	-15 min.	-20 min.
Dif. in cost (auto-transit)	$1.00	$1.50

SOURCE: Charles River Associates, 1976.

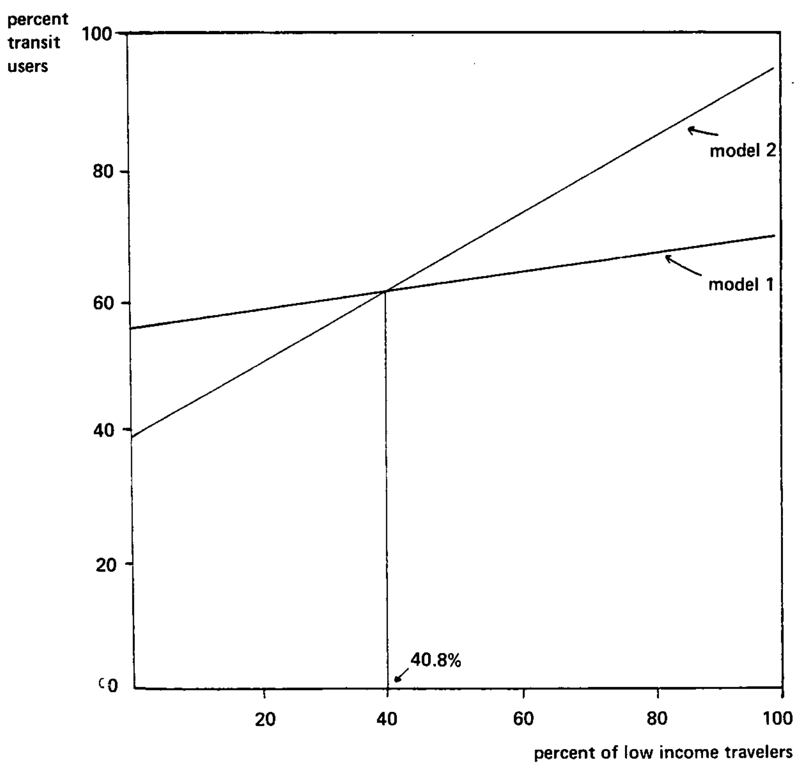
Applying our first model, where no income shift effect is incorporated, to these two household types yields transit use probabilities of 0.669 and 0.589 for low- and high-income households respectively. The second model (including the income variable) predicts low-and high-income household transit use probabilities of 0.950 and 0.395 respectively. It is clear that the second model, which takes account of an income shift effect, is more discriminating in this example in the sense that the mode choice probabilities are further out on the "tails" of the logistic function.

Moreover, the results here indicate that predictions of <u>population</u> mode choice splits are sensitive to the distribution of income in the sample. Figure D-2 displays the percent transit of users among a sample of two homogeneous income groups. Our second model shows a relatively large variation in mode splits as a function of the distribution of income in the forecast sample. The two predict equal mode splits for a sample with 40.8 percent low-income travelers. This is approximately equal to the percent of low-income travelers (income < $7,000) in the sample used for estimation of the work mode split models.

The example presented in this section is representative of the general problem of <u>specification</u> <u>error</u>. Omitting variables (in our case, income) that significantly influence the choice behavior of individuals may have two unfortunate consequences:

Figure D–2

POPULATION MODE CHOICE PREDICTIONS FOR INCOME-DIFFERENTIATED HOMOGENEOUS TRAVEL GROUPS



- The estimated model will yield large forecasting errors when applied to a sample whose distribution of the omitted variable differs significantly from the estimation sample; and
- The parameter estimates of those included variables significantly correlated with the omitted variable(s) will be biased.

Our tentative results presented here suggest the importance of improving model specification. Data that are somewhat difficult to collect and forecast accurately (e.g., income) may still be employed in some proxy fashion to improve the predictive ability of disaggregate choice models.

It also should be noted that income data are often difficult to collect accurately. In the sample used to estimate the models presented earlier in this section, nearly half of the respondents refused to report their family income. In these cases, the interviewers estimated household income (based on quality of the housing unit, subjective evaluations of the cost of furnishings, etc.). While the income data are somewhat suspect in measuring income precisely, their ability at least to distingish between "low" and "high" wage earners proved to be significantly explanatory of choice behavior.

Income Effects on the Valuation of Travel Time and Cost

In addition to testing for the significance of an "income shift effect" on mode split travel behavior, the Pittsburgh work and shopping mode split models were used to investigate the effect of income on commuter's valuations of travel times and costs. In this instance, the models took the form as shown in Tables D-12 and D-13 for work and shopping mode choice respectively.

It may be noted from Tables D-12 and D-13 that auto availability is measured by the single term autos per worker (APERW) rather than the two separate auto availability terms DIRCOMP and INDCOMP used in the models presented earlier in this section. Because of the high correlation between household income and automobile ownership, it was not possible to obtain significant parameter estimates of both worker competition (DIRCOMP) and nonworker competition (INDCOMP) auto availability effects when income was entered in the travel cost variable. Section 4 of this appendix will discuss in detail the effects of auto availability on mode choice behavior.

Income now enters the model both as an alternative-specific auto preference shift term and as a generic variable combined with the travel cost term. The variable C/INC (travel cost divided by income) is intended to reflect the hypothesis

Table D-12

WORK MODE SPLIT MODEL WITH INCOME-DIFFERENTIATED TRAVEL COST TERM

$$\ln \frac{P\,(auto)}{P\,(transit)} = \underset{(3.78)}{-5.72} + \underset{(1.44)}{1.37}\,HINC + \underset{(3.74)}{4.07}\,APERW - \underset{(-2.18)}{.117}\,OVTT - \underset{(-1.88)}{.0348}\,INVTT$$

$$- \underset{(-4.28)}{9.06}\,C/INC$$

NOBS = 115
ρ^2 = .66
$L(0) = -79.71 \quad L(\theta) = -25.13$
Percent Estimated Correctly = 93.04

P(auto) = *probability of choosing auto*
P(transit) = *probability of choosing transit*
HINC = 1 *if household income exceeds $7000/year; 0 otherwise*
APERW = *autos per worker*
OVTT = *difference in (auto-transit) out-of-vehicle travel time (in minutes)*
INVTT = *difference in (auto-transit) in-vehicle travel time (in minutes)*
C/INC = *difference in (auto-transit) cost in dollars divided by income code (see Table D-15)*

SOURCE: Charles River Associates, 1976.

Table D-13

SHOPPING MODE SPLIT MODEL WITH INCOME-DIFFERENTIATED TRAVEL COST TERM

$$\ln \frac{P(auto)}{P(transit)} = \underset{(-4.36)}{-8.79} + \underset{(2.99)}{3.16}\ APERW + \underset{(2.06)}{2.05}\ HINC - \underset{(-3.73)}{.44}\ OVTT - \underset{(-1.58)}{.04}\ INVTT$$

$$\underset{(-3.70)}{-15.01}\ C/INC$$

NOBS = 140

$\rho^2 = .77$

L(0) = -97.04 L(θ) = -22.19

Percent Estimated Correctly = 94.29

P(auto) = probability of choosing auto

P(transit) = probability of choosing transit

APERW = autos per worker

HINC = 1 if household income exceeds $7,000/year; 0 otherwise

OVTT = difference in (auto-transit) out-of-vehicle travel time (in minutes)

INVTT = difference in (auto-transit) in-vehicle travel time (in minutes)

C/INC = difference in (auto-transit) cost in dollars divided by income *(see Table D-16)*

SOURCE: Charles River Associates, 1976.

that travelers with different income circumstances value travel costs differently. Thus, this form allows for a separate measure of travel time and cost elasticities and value of time by income group. As noted earlier in this section, several different forms of "weighting" LOS terms by income may be employed to test this hypothesis.

In both the shopping and work mode split models, all parameters were of the right sign and significant at the 10 percent level. In order to interpret these estimation results in terms of their implied income-differentiated travel behavior, it is convenient to derive respective <u>value of time</u> measures obtained from the mode choice models.

The derivation of the value of time (VOT) measure from a logit mode split model is straightforward if VOT is interpreted as the marginal rate of substitution between travel time and travel cost for consumer indifference between two alternative modes. We set the total derivative of the choice probability, p_i to zero:

$$dp_i = 0 = d\ \frac{e^{V_i}}{e^{V_i} + e^{V_j}} = 0 = \qquad \text{(D-2)}$$

$$\frac{\partial p_i}{\partial t}\, dt + \frac{\partial p_i}{\partial c}\, dc = 0$$

Thus,

$$VOT = \frac{dc}{dt} = \frac{\frac{\partial p_i}{\partial t}}{\frac{\partial p_i}{\partial c}} = \frac{\alpha_t p_i (1-p_i)}{\frac{\alpha_c p_i (1-p_i)}{INC}} = \frac{\alpha_t}{\alpha_c} INC \qquad (D\text{-}3)$$

where:

- p_i = probability of choosing mode i;
- V_i = the utility function associated with mode i;
- dt,dc = travel time and cost differentials, respectively;
- α_t = coefficient of the travel time variable;
- α_c = coefficient of the travel cost variable; and
- INC = household income.

Using the parameter estimates from the work and shopping mode split models in Equation D-3, Table D-14 summarizes the value of time measures for two representative income groups, a high-income household (with assumed income of $16,000 per year) and a low-income household (with assumed income of $8,000 per year). Income was entered in the work and shopping mode split models with a code value. See Tables D-15 and D-16 for the respective income codes used in the two models.

The results indicate that value of time is an increasing function of household income for both work and shopping mode choice behavior. This finding is consistent with the hypothesized income-differentiated travel response behavior displayed in Figure D-1.

Table D-14

VALUE OF TIME MEASURES FOR WORK- AND SHOPPING-TRIPS*

(1967 Dollars per Hour)

	Low-Income Households (income=$8,000/year)	High-Income Households (income=$16,000/year)
Work Mode Split	$0.92	$1.38
Shopping Mode Split	$1.26	$2.76

*For in-vehicle travel time.

SOURCE: Charles River Associates, 1976.

Table D-15

INCOME CODES USED IN MODELS ESTIMATED WITH THE PITTSBURGH WORK DATA BASE

Code	Household Income Range (1967 Dollars)
1	< $3,000
2	$3,000 - $4,999
3	$5,000 - $6,999
4	$7,000 - $9,999
5	$10,000 - $14,999
6	$15,000 - $19,999
7	$20,000 - $24,999
8	$25,000 - $29,999
9	$30,000 or more

SOURCE: Charles River Associates, 1976.

Table D-16

INCOME CODES USED IN MODELS ESTIMATED WITH THE PITTSBURGH SHOPPING DATA BASE

Code*	Household Income Range (1967 Dollars)
1.5	< $3,000
4.0	$3,000 - $4,999
6.0	$5,000 - $6,999
8.0	$7,000 - $8,999
8.5	$9,000 - $9,999
12.5	$10,000 - $14,999
17.5	$15,000 - $19,999
22.5	$20,000 - $24,999
27.5	$25,000 - $29,999
35.0	$35,000 or more

*Code is the midpoint of the corresponding income range.

SOURCE: Charles River Associates, 1976.

4. AUTO AVAILABILITY EFFECTS ON MODE CHOICE

Introduction

It has been a frequent finding of empirical research on disaggregate demand modeling that household automobile ownership significantly influences mode choice to work (D1, D8). There are two basic problems with the use of auto ownership variables in mode choice models.

- Travel decisions are not independent of household mobility decisions. Consequently, parameter estimates of auto ownership variables in disaggregate mode choice models will probably be biased.
- It is not so much the number of autos in a household as the availability of an auto at the time a trip is made that influences not only mode choice, but also decisions on where, when, and how often to travel.

Research on this project has focused on the latter issue -- understanding the distinction between automobile ownership and automobile availability, and how these factors influence travel behavior. In the short run, improvements in the performance of disaggregate demand models can be gained by improving their representation of auto availability effects. The larger question of modeling the interaction of household location, automobile ownership, and household travel remains as a long-range research issue. The Baltimore data set does offer the opportunity to evaluate this issue in some ways, but was beyond the scope of this project.

In understanding the influence of automobile availability on travel behavior, it is useful to introduce the notion of competition for use of a car within the household. Generally, the greater the number of drivers in a household, the greater will be the competition for use of the household's automobiles. (This is a short-run analysis in the sense that we are considering the number of autos in the household fixed.) Stated another way, as the competition for use of a household's automobile(s) increases, the probability that an auto is used for any given trip should decrease. This is particularly true for a household's worktrips where there is little flexibility on when the trip can be made. For other types of trip purposes (e.g. shopping), competition for use of the household's autos may be resolved by adjusting the timing as well as the choice of mode for travel.

The original work mode split models appearing in CRA (D1) incorporated a variable measuring automobiles per worker in the household. This definition partly accounts for worktrip competition for use of the household's automobiles for worktrips. As expected, the parameter estimate of this term was positive suggesting that (other factors held constant):

- An increase in the number of automobiles increases the probability of auto mode choice for work travel; and
- Conversely, if an additional household member enters the work force, the probability that any given worker uses auto for his worktrip decreases.

A Revised Model Specification

A revised specification of the work mode split models using the Pittsburgh disaggregate data set was estimated in this project to assess the effects of both worker and nonworker competition for use of a household's automobiles. In this context, we refer to worktrip competition as the number of workers in a household who may make exclusive use of the household's auto(s) for their worktrip. Nonworktrip competition expresses the possibly mutually exclusive uses of the household's auto(s) between licensed workers and licensed nonworkers.

Formally, we define

DIRCOMP = worker competition effect: 1 if the number of autos in a household is greater than or equal to the number of full-time workers; 0 otherwise; and

INDCOMP = nonworker competition effect: 1 if the number of autos is greater than or equal to the number of licensed household residents; 0 otherwise.

To see how these variables express the competition for use of a household's autos, consider the following household types shown in Table D-17. In the first household type, there is only one licensed driver. His use of the car for the worktrip does not compete with other household members' potential need for auto driving. From the definitions above, the values of both INDCOMP and DIRCOMP are 1 in this case. Household type 2 is representative of nonworker auto competition. Although there is still only one worker and one auto, this household has a second licensed driver. In this instance, the variable INDCOMP assumes the value of 0. Finally, household type 3 characterizes a case of both worker and nonworker competition. This household's one auto must be shared between two licensed workers and one additional licensed nonworker. The estimated model incorporating these two auto availability/competition variables is shown in Table D-18.

The estimation results are consistent with our a priori expectations of the effects of auto availability. The parameter estimates of DIRCOMP and INDCOMP are positive and significant (at the 5 percent level) suggesting that both worker and nonworker competition for auto use within a household affect work mode choice. Moreover, it is apparent that the direct competition of two workers for one auto is more

Table D-17

HOUSEHOLD TYPES DIFFERENTIATED BY AUTO AVAILABILITY

Household Type	Number of Autos	Number of Workers	Number of Licensed Drivers	Value of DIRCOMP	Value of INDCOMP
1	1	1	1	1	1
2	1	1	2	1	0
3	1	2	3	0	0

SOURCE: Charles River Associates, 1976.

Table D-18

WORK MODE SPLIT MODEL WITH INCLUSION OF

AUTO USE COMPETITION TERMS

$$\ln \frac{P\ (auto)}{P\ (transit)} = \underset{(-3.60)}{-5.67} + \underset{(2.39)}{2.39}\ HINC + \underset{(3.48)}{3.30}\ DIRCOMP + \underset{(2.12)}{2.02}\ INDCOMP$$

$$+ \underset{(-1.72)}{-.10}\ OVTT - \underset{(-2.19)}{.04}\ INVTT \underset{(4.44)}{-2.24}\ COST$$

NOBS = 115

ρ^2 = .70

$L(0) = -79.71$ $L(\theta) = -23.82$

Percent Estimated Correctly = 93.04

P(auto) = *probability of choosing auto*

P(transit) = *probability of choosing transit*

HINC = *1 if household income exceeds \$7000/year; 0 otherwise*

DIRCOMP = *1 if number of autos is ≥ number of workers; 0 otherwise*

INDCOMP = *1 if number of autos = number of licensed drivers; 0 otherwise*

OVTT = *difference in (auto-transit) out-of-vehicle travel time (in minutes)*

INVTT = *difference in (auto-transit) in-vehicle travel time (in minutes)*

COST = *difference in (auto-transit) cost in dollars with auto costs computed at 3 cents per mile*

SOURCE: Charles River Associates, 1976.

significant in determining work mode choice than indirect competition for household auto use by nonworkers as evidenced by the relative magnitude of the parameter estimates of DIRCOMP and INDCOMP. This finding is consistent with our earlier comments on the flexibility of scheduling nonwork travel around availability of the household's autos.

In order to focus on the effects of auto availability on work mode choice, the estimated model was applied to the three prototypical households shown in Table D-17. Values of the LOS variables in this analysis were taken at the sample averages of the Pittsburgh data base. The results of this analysis are summarized in Table D-19. Two points should be noted here:

- Both worker and nonworker competition for use of household autos influence work mode choice; and
- Failure to account for either effect may seriously alter mode split forecasts. The use of a single measure of autos per worker, for example, would fail to differentiate between the circumstances of the household type 1 and 2 (see Table D-17). Our analysis here has shown that this omission is important.

While more research is needed to account fully for the interaction of household mobility and travel decision processes, our research has indicated that models for

Table D-19

MODE CHOICE PREDICTIONS FOR THREE DIFFERENT

AUTO OWNERSHIP HOUSEHOLD TYPES

Mode Split Estimates for Three Household Types

Household Type 1: No Competition for Use of Auto by Primary Worker

Household Type 2: Competition for Use of Auto by Primary Worker

Household Type 3: Secondary Worker and Nonworker Competition for Use of Auto by Primary Worker

PREDICTED MODE SPLITS

(Estimated at Sample Average)

	Auto*	Transit	Percent Change of Auto Probability from Household Type 1 Case
Household Type 1	0.71	0.29	-
Household Type 2	0.44	0.56	0.47
Household Type 3	0.02	0.98	0.19

*Driver-only auto trips.

SOURCE: Charles River Associates, 1976.

predicting (short-run) mode choice can be improved by incorporating better indicators of automobile availability. This finding must, of course, be balanced against the increased data burden in aggregate forecasting to identify the percentage of the population that falls in the three categories of household types.

5. ALTERNATIVE SPECIFICATIONS OF THE LOGIT MODELS

Introduction

In Chapter 2 it was noted that LOS variables (e.g., travel time, cost, etc.) could be introduced in a variety of complex transformations (e.g., logarithms, ratios, etc.) to represent different travel behavior hypotheses.

One such transformation that was tested in models estimated on the Pittsburgh mode split data base was a logarithmic transformation of the travel (in-vehicle) time and cost variables. Behaviorally, this transformation corresponds to the hypothesis that a traveler's sensitivity to absolute changes in travel time and cost decreases for longer (or more expensive) trips. For example, for a 10-minute travel time difference between auto and transit when auto travel time is 20 minutes, the logarithm of the time difference is 0.405. The same 10-minute time differential when auto travel is 50 miutes yields a log time differential of only 0.812. Thus, the logarithmic transformation tends to "deflate" LOS differentials between auto and transit for longer trips.

Estimation Results for the Two Model Forms

Tables D-20 and D-21 summarize the estimation results for two binary mode split models estimated with the Pittsburgh disaggregate data set. Both models have identical specifications except for the representation of in-vehicle travel time and travel cost. In the model in Table D-20 these variables entered linearly, while in Table D-21 travel time and cost entered logarithmically. In both specifications, all the estimated parameters were of the correct sign and significant at the 5 percent level. However, the linear LOS specification yielded a slightly better goodness of fit as measured by the respective coefficients of determination, ρ^2 and the percent of correct predictions. (See Chapter 2 for a detailed discussion of these model evaluation measures.) For the linear LOS specification model, ρ^2 was equal to 0.70, and 93.04 percent of the observed mode choices in the estimation sample had a predicted mode selection probability greater than 50 percent. This compares to ρ^2 of 0.63 and 90.43 percent "correct" predictions for the model with logarithmic LOS representation.

The differences in the goodness of fit measures of the two model forms are not significant. Accordingly, the estimation results do not strongly support either hypothesis on the

Table D-20

WORKTRIP MODE SPLIT ESTIMATION
RESULTS FOR LINEAR REPRESENTATION OF
IN-VEHICLE TRAVEL TIME AND COST

$$\ln \frac{P\,(auto)}{P\,(transit)} = \underset{(-3.60)}{-6.47} + \underset{(2.39)}{2.05\ HINC} + \underset{(3.48)}{3.12\ DIRCOMP} + \underset{(2.12)}{1.67\ INDCOMP}$$

$$\underset{(-1.72)}{-.10\ OVTT} - \underset{(-2.19)}{.04\ INVTT} - \underset{(4.44)}{2.24\ COST}$$

NOBS = 115
ρ^2 = .70
$L(0) = -79.71$ $L(\theta) = -23.82$
Percent Estimated Correctly = 93.04

P(auto)	= probability of choosing auto
P(transit)	= probability of choosing transit
HINC	= 1 if household income exceeds $7000/year; 0 otherwise
DIRCOMP	= 1 if number of autos is $\geq$ number of workers; 0 otherwise
INDCOMP	= 1 if number of autos = number of licensed drivers; 0 otherwise
OVTT	= difference in (auto-transit) out-of-vehicle travel time (in minutes)
INVTT	= difference in (auto-transit) in-vehicle travel time (in minutes)
COST	= difference in (auto-transit) cost in dollars

SOURCE: Charles River Associates, 1976.

Table D-21

WORKTRIP MODE SPLIT ESTIMATION
RESULTS FOR LOGARITHMIC REPRESENTATION
OF IN-VEHICLE TRAVEL TIME AND COST

$$\ln \frac{P(auto)}{P\ transit} = \underset{(-4.36)}{-6.47} + \underset{(2.37)}{2.05\ HINC} + \underset{(3.44)}{3.12\ DIRCOMP} + \underset{(2.15)}{1.67\ INDCOMP}$$

$$\underset{(-2.07)}{-0.11\ OVTT} \quad \underset{(-2.39)}{-2.03 LOGTIME} \quad \underset{(-3.74)}{-1.34\ LOGCOST}$$

NOBS = 115
ρ^2 = .63
$L(0) = -79.71$ $L(\theta) = -29.39$
Percent Estimated Correctly = 90.43

P(auto)	= probability of choosing auto
P(transit)	= probability of choosing transit
HINC	= 1 if household income exceeds $7000/year; 0 otherwise
DIRCOMP	= 1 if number of autos is $\geq$ number of workers; 0 otherwise
INDCOMP	= 1 if number of autos = number of licensed drivers; 0 otherwise
OVTT	= difference in (auto-transit) out-of-vehicle travel time (in minutes)
LOGTIME	= log of difference in (auto-transit) travel time in minutes
LOGCOST	= log of difference in (auto-transit) cost in dollars

SOURCE: Charles River Associates, 1976.

marginal valuation of travel times and costs (namely, constant marginal valuation versus decreasing marginal valuation of travel times and costs). On a priori grounds, the hypothesis of decreasing marginal valuation of travel time and costs is appealing. A 10-minute travel time difference between auto and transit may be more a significant factor in determining mode choice for a one-mile journey to work than for a 40-mile worktrip. However, the range of travel times and costs in the sample used for our model estimation was limited. (For example, the average transit travel time in the estimation data sample was 55.7 minutes with a standard deviation of only 29.02. For transit cost the average and standard deviation were .63 and .11 respectively.) Within this range the data do not provide a powerful test of the hypothesis of decreasing marginal valuation of LOS.

Properties of the Two Model Forms

While the overall goodness of fit does not differ appreciably between the two model forms discussed above, predicted mode selection probabilities and elasticities of demand with respect to travel times and costs appear to be highly sensitive to the specification of variables. This can be seen by referring to Table D-22 which presents predicted probabilities and travel time elasticities for the two models

Table D-22

TRAVEL TIME ELASTICITIES FOR
TWO ALTERNATIVE MODEL SPECIFICATIONS*

Auto Travel Time In Minutes	Log Form p_a	Log Form $\eta_{p_a \| t_a}$	Linear Form p_a	Linear Form $\eta_{p_a \| t_a}$
10	0.899	-0.205	0.834	-0.073
20	0.685	-0.640	0.764	-0.210
30	0.488	-1.040	0.675	-0.433
35	0.411	-1.197	0.624	-0.583
40.675**	0.340	-1.342	0.564	-0.787
45	0.295	-1.423	0.516	-0.966
50	0.253	-1.519	0.461	-1.196
60	0.189	-1.648	0.354	-1.719
70	0.146	-1.736	0.260	-2.297

*Evaluated at the sample means of the independent variables.

**Represents sample average of the auto travel time variable.

SOURCE: Charles River Associates, 1976.

evaluated at the sample mean values of the independent variables with a parametric variation of the auto travel times. Auto mode choice probabilities (denoted as p_a in Table D-22) are generally higher for the linear LOS model than for the logarithmic model form. Conversely, for most of the range of auto travel times, the elasticities of auto mode choice with respect to auto travel time (denoted as $\eta_{p_a|t_a}$ in Table D-22) is greater in the logarithmic LOS model. The same pattern may be observed with respect to auto cost variation (see Table D-23). The linear LOS model generally predicts higher auto choice probabilities with a somewhat "flatter" logit response function over most of the range of the variation in travel times and costs found in the estimation sample.

6. EMPIRICAL CHOICE SET FORMATION

Introduction

The Baltimore Disaggregate Data Set (BDDS) included a novel experiment in developing choice sets for disaggregate demand modeling. Most disaggregate demand models estimated to date have relied upon the analyst to assign choice sets to individuals and develop LOS values for alternatives not chosen. In the BDDS detailed trip reports respondents were queried about alternatives to the detailed round trip. LOS data were gathered and reported for alternatives actually chosen during the previous six months. We have termed this approach "empirical choice set formation."

Table D-23

TRAVEL COST ELASTICITIES FOR
TWO ALTERNATIVE MODEL SPECIFICATIONS*

	Log Form		Linear Form	
Auto Travel Cost In Dollars	p_a	$\eta_{p_a\|c_a}$	p_a	$\eta_{p_a\|c_a}$
0.30	0.740	-0.350	0.879	-0.081
0.60	0.528	-0.635	0.788	-0.286
0.80	0.432	-0.765	0.703	-0.534
0.90	0.393	-0.817	0.654	-0.700
1.00	0.360	-0.862	0.602	-0.895
1.070**	0.340	-0.890	0.564	-1.050
1.20	0.306	-0.935	0.491	-1.374
1.30	0.283	-0.965	0.435	-1.651
1.70	0.216	-1.056	0.239	-2.910
2.00	0.181	-1.103	0.138	-3.877

*Evaluated at the sample means of the independent variables.

**Represents the sample average of the auto travel cost variable.

SOURCE: Charles River Associates, 1976.

This section describes our findings on the empirical approach. In general, this approach does not appear to be an efficient way to collect data. The reasons why it is not efficient are illuminating. First, there is comparatively little variability in travel behavior. Second, when travel behavior does vary, it often varies in nonstandard ways (e.g. the alternative to driving is not simply transit but a combination of shared ride and transit). Our discussion of this process starts with a brief description of the BDDS alternatives identification process, followed by a discussion of the findings from the detailed link file and a statement of conclusions.

BDDS Alternative Generation Process

In collecting the BDDS, a one-day trip diary was collected for every household member. One of these household members was randomly selected to be the "Primary Respondent." One trip from the primary respondent's diary was randomly selected for detailed reporting. From this detailed report the interviewer asked questions to generate alternatives to the detailed trip. This process for choice set formation is decribed below.

Trip Selection for Detailed Reporting. Once the interviewer completed the diary with the primary respondent, he selected a one-way trip for detailed reporting. There were

four steps to the trip selection process. First, the interviewer counted the total number of one-way trips by the primary respondent for work/school, shopping, visiting friends, or related purposes. These were referred to as "special" trips (S). Second, he totaled the number of "other" one-way trips (T). (Trips for the purpose of returning to home or work were excluded.) Third, he used a special random number table similar to Table D-24 and selected the appropriate category of trip, S or T.

Detailed Trip Report. The Detailed Trip Report (DTR) investigated this single randomly selected trip in great detail. First, the interviewer asked about all other one-way trips associated with this trip necessary to chronicle a complete "round trip" from home or work (school). Complete round trips were always recorded in the DTR.

The unit of analysis for the DTR was the link. A link was considered any part of a trip wherein one mode was used. When mode changed, one link ended and another began. Every one-way trip contained at least one link and could be composed of many more. For instance, the walk to a bus stop from home would be one link, the ride on the bus another link, the walk after debarking a third link, etc. Because the DTR trip was required to both leave and return to home or work/school, the reported

Table D-24

SAMPLE TRIP SELECTION TABLE

Number of "Other" Trips	Number of Special Trips															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	-	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
1	T	S	S	S	S	S	S	S	S	S	S	S	S	S	T	T
2	T	T	T	T	S	S	S	S	S	S	S	S	S	S	S	S
3	T	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
4	T	T	S	T	S	S	T	S	S	S	S	S	S	S	S	S
5	T	S	S	S	S	T	S	T	S	S	S	T	S	S	S	S
6	T	T	S	S	S	T	T	S	S	S	S	S	S	S	S	S
7	T	S	S	S	S	S	S	S	T	S	T	S	S	S	S	S
8	T	T	T	S	T	S	S	S	S	S	S	S	S	S	S	S
9	T	T	T	T	S	T	S	S	S	T	S	S	S	T	S	S
10	T	T	T	S	S	T	S	S	S	S	S	T	T	S	S	S

The circled cell above (3,3) has a probability of getting a T:

$$P = \frac{3}{(4x3)+3} = \frac{3}{15} = \frac{1}{5} = 0.2$$

Since the entry is an S, we can conclude that the random number which was generated for that cell was greater than 0.2.

SOURCE: Charles River Associates, 1978.

round trip could have been a chain of trips or a complex tour. A trip chain occured whenever one or more purposes were accomplished by stopping at several locations enroute before returning to the home or work origin.

In practice, the trip selection process yielded three basic types of trips for DTR collection:

- The home-based work-/school-trips including all links on the round trip between home and work/school including stops for shopping and other purposes made enroute between home and work/school;
- The work-based trips including all links of a round trip for any purpose starting and ending at work/school; and
- Home-based other trips including all links of a round trip for any nonwork/-school purpose.

For each link, the interviewer determined the origin, mode, route, perceived travel and distance time on the link, destination, and arrival time. If a private automobile was the mode for the link, the interviewer asked about tolls, unparking time (for drivers), waiting time (for passengers), parking time and cost. If it was a shared ride by auto, the interviewer determined how costs were shared. For transit links, the respondent provided data on perceived waiting time, fares, seat availability, and covered shelters at transit stops.

Once data were collected for each of the links, the interviewer asked the respondent if he had carried packages or other articles, and if so, whether a car or taxi was necessary or whether the choice of mode was otherwise affected. If the chain included a shopping trip, the interviewer obtained an expenditure estimate.

Identification of Alternatives. The alternative identification process consisted of a hierarchy of questions about the Detailed Trip designed to identify alternative modes for all links, alternative trip configurations for trip chains, and alternative destinations for non-work-/school-trips. All alternatives that the primary respondent had actually used in the last six months were selected for detailed reporting using the Alternative Trip Report (ATR).

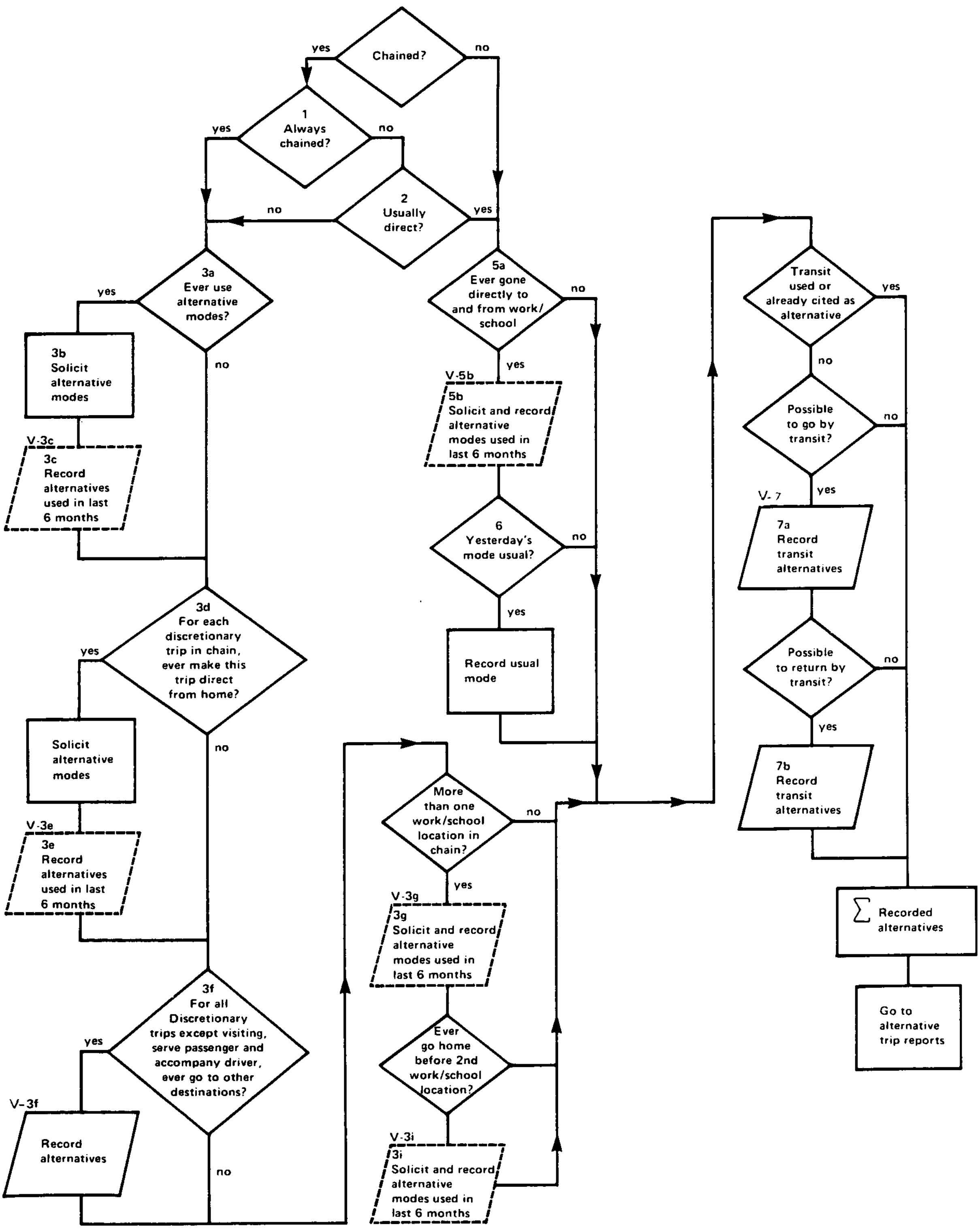
Work-/School-Trips -- The interrogatory sequence for identifying alternatives is shown in Figure D-3. If the DTR trip was a simple home-based work-/school-trip, the interviewer inquired about alternative modes used to make this trip. He recorded all identified alternatives and marked those used during the previous six months (V-5b). If the DTR was part of a chain, the line of questioning was more complex. The interviewer first sought to identify alternatives that maintained the same trip sequence but used alternative modes

(V-3c). For discretionary trips in the chain he asked about the possibility of serving these trip purposes with a simple home-based trip (V-3e). If the chain included more than one work/school location, the interviewer asked about alternative chains using different modes between the two work/school locations (V-3g). He also explored the possibility of traveling to the second work and school location directly from home (V-3i). Each of these different alternatives used during the last six months was marked for the ATR. The interviewer also identified alternative destinations for discretionary links (V-3f) and explored the possibility of transit alternatives for respondents with auto-dominated travel patterns (V-7), but these alternatives could not be selected for the ATR.

Non-Work-/School-Trips -- The interrogatory scheme for non-work-/school-trips is shown in Figure D-4. If the DTR trip was a simple round trip, the interviewer inquired about alternative modes to that destination for that purpose (VI-56). If the DTR trip was part of a chain, the interviewer asked about alternative modes for some or all trips in the chain (VI-2c). He also asked if the destinations had ever been visited in an alternative sequence (VI-2e). He then determined if the DTR destination had ever been visited as a simple

Figure D–3

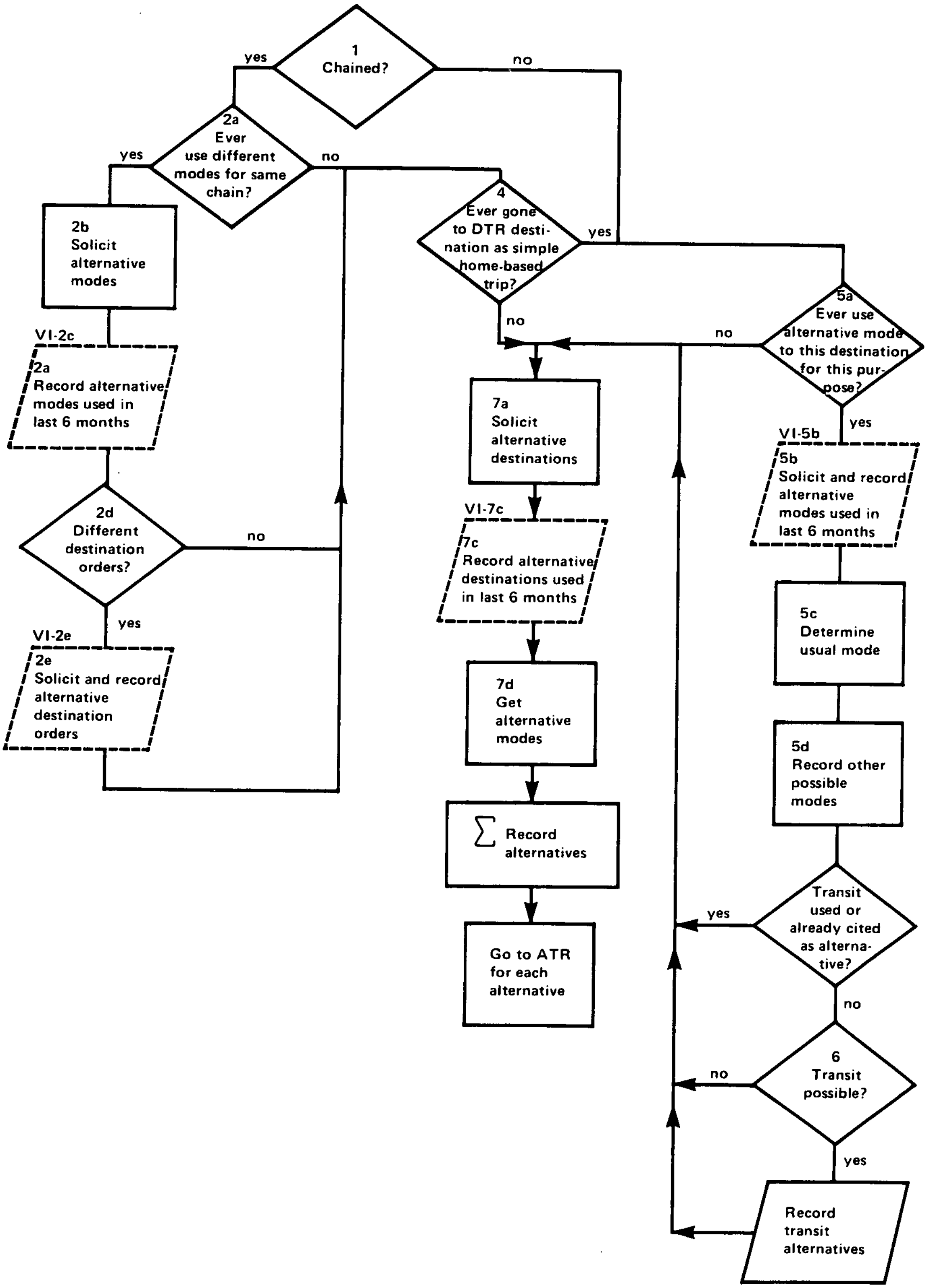
IDENTIFICATION OF ALTERNATIVES FOR HOME-BASED WORK-/SCHOOL-TRIPS (BDDS)



SOURCE: Charles River Associates Incorporated, November 1980.

Figure D—4

IDENTIFICATION OF ALTERNATIVES FOR NON-WORK-/
SCHOOL-TRIPS (BDDS)



SOURCE: Charles River Associates Incorporated, November 1980.

home-based trip. If so, he asked about alternative modes for this simple trip (VI-5b). For both chains and simple tours, he asked about alternative destinations for this trip purpose (VI-7c). He also asked about the mode(s) used for alternative destinations (VI-7d) and possible transit alternatives for persons with auto-dominated travel patterns (VI-6); however, these alternatives could not be selected for the ATR.

Data collected during the process of identifying alternatives using Forms V or VI are stored in the household file. Much of this data was miscoded or otherwise scrambled during transcription and may be permanently lost.

Alternative Trip Reports. A separate ATR form was completed for each alternative trip identified during the alternative identification process. The ATR is virtually identical to the DTR. Complete round trips are reported. The unit of analysis for the ATR is the link.

Summary. Thus, the BDDS detailed and alternative trip reporting process was designed to generate empirically based choice sets for estimation of disaggregate demand models of mode and destination choice. The results of this process as shown below, however, were less than encouraging.

Findings

Table D-25 shows the contents of the detailed link file data on alternative trips. It is noteworthy that of the 966

households interviewed, 135 took no trips on the travel day; of the remaining 831 households, only 577 identified an alternative to the chosen trip. For the 577 households who were identified in the household file as having an alternative to the chosen trip, CRA only found ATRs for 389 households. More than one alternative could be recorded for a single respondent, explaining how 389 households could have 779 trips.

As can be seen from the Table D-25, most of the alternative trips fall into three categories -- alternative modes for simple work- and nonworktrips (or simple alternatives to chains), and alternative destinations for discretionary trips. There are insufficient cases in the data set to model the other types of alternatives.

Table D-26 shows the mode for each of the links in the detailed link file. Most of the links were by auto, bus, or walk. It is not clear that there are sufficient cases with the other modes chosen or as alternatives to enable them to be included in the choice set for mode choice models.

Sources of Bias. The procedures for selecting the primary respondent and the detailed trip and for identifying the alternatives have ramifications for mix of trips and

Table D-25

NUMBER OF ALTERNATIVE TRIPS REPORTED IN THE BDDS BY TYPE AND NUMBER OF LINKS

	Number of Links							
Alternative Types	2	3	4	5	6	7	8	Total
I. Alternative mode for work chain	--	--	2	--	--	--	--	2
II. Alternative mode for work chain -- some links removed	6	--	--	--	--	--	--	6
V. Simple worktrip -- alternative mode*	85	7	19	1	5	--	--	117
VI. Alternative modes for nonwork chain	2	9	2	--	1	--	--	14
VII. Alternative Order of Links in nonwork chain	3	2	--	--	--	--	--	5
VIII. Alternative modes simple nonwork	195	5	12	--	4	1	--	217
IX. Alternative destinations	382	12	15	1	7	--	1	418
								779

This also may be a simple alternative to a complex tour.

SOURCE: Charles River Associates, 1980.

Table D-26

MODE FOR LINKS IN BDDS DETAILED LINK FILE

	Number	Percent
Auto driver	1,736	40.8
Auto passenger	859	20.2
Walk	997	23.4
Bus	505	11.9
Taxi	71	1.7
Bike	49	1.2
School bus	19	0.4
Motorcycle	19	0.4
Boat	2	0.0
TOTAL	4,257	99.9*

*Rounding error.

SOURCE: Charles River Associates, 1980.

alternatives found in the data set. Since a primary respondent was randomly selected first, and only then was a trip randomly selected from his trip summary, trips by persons who made few trips on the travel day were more likely to be selected for detailed reporting than were trips by persons who made many trips.

This can be made clearer by a simple example. Suppose household contains only two members, John and Mary. John only went to market on the travel day whereas Mary went to work and to her welding class. John and Mary have equal probabilities of being selected as primary respondents (0.5). If John is selected as the primary respondent, there is a 100 percent chance his market trip will be selected for detailed reporting. In Mary's case each of her trips has only a 0.50 chance of selection. Consequently, the marginal probability of her trip to welding class being selected for reporting is 0.25 whereas John's market trip has a 0.50 chance of selection at the outset of the household interview. The full ramifications of this selection bias for trips by less frequent travelers in the household has not been explored. We believe it may reduce the representativeness of the sample but does not preclude the estimation of consistent disaggregate demand model parameters.

A comparable selection bias arises from the trip selection process once the primary respondent was chosen. Since an individual one-way trip for any purpose was selected as the basis for compiling a "complete round trip" for detailed reporting, the probability that chains or tours would be selected for detailed reporting was enhanced by a factor equal to the number of links in the chains.

Complex Tours. Related to this problem on the overrepresentation of chained trips is a simplification in the alternative identification process. The alternative identification process broke complex chains into simple alternatives but did not identify complex alternatives to simple trips. Consequently, the data set contains a certain number of "apples and oranges" cases, where the trip taken served two or more purposes but the trip alternative served only one.

Multiple Modes. In order to estimate models of single-purpose round trips, CRA collapsed the links in the detailed and alternative trips into single data vectors representing the entire round trip. The mode and purpose variables were collapsed into dummy variables to preserve information to select specific alternative types. This raised a new problem with the occasional cases where the detailed trip or alternative used a relatively unorthodox mix of modes, such as auto passenger to work and bus back home. Since the state of the art offers little guidance to the modeler in these cases, CRA also eliminated these alternatives from analysis. Similarly, CRA also eliminated alternatives using infrequently used modes such as boat, bicycle, and taxi, since insufficient cases were available to create a data base with these alternatives.

Worktrip Mode Choice. In order to estimate a model of work mode choice with three alternatives, auto, walk, and transit, CRA edited the contents of the detailed link file to develop a data set containing worktrip DTRs and one or more ATRs that met the simple criteria required by the simplified assumptions of traditional travel choice models. These screening criteria included:

- No mode switching within round trips;
- No multiple purpose trips;
- No round trips that do not end at the original origin; and
- One or more alternatives identified to the selected round trip.

With respect to modal alternatives, only one ATR and the DTR were sufficient for inclusion in the final sample. (Strictly speaking, two ATRs would be required but some logit

packages allow for missing values in the choice set.) The final data set that met these criteria contained only 30 usable observations from an original data set of 966 interviews. Clearly, this particular empirical approach to choice set formation is not data-efficient.

Conclusions

Two competing conclusions can be drawn from this experience. The first conclusion would be that clearly the alternatives generation process used in collecting the BDDS was flawed and contributed to the attrition of travel data. The second, more radical conclusion would be that the simplifying assumptions made in most disaggregate demand models are so abstracted from the reality of individual travel patterns that it is difficult to find individuals in the real world whose true choice process conforms to that supposed by modelers. While it is tempting to accept the latter possibility, the explanation of the problems may be found in the unworkability of using the interview process to generate choice sets in the manner employed in the BDDS.

7. DISAGGREGATE DATA SET COLLECTION: LESSONS FROM THE BALTIMORE EXPERIENCE

The Baltimore Disaggregate Data Set was intended to break new ground both in the quality and type of data available for disaggregate demand models. It was designed with the aid of at least 50 active reseachers in the transportation field and benefitted from the advice of some of its most experienced pracititioners (this is more fully described in Section 10 of this appendix). It is not clear that the data set realized its promise. Our detailed familiarity with the BDDS is confined to the detailed link (DETLNK) file. The scope of critical comments therefore will be accordingly limited.

Our comments are divided into two broad areas: issues of <u>quality control</u> and issues of <u>questionnaire design</u>. The first area covers problems of accuracy that have plagued the project and would have created difficulties, even in the absence of design problems. We also include documentation and user accessibility problems under this heading. The second area is substantive and reflects on what we have learned about questionnaire design and content for disaggregate travel data sets.

<u>Quality Control (Editing)</u>

Interviewers executed various field edit procedures as the data were first collected. However, it is evident that the public use tape required a more sophisticated editing procedure to eliminate errors not found in the initial edit. When the data were originally coded, quality control consisted of

checking for extreme values on a variable by variable basis. In addition, CRA performed further checks for within-record or within-household inconsistencies. As a result, while individual variables may all appear to have reasonable distribution, because of joint distribution many pairs of variables could still contain inadmissable combinations. Among the apparent coding and other problems, we have found the following.

- Inconsistent identification (ID) fields are reported in the pointers that link travel and personal records. In 40 percent of all the cases for the household and detailed link files these pointers did not match (this problem was later resolved).
- Inconsistent ID fields are reported in the person file (this problem was later resolved).
- Several individuals are reported as having spent negative amounts of time at a destination, i.e., the time of arrival of the first trip is later than the time of departure of the immediately subsequent trip (many of these problems were resolved).
- In the regular working hours questions, some individuals leave work before they arrive.

- Bus fares are reported for some links that supposedly used an auto mode.
- Many questions were answered when the questionnaire indicated that the questions should not have been asked.
- There were several persons who could not be confidently assigned to households. This would limit researchers' abilities to impute the household composition of any of the households in the data set (this problem was resolved by FHWA).
- Travel times and distances are reported for each link in the DETLNK file; the ratios of these times and distances often implied implausible travel speeds.
- DETLNK also reported perceived values, and of course respondents frequently badly misestimated distance traveled.
- There were duplicate records in the DETLNK file. It appears the the duplicates were intended to be corrections, but both the original and corrected records ended up on the tape. Without access to the questionnaires it was not possible to determine which record was correct.

- Apparently records are missing in the coded files: in the DETLNK file more persons do not return home at the end of trip than seems reasonable given the other information we have.
- Age should not be inconsistent with education, but the data report 12 year old children with graduate degrees.

In some cases the complex relational design of BDDS may have promulgated inconsistency problems. This possibility should be recognized in planning data set collection efforts. Critical checks that may be more difficult to design could be included with help from the agent designing the questionnaire.

Baltimore was chosen as the site for collecting the FHWA disaggregate data set, in part because it had recent high quality systems data for the highway and transit networks. Yet the distribution of some network times and distances are not believable. There are 5-hour bus rides, implied vehicle speeds ranging from 2 to 960 miles per hour, individuals living 99 miles from the nearest highway or transit network node, and so forth. CRA and FHWA worked closely to resolve many of these problems.

Currently, the BDDS is a relatively "clean" and consistent set of travel records. Much of the efforts in the third phase of this project were devoted to straightening out the logical and editing problems of the original BDDS release.

Questionnaire Design

This section will discuss guidelines for future surveys of travel data based on the Baltimore experience.

Sample Attrition. Sample attrition from the alternative trip generation process is the most salient problem in the survey design for disaggregate demand modeling. DTRs were obtained for trips randomly selected from four trip categories. Dividing the sample in this way obviously limits the number of trips available for modeling any one purpose. In addition, only 389 of 966 households reported any alternative trips. This is disappointing. The questionnaire appears to probe systematically for trip alternatives, but ATRs were only completed if the primary respondent reported using an alternative mode or destination to satisfy the purpose of the DTR in the last six months. This limit of six months undoubtedly contributed to sample attrition.* The "other" category includes trips for unusual purposes. Insufficent data were gathered for them to be modeled separately; they cannot be included in data sets for modeling different trip purposes.

*Whether using a longer time period would on balance have produced better alternative data is a design question. Respondents resent being pressed for details that are hard to recall or strike them as having a hypothetical nature. Another related design choice was the decision to base perceived alternatives on actual previous use, but, obviously, reporting of trip detail implies recollection of actual experience.

Complex Tours. The inclusion of chained trips in the data set, discussed above and in previous sections, also contributed to this data attrition problem since the models we are working with deal with single-purpose round trips that are not part of chains. Future attempts to gather data systematically for disaggregate modeling should review with special care the problem of defining and generating alternatives. Some approach other than interviewing for retrospective data may be required.

Fatigue. It also should be recalled that alternative trip reporting came at the end of a very long interview. Interviewer and respondent fatigue may have contributed to the apparently small number of alternatives identified. Fatigue may explain the discrepancy mentioned above, between the 577 households reporting alternative trips according to the household file (as determined earlier in the interview) and the 389 households for which we found ATRs in the DETLNK file. (At some future time, it may be desirable to explore uses of the information about the reported availability of alternatives -- mode, destination, etc. -- for the 188 households for whom we believe, at this time, ATRs were not collected.) ATRs were necessarily completed at the end of the interview. Of course, given the care with which alternatives of all kinds were solicited, we may be confronting a limitation in respondents' capacity to absorb and respond to the concept of a trip alternative under direct questioning.

Multiple LOS Measures. The BDDS aimed to provide objective and perceived time and cost information, in rather full detail, on both actual and alternative trips. With respect to perceived data, the detail was requested to separate in-vehicle travel time (IVTT) from out-of-vehicle travel time (OVTT) and from implicit out-of-pocket costs. However, perceived IVTT is missing from the public use tape. Objective information would allow reseachers to estimate different disaggregate demand models using objective or perceived systems data and compare the difference in parameter estimates. Inclusion of objective data was intended, but some of the objective systems information was lost in the transmission process or conceivably was never acquired.

Confusion. The complexity of the survey instrument and the data it sought probably contributed to the high rate of nonresponse to the alternative trip questions. The instrument contained three separate components for gathering this

information. First, the survey gathered link-by-link information on the actual selected round trip (DTR) including the time and cost dimensions of each link. Second, a series of questions followed, designed to identify alternatives to the DTR. Third, a full alternative trip report required link-by-link information on each eligible alternative.

This careful but possibly confusing design yielded little additional information. Only 18 households answered affirmatively to questions generating alternatives to chained trips. Less than half of all households reported alternative trips at all.

Trips versus Links. As a further suggestion, we note that over 90 percent of the one-way trips in the data set were composed of a single link. It might be more convenient to organize the data on a trip basis, with a separate file for the households having multilinked trips.

Specific Recommendations. The following changes are suggested for future data collection efforts.

- The primary respondent should be drawn from the over 16 age group rather than the over 12 age group.
- Questions about employment status should be drawn from the Current Population Survey of the Bureau of the Census so that the various categories would have definitions compatible with those of the Department of Labor and the Census Bureau.

- The marital status question should be sufficiently detailed and should use categories borrowed from the Census.
- A distinction should be made between earnings (compensation for labor) and income (which includes unearned income and transfers). The earnings figure is necessary to determine a shadow price of time; future instruments should try to determine a person's wage or salary as well as income.
- The coding of the length of residence should be continuous, instead of having the top category be "more than two years."
- Special attention to travel time and cost should have resulted in a follow-up question to extra long times or extra high costs to insure accuracy of outlying observations. For example, for time to park or time to unpark, a follow up asking why the time was so long could have been asked for durations over 20 to 30 minutes.

Conclusions

The Baltimore Disaggregate Data Set was a bold experiment in transportation planning data collection. In the context of an experiment it was a considerable success. Many valuable lessons have been and will be learned from the data source.

However, such an ambitious approach to data gathering is not recommended for the calibration or adjustment of planners' applied choice models. It is simply too complex. Rather, for applied purposes the data should be collected with a short and simple questionnaire. Accurate systems data should be used for LOS measures.

The empirical choice set formation process seems to have been too complicated for respondents. However, it may have yielded more useful data if it had been structured differently. For instance, if it only focused on worktrips and developed ATRs whether or not the alternative was ever chosen or only sought destination alternatives for discretionary travel, the data may have been more robust.

8. ESTIMATION OF THE PITTSBURGH SHOPPING-TRIP MODE CHOICE MODELS

The Shopping Mode Choice Model

The shopping-trip mode choice model represents a conditional probability structure of auto and transit choice. The estimation sample consists of 140 observations randomly drawn from specific travel corridors in the Pittsburgh metropolitan area. As in the worktrip mode split analysis discussed earlier, mode choices were limited to auto driver and bus (with walk access). The LOS data reflected travel conditions for the time of day the shopping trip was actually taken. Since few of the observations represented peak-hour shopping travel to the CBD, the auto LOS reflected generally uncongested travel speeds with minimal parking charges (if any) at the shopping destination. Bus LOS in this sample was generally poorer than in the worktrip estimation sample, since most trips encountered off-peak, longer service headways.

Estimation results for the shopping-trip conditional probability mode choice model are summarized in Table D-27. The model specification includes three LOS variables: (transit) walk access time, auto and transit in-vehicle travel times, and modal costs divided by income. A pure income shift variable, HINC, and a term representing household autos per licensed driver were included as socioeconomic descriptors. Although the model specification of the shopping mode split model is similar to the work mode split model, no deliberate attempt was made to parallel the two.

The parameters of all the variables in the shopping mode split model were of the correct sign and significant at the 95 percent level. The relative magnitudes of the parameters for transit walk (OVTT) and in-vehicle travel time are on the order of eight to one, suggesting that travelers find walk access considerably more onerous than in-vehicle travel time on a shopping journey.

Table D-27

ESTIMATION RESULTS: SHOPPING MODE SPLIT MODEL

$$\ln \frac{P\ (auto)}{P\ (transit)} = \underset{(-4.62)}{-6.63} + \underset{(2.48)}{2.16}\ HINC + \underset{(2.13)}{2.03}\ APERDR \quad \underset{(-3.71)}{-.34}\ OVTT \quad \underset{(-2.02)}{-.04}\ INVTT$$

$$\underset{(-3.54)}{-13.50}\ C/INC$$

NOBS = *140*

ρ^2 = *.71*

L(0) = -97.04   L(θ) = -28.26

Percent Estimated Correctly = 92.86

p(auto) = probability of choosing auto

p(transit) = probability of choosing transit

HINC	*= 1 if household income exceeds $7000/year; 0 otherwise*
APERDR	*= autos per licensed driver*
OVTT	*= difference in (auto-transit) out-of-vehicle travel time (in minutes)*
INVTT	*= difference in (auto-transit) in-vehicle travel time (in minutes)*
C/INC	*= difference in (auto-transit) cost in dollars divided by income code (see Table D-16)*

SOURCE: Charles River Associates, 1976.

As in the work mode split model, the coefficient of the income shift variable was positive, indicating that, all other factors equal, high-income travelers exhibit a preference for auto travel. Auto availability as represented by the term autos per licensed driver (APERDR) also positively influenced auto mode choice.

The model correctly predicted the mode choice of 92.86 percent of the travelers in estimation sample. The overall goodness of fit of the model is good as indicated by the coefficient of determination, ρ^2 equal to 0.71.

9. MODEL ESTIMATION WITH THE TWIN CITIES DISAGGREGATE DATA SET

Introduction

Mode split model development reseach was undertaken using data collected in the Twin Cities (TC) metropolitan area to accomplish the following objectives.

- To assess the feasibility of applying a disaggregate demand modeling framework to metropolitan area household interview survey (HIS) data augmented with aggregate network LOS data. From the standpoint of ensuring the practicality of the disaggregate model approach our experience with the TC data base represents the types of problems that may be encountered in the great number of urbanized areas where little or no data exist that are specifically designed for disaggregate analysis.

- To evaluate the quality and usefulness of HIS data in order to recommend improved procedures for collecting and forecasting disaggregate data.
- To explore further variable selection issues for disaggregate model specification.
- To explore further sample design issues as they affect parameter estimation and model application.
- To investigate the transferability properties of disaggregate demand models.

Research findings on each of these topics will be discussed in turn below.

Feasibility of Using Home Interview Survey Data for Disaggregate Model Estimation

In several respects, the Twin Cities (TC) data base available through the Metropolitan Council was better suited for disaggregate model estimation and analysis than the travel data commonly available in most urban areas. In particular, the TC data base had already augmented the traditional HIS household and trip data with aggregate network LOS information for both highway and transit. Moreover, the TC data set was already partitioned into a subsample of trips for which transit was available. Nonetheless, the following problems were encountered in preparing the data for logit estimation:

- The trip data represented one-way travel, requiring processing to link up round trips;
- The network LOS data represented 24-hour average travel times rather than peak and off-peak travel times;
- LOS data were available only for transit and "highway"; thus, no LOS data were directly coded for carpools, taxis, or other modes; and
- Some household descriptors useful in demand analysis (e.g., number of licensed drivers) were not collected.

Starting with an initial base of 3,640 one-way trips by all modes and purposes, the data were screened and processed, resulting in 721 linked round trips available for logit estimation. The mode and purpose breakdown of the linked trips was presented earlier in the Table D-2. Of the 721 trips, 80.2 percent were by the auto-drive mode, 15.9 percent by the auto-passenger mode, and 3.9 percent by the transit mode. The overall auto-transit split was 693 (96.1 percent) auto and 28 (3.9 percent) bus. The 721 trips included 406 work trips, 111 personal business trips, 8 medical trips, 68 social/recreation trips, 91 shopping trips, and 37 school trips.

One immediate finding here was that for all nonworktrip purposes, there were too few transit trips to allow for disaggregate demand model estimation. The implications of this

finding are two-fold. First, lack of sufficient data in a given metropolitan area for estimating nonwork mode choice models places increased importance on investigating the transferability of disaggregate demand models (from cities with sufficient data to estimate disaggregate demand models for several trip purposes to cities like Minneapolis/St. Paul). Second, the low patronage rates of certain modes for nonwork trip purposes suggest that if models are ever to be estimated to explain this choice behavior, changes in data collection procedures must be undertaken to sample more heavily from areas where transit is relatively heavily used (cluster sampling). It should be stressed that a lack of significant nonauto ridership at present does not obviate the need for mode choice models. Such models are required in order to gain a better understanding of why current services are little used as well as providing a tool to assess the market potential of new transport services.

Quality and Usefulness of Existing HIS Data

Some of the problems with existing HIS data and network LOS data for disaggregate model estimation have already been cited. Experience to date with both the Twin Cities and Pittsburgh data bases have brought to focus several weaknesses with existing disaggregate data. On the basis of our experience, we have prepared a number of recommendations to guide future disaggregate data collection efforts.

The principal recommendations can be summarized in four basic categories.

1. Improvements in the quality of the LOS data to be used in disaggregate model estimation:
   - Use of household-specific rather than zonal average LOS data; and
   - Use of peak and off-peak data rather than some measure of average daily conditions.
2. Improvements in the definition of the feasible choice set for individuals:
   - More precise knowledge of the destinations and modes considered by the traveler at the time of his trip rather than assumptions based on judgment.
3. Improved sampling procedures:
   - Cluster sampling to ensure a diversity of traveler choices.
4. Improved definitions of trip purpose categories and mode alternatives.

Variable Selection Issues

Findings. The Twin Cities data base has been used to estimate several alternative specifications of binary choice (auto/bus) work mode split models. These modeling efforts, in addition to our work on the Pittsburgh mode split models, have

provided further insight into the factors that significantly influence travel behavior. In qualitative terms, the empirical findings suggest the following.

- An expanded range of socioeconomic descriptors may be significant in explaining mode choice behavior. Income, life cycle indicators, sex, and auto ownership have all proved to be significantly explanatory of mode choice.
- Travelers' responses to travel time changes depend on the way in which the time is spent. In terms of increasing value of time, travelers appear to be most sensitive to in-vehicle travel time, waiting time, and walk access time in that order. This finding generally confirms estimation results obtained on the Pittsburgh disaggregate data set. However, in empirical research with the Twin Cities data set, transit wait time proved to be a significant variable in its own right. As might be expected, the Twin Cities estimation results suggest that travelers view increases in transit wait time as being more onerous than increases in transit in-vehicle time.
- The relative valuations of the cost of alternative modes depend on the income of the traveler. As expected, the higher a traveler's income, the less significant is the cost difference between modes in determining his choice

behavior. Here again, this finding was consistent with estimation results on the Pittsburgh work mode choice data set.

Description of the Variables. The types of variables employed in the binary choice mode split model included socioeconomic variables to account for differences in travelers' tastes and LOS variables to measure the relative impedances of the auto and transit modes. In particular, four socioeconomic variables were included to investigate the effects of household life cycle, sex of the traveler, auto availability, and income on mode choice behavior.

The inclusion of *life cycle* variables is based on the hypothesis that households have different underlying travel behavior depending on their present stage of life. Life cycles may be broadly defined in terms of the marital status, age, number of children, and occupations of the household members. Households in differing stages of life may be considered as distinct market segments. There are two reasons why travel behavior may differ according to life cycle status. First, *values and tastes* may differ between life cycle groups. For example, a household composed of elderly retirees would probably value comfort and convenience modal attributes more highly than a household in a younger (age) life cycle. Second,

household travel patterns differ and result in competition for use of the household's autos. For example, in households with preschool children, a nonworking adult household member may not have the opportunity to travel often, thus "freeing up" the household's automobile for use by a primary worker.

We may view worktrip modal choice decisions being made in the context of interdependent *household* travel pattern adjustments (D8). Modal choice decisions are based, to some extent, on the availability of automobiles for individual trip use and on the nature of competing uses for the household's autos by all household drivers. Life cycle segmentation is one means of identifying the effect of any systematic differences in travel patterns by different household types on worktrip modal choice.

A single life cycle indicator was employed in the mode split models estimated with Twin Cities data, a dummy variable whose value was 1 (in the utility function of the *auto* mode) if the household was a young married couple (household head under 45 years of age) with one or more children under five years of age, and 0 otherwise. From our earlier comments on differences in household travel patterns, it was expected that this variable would positively influence the probability of auto choice for worktrips.

A second socioeconomic variable in the model represented the sex of the traveler. As with household life cycle indicators, the sex of a traveler may influence mode choice behavior because of differences in values and tastes and/or differences in the intrahousehold allocation of automobile use among household members.

The number of automobiles relative to household size and composition was entered in the model as the single variable autos per (household) resident. The Twin Cities household interview survey did not collect information on the number of workers or the number of licensed drivers per household.

The fourth socioeconomic variable employed in the worktrip mode split models was household income. It was hypothesized in this appendix that income might influence traveler mode choice in both a pure (auto preference) shift effect and through differences in the value placed on travel times and cost. The presence of a pure income shift effect was not statistically significant in models estimated on the Twin Cities data. However, the effect of income on travelers' valuations of LOS was represented by the variable cost divided by income. This representation is based on the hypothesis that, all other factors equal, a high-income traveler is less sensitive to cost differences between modes than is a low-income traveler.

The LOS variables employed in the model distinguished between three individual components of door-to-door travel time: in-vehicle (linehaul) time, walk access time, and wait time. The latter variable was computed as one half the average bus headway for transit travel and zero for auto travel. The walk access time (for both transit and auto) assigned to each individual represented a measure of zonal average access in the household's residence zone. Similarly, in-vehicle travel times were coded as zone centroid to zone centroid average travel times.

Travel time data were <u>not</u> stratified into peak and off-peak conditions. Rather, the data were representative of average daily conditions. Since most worktrips occur in peak hour periods, the travel time recorded in the Twin Cities data set probably underestimates actual conditions for the majority of travelers.

The travel <u>cost</u> variable recorded round trip bus fares for the transit mode and parking plus operating costs for auto. The cost variables were divided by an income code to reflect income-dependent cost valuation differences.

<u>Presentation of the Model</u>. Table D-28 summarizes the basic form of mode split model specification estimated with the Twin Cities disaggregate data set. The estimation results,

Table D-28

SPECIFICATION OF THE MODE SPLIT LOGIT MODEL FORM ESTIMATED WITH THE TWIN CITIES DATA SET

	Mode-Specific Constant	Life Cycle	Sex	Auto Availability
Variables Appearing in the Auto Utility Function	1	1 if household head is under 45 with a child under 5; 0 otherwise	1 if traveler is male; 0 if traveler is female	autos per household resident
Variables Appearing in the Transit Utility Function	0	0	0	0

	Walk Access Time	Wait Time	In-Vehicle Time	Travel Cost
Variables Appearing in the Auto Utility Function	round trip walk time to and from parking lots	0	round trip auto in-vehicle travel time	round trip auto operating plus parking cost dividend by income
Variables Appearing in the Transit Utility Function	round trip to and from bus stops	round trip bus wait time (computed as 1/2 average headway each route used)	round trip bus in-vehicle travel time	round trip bus fare divided by income

SOURCE: Charles River Associates, 1976.

coefficients, t-statistics, and other goodness-of-fit measures are shown in Table D-29. All the estimated coefficients have the anticipated sign and four of the six are significant at the 90 percent level. The coefficients of the LOS variables (time and cost) are, as expected, all negative, although the resulting value of time measures are somewhat higher than might be expected. The results confirm our hypothesis on the relative valuations of the individual components of door-to-door travel time. The value of (transit) wait time is 50 percent higher than the value of in-vehicle time. Walk access time is valued even higher; three times greater than the value of transit wait time.

The parameter estimates of the socioeconomic variables all conformed in sign to our a priori expectations. Auto availability, as represented by the variable autos per household resident, had a positive and significant (at the 95 percent level) coefficient, indicating that the greater the number of autos relative to household size, the greater the probability of auto use for the worktrip.

The life cycle variable also had a positive parameter estimate. This result corroborates the previously noted hypothesis that in households with preschool children, the intrahousehold competition for use of household auto(s) during

Table D-29

WORKTRIP MODE SPLIT MODEL
ESTIMATION RESULTS:
TWIN CITIES DATA SET

$$\ln \frac{P(auto)}{P(transit)} = \underset{(-1.05)}{-.95} + \underset{(2.12)}{2.23}\ APERR + \underset{(2.02)}{.93}\ SEX + \underset{(1.31)}{.72}\ LCYCLE - \underset{(-1.08)}{.02}\ INVTT$$

$$\underset{(-2.49)}{-.09}\ OVTT - \underset{(-1.00)}{.03}\ WAITT - \underset{(-1.11)}{.64}\ C/INC$$

NOBS = 350

ρ^2 = .68

L(0) = 242.60   *L(θ)* = -78.24

Percent Estimated Correctly = 93.14

P(auto) = probability of choosing auto

P(transit) = probability of choosing transit

APERR = autos per resident

SEX = 1 if male; 0 if female

LCYCLE = 1 if household is young married couple (household head under 45) with child under 5; 0 otherwise

INVTT = difference in (auto-transit) in-vehicle travel time (in minutes)

OVTT = difference in (auto-transit) out-of-vehicle travel time (in minutes)

WAITT = wait time (for transit) in minutes

C/INC = difference in (auto-transit) cost in dollars divided by income code

SOURCE: Charles River Associates, 1976.

the day may be less than in other household types. Consequently, households with children under the age of five exhibited a higher auto preference for work travel.

The sex of traveler variable, defined as 1 if male, 0 if female in the auto utility function, also had a positive and significant coefficient estimate, suggesting that, all other factors equal, males have a higher probability of auto use than females for worktrips. This result may indicate that in multiworker households, the male household member makes more frequent use of the auto for worktrips than does the female.

Sample Design Issues

Our empirical research using the Twin Cities and Pittsburgh disaggregate data sets has raised three issues concerning the effects of sample design on the efficiency and bias properties of resulting logit model parameter estimates. Specifically, the three issues concern the modeling implications of the following:

- Corridor sampling, where the sample is drawn randomly only from corridors where transit service is good relative to the region as a whole;
- Nonrandom sampling, where the probability of all members of a population within a specified area being in the sample is not equal. In this case the mode split in the

estimation sample will not be representative of the area or corridor as a whole; and

- Stratified sampling, where the population is divided into subgroups on the basis of one or more (independent) variables and each group is sampled randomly at different sampling rates.

Each of these issues is discussed below in the context of the empirical research conducted during the project.

Corridor Sampling. The Pittsburgh disaggregate data sample was drawn randomly from a prespecified travel corridor where the quality of transit service was good relative to the region as a whole. This sampling procedure may yield biased parameter estimates in a mode choice model due to the interactions with land-use choices.

Biased parameter estimates may be caused by the presence of significant correlations between included and omitted (unobserved) variables in a choice model. If "transit lovers," whose preferences are explained by factors omitted in the specification of a choice model, tend to locate in areas where good transit service is provided, corridor sampling may indeed lead to biased estimates.

To the extent that a disproportionate number of travelers with these types of (unobserved) transit preferences tend to locate in well serviced transit corridors, mode choice models estimated on data from these corridors may have two related, unfortunate properties.

- The mode-specific constant may exhibit a "transit bias." This estimated constant represents the average effect of omitted factors in the mode split model specification. As such, applying the estimated model to mode split forecasts in other areas where there is not a concentration of "transit lovers," may lead to overestimates of transit patronage.
- Some of the coefficients of the LOS variables may be biased. This will occur if an unobserved attribute, e.g., preference for transit because of personal tastes, is correlated with an observed attribute, e.g., travel time. In this example, the resulting travel time coefficient would be overestimated.

Nonrandom Sampling. This type of procedure involves sampling on the dependent choice variable. (Manski and Lerman (D9) refer to this procedure as choice-based sampling.) An example of nonrandom sampling would be the collection of disaggregate data from a prespecified number of on-board transit surveys and roadside (auto) interviews. In such cases,

the relative shares of transit and auto users in the sample would not be representative of the modal split in the region as a whole.

In general, nonrandom samples will produce parameter estimates in a logit choice model that are different from parameters estimated on a random sample. The differences in parameter estimates will probably be most pronounced for mode-specific constant (MSC) terms and any other variables highly correlated with the MSC(s). (Manski and Lerman (D9) report that McFadden has proved that MSC terms are inconsistent in logit models with a full set of MSC(s). Therefore, in large samples, differences in parameter estimates are essentially confined to MSC terms.)

The Twin Cities data base was employed to investigate the importance of nonrandom sampling on the differences in the resulting model parameter estimates. Table D-30 displays the parameter estimates of a binary work mode split model estimated on three alternatively constructed samples from the Twin Cities data. The definitions of variables and functional form of these models are the same (except for the omission of the life cycle and transit wait time variables) as described earlier in Table D-28. The first model was estimated on the entire random sample of 350 work trips (of which 25 trips were by transit).

Table D-30

MODE SPLIT MODEL PARAMETER ESTIMATES
FOR RANDOMLY AND NONRANDOMLY DRAWN SAMPLES
(Standard Errors in Parentheses)

	Model 1	Model 2	Model 3
Number of transit users	25	25	25
Number of auto drivers	325	150	75
Mode-specific constant term	-0.42 (0.79)	-1.26 (0.88)	-2.87 (1.18)
Sex	1.03 (0.45)	1.08 (0.52)	1.84 (0.65)
Autos per resident	1.54 (0.84)	1.93 (0.95)	2.50 (1.10)
In-vehicle travel time	-0.014 (0.01)	-0.012 (0.01)	-0.012 (0.013)
(Out of pocket cost)/income	-0.737 (0.55)	0.835 (0.79)	-0.484 (0.91)
Wait time	-0.038 (0.03)	-0.025 (0.03)	-0.038 (0.04)
Access time	-0.088 (0.03)	-0.088 (0.03)	-0.089 (0.03)

SOURCE: Charles River Associates, 1976.

(The sample was drawn only from areas where transit was available. Nonetheless, within these corridors the sampling was random.) The second and third models retain all 25 transit trips from the larger sample, but include only a subset of the total number of auto users (see Table D-30). As such, the two latter samples are nonrandom; our procedure has "artificially enriched" the data with transit users.

In comparing the parameter estimates from the three models, as expected, the MSC term becomes increasingly negative (exhibiting transit bias) as the proportion of transit users in the estimation samples increases. It also may be noted that the parameter estimates of the LOS variables are relatively insensitive to the mode choice composition of the estimation sample. This may indicate that in the Twin Cities data base, LOS coefficients are more transferable than the MSC.

Stratified Sampling. One of the problems encountered in estimating models with the Twin Cities data set was that few of the observations represented "traders" -- travelers whose travel cost by one mode was higher and travel time lower than the competing mode (or vice versa). For most of the observations, auto was the dominant mode in all LOS characteristics. This partly explains the reason for the relatively low t-statistics on the coefficients of the LOS variables shown in Table D-29.

The effect of a limited variation of the independent (LOS) variables in the estimation sample on the goodness of fit of the estimated logit model is shown schematically in Figure D-5. As noted above, for most of the observations in the Twin Cities data set, the LOS data were "clustered" in a region where the auto mode was favorable (cheaper and faster) to transit. This is indicated in Figure D-5 by the circled cluster of data points representing a particular LOS variable (time or cost). As shown, when the independent variables and choices (93 percent of the travelers in the sample chose auto) have limited variation, several different logit response curves may "fit" the data relatively closely.

These comments suggest that future data collection efforts should consider a cluster sampling technique wherein a wide variation in the independent LOS variables would be ensured. As shown in Figure D-5 (by the data clusters in the boxes), having data over a wider range of the independent variables may serve to provide a better estimate of the logit response function.

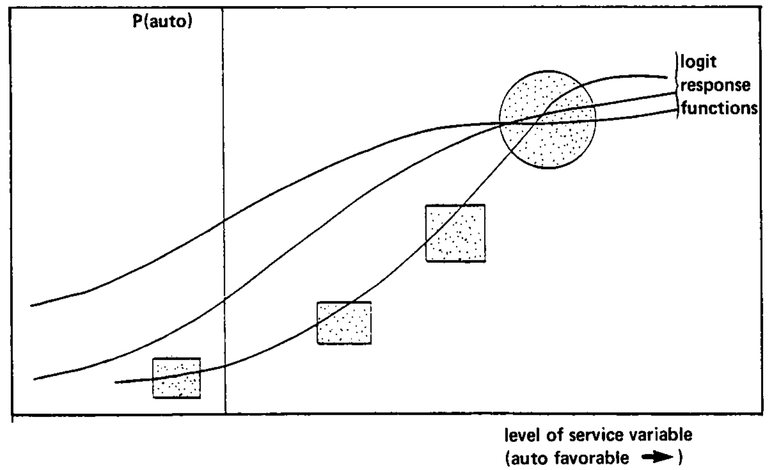
10. MODELING WITH THE BALTIMORE DISAGGREGATE DATA SET

Introduction

This section describes disaggregate modeling research for the Baltimore Disaggregate Data Set (BDDS). The discussion is

Figure D-5

SCHEMATIC REPRESENTATION OF THE EFFECT OF LIMITED VARIATION IN THE INDEPENDENT VARIABLES ON GOODNESS OF FIT OF THE LOGIT MODEL



broken down into several subsections. First, it discusses the data interests of disaggregate modelers that prompted the data collection effort. In the second section computer data files are described, followed by a discussion of how the data files were reorganized to estimate mode choice models. A conceptual model of mode choice is presented in the fourth section, followed by a discussion of model specification. The final section presents empirical results.

This section has two objectives: first, to introduce researchers to the BDDS as a tool in disaggregate modeling research; and second, to present the results of CRA's demand modeling work with the BDDS.

The BDDS was collected during May and June of 1977. It describes 966 households and the travel of all responsible members of these households for a 24-hour period. The data set consists of six separate files tied by common identification numbers. The separate files describe the households, their vehicles, the household members, trip summaries for all household trips, detailed links (including alternatives) for a randomly selected trip, and the land use in each Baltimore

transportation zone. This set of data files affords many opportunities to model choice behavior of interest to transportation analysts and planners.

Data collection and file development was sponsored by the U.S. Department of Transportation's Federal Highway Administration. The research objective was to interview 1000 households and obtain one-day summary records of all trips taken in a 24-hour period for all household members over 12 years of age located in these households. In addition, one individual 16 years of age or over was selected at random as the "primary respondent." Trip details by link were obtained for a single round trip selected at random from a single individual's trip summary. Alternatives to this detailed trip were identified by asking the primary respondent about his alternative travel for this purpose in the last six months.

Data Interests of Disaggregate Modelers

At the outset of developing the disaggregate data set, CRA undertook a major effort to solicit recommendations on the household survey from researchers in the field of disaggregate travel demand modeling to ensure that the data set would reflect the needs of a variety of researchers with different approaches and hypotheses to test.

It is convenient to classify the results of the investigation into six categories: socioeconomic; neighborhood and housing; auto ownership; basic trip record; alternative trip record; and nonworktrips.

Socioeconomic Data. Most emphasized the need for complete data on the household's life cycle stage. It was agreed by most respondents that data should be collected for all household members. In general, most researchers seemed to be interested in collecting the usual socioeconomic data already known to be useful in modeling travel behavior.

Neighborhood and Housing Descriptors. Many researchers did not consider detailed housing and neighborhood descriptors to be of much value in modeling travel behavior, but most researchers attached some importance to structure type (e.g., one-family, duplex, etc.). Since few neighborhood and housing descriptors were included in the set, it should not be much more useful than existing data sources are to those economists and planners using transportation data sets to model housing demand.

Auto Ownership. Most indicated a strong interest in knowing the number and types of vehicles owned and/or available to the household to develop measures of auto availability. A majority indicated interest in having information on the household's home parking situation, particularly the type of parking (garage, street, lot, etc.) and distance to the parking location.

Basic Trip Record. Most researchers agreed on a core of important items for inclusion in the basic trip record

including trip purpose and times of departure and arrival. For each link, items included mode of travel, route, and activity at start and end of link. Perceived travel time and cost data were considered important.

Alternative Trip Record. The most consistent comment made by researchers about alternative trip data was the difficulty of collection. Nonetheless, many indicated an interest in collecting it. Some researchers feared that asking the traveler for his perceptions of travel times and costs for all of the combinations of possible alternatives for a given discretionary trip (mode, destination, time of day, route) would make the survey instrument unmanageable. They suggested a compromise of asking the traveler what the alternatives were and then computing the LOS data from a system network. Some researchers also felt that the current state of the art in modeling discretionary trips was not sufficiently developed to warrant the collection of alternatives for discretionary trips. As was described in Section 6 of this Appendix, many of these fears were realized.

Nonworktrip Purpose. Researchers were asked to indicate which nonworktrip types were most interesting to them. Three types of nonworktrips were mentioned most often: grocery shopping, other shopping, and recreational.

A number of respondents made the point that the existence of multipurpose trips make it difficult to concentrate on selected trip types. For instance, a journey from home to a friend's house, then on to a department store, then over to pick up the children from school, and back home again would be difficult to classify for trip purpose; yet this pattern is not uncommon. Attempting to separate this type of trip into its component parts and considering each part as a distance trip may well distort the time and character of the travel behavior.

The Data Files

The home interview and systems data collected by the survey effort are located in six separate computer files. The files are organized to efficiently utilize storage space; there are a minimum of blank records. The files are not organized to be immediately loaded into a logit package and to estimate coefficients. There are pointers in each file that allow the user to reference observations across files. This section will briefly describe each of the data files. It will then describe how the files were merged to create a file for estimating mode choice models. For more detail on each file, the reader is referred to the BDDS User's Guide (D10).

The Household File. This file is the hub file of the data set. In FHWA/CRA clean-up efforts, not all this data was screened for quality. It contains each household's census tract and block as well as the sampling strata, household number, and sample weight as calculated by CRA for the entire sample. It contains the number of persons in the household, the total number of travelers and the total number of persons eligible as respondents. Also, it contains the number of the primary respondent. The household file also contains the data on distances to the nearest public transit, arterial, and expressway; home tenure; household income; dwelling type; and number of rooms.

Some questions from the Detailed Trip Report were asked only once of the primary respondent -- Were packages carried? Tools? If so, did they make a car or taxi necessary? How much was spent on shopping trips? These responses are provided in the household file.

The remaining variables in the household file are the responses to the alternative trip identification process. These data were not screened in clean-up efforts.

The Vehicle File. The vehicle file data on each vehicle include: vehicle type, Wharton model code, year of manufacture, ownership, seating capacity, regular commuting use, regular driver(s), parking place and cost, and conditions of availability for employer-owned cars.

The Person File. This file contains data on each person in the household including the person number, age, and sex. The file also contains the number of one-way trips made by each person and four variables describing physical handicaps, if any.

The remaining variables in the file include employment status, occupation, industry, wage earner or salaried, work hours, and work time flexibility. The file also includes length of residence at present address, former residence, possession of drivers license, when person last drove, person income, educational attainment, marital status, race, and fluency in English.

The Trip File. This data file contains the household number, the person number, and the trip number. It also contains the trip type, origin and destination, trip purpose, mode, and household members making the trip. For automobile trips it contains the number of persons in the vehicle. The file contains the departure and arrival times, a frequency estimate, and the land use of the destination. From the coded network, the file contains centroid-to-centroid travel time and distance by chosen mode. It also contains highway and transit zone-to-zone network data for each trip recently appended by FHWA.

The Detailed Link File. The file contains a variable uniquely identifying each link by household, alternative type, and link within the alternative. There is also a variable indicating the question that generated the alternative trip. It will be recalled that there are nine different types of alternative trips identified in the BDDS.

For work-/school-trip:

1. Alternative modes for chains -- all discretionary trips retained in chain;
2. Make discretionary trips(s) in chain as separate home-based trip(s);
3. Alternative mode for trip from one work/school location to a second work/school location;
4. Make trip to second work/school location in chain as a separate home-based trip; and
5. Make home-based work-/school-trip from DTR by alternative mode or make work-/school-trip of chain as a separate home-based trip.

For non-work-/school-trips:

6. Alternative modes for chains -- all trips retained in chain, same order;
7. Alternative chain configuration and modes -- all trips retained in chain;

8. Make home-based trip from DTR by alternative mode or make the DTR trip of chain as a separate home-based trip; and
9. Alternative destination.

The remaining variables describe the link. For automobile links they include the number of occupants on alternative links, tolls, unparking and parking time, parking type and cost, and cost sharing data. For transit alternatives there is a wait time estimate, fare, shelter, and seat availability. For all alternatives there is perceived travel time and distance, the origin and destination census tract, block and traffic analysis zone, as well as systems data describing transit wait time, trip distance, and travel time. The file contains other zone-to-zone network data appended recently by FHWA. This recently added data is at the trip level of detail rather than the link level.

Zone File. The sixth data file includes information on population, land use, and employment in each of the transportation zones in the Baltimore planning network.

Data Manipulation

It should be apparent to the reader from the discussion above that the public use tape of the Baltimore Disaggregate Data Set is not a single data file from which choice models can be simply estimated with a minimum of preprocessing. Rather,

extensive data manipulation and analyses were required. This section briefly describes the edits and merges that were required to develolp two mode choice data sets from the detailed link file.

It will be recalled that the detailed link file (DETLINK) contains link-level data on a single round trip (DTR) and data on up to five different classes of alternatives to that trip (ATRS) for worktrips (four alternative types for nonworktrips). It also will be recalled that many DTRs did not have coresponding ATRs; many DTRs were complex tours and some DTRs and ATRs used combinations of modes (e.g., bus-out, walk-back) or exotic or rare modes (e.g., horse, boat, bike). Given the state of the art and the configuration of the data, it was necessary to identify usable trips (e.g., none of the complications mentioned above) and collapse those link records into round trips.

There were only sufficient observations available in DETLINK file to investigate three dimensions of choice: work-/school-trip mode choice, discretionary trip mode choice, and discretionary trip destination choice. Very few alternatives were identified by the other six alternative identification schemes.

First Approach. Using the DETLINK file as a base, CRA compressed link-level data to develop summary round trip information. The new file, INDEX, is organized by household and round trip: household information, detailed trip information, alternative trip information (for up to seven alternatives), and a single code to identify whether the observations included an alternative trip. The information on detailed and alternative trips (up to seven ATRs) consisted of codes identifying:

- The last mode used by the tripmaker;
- The number of one-way trips in each round trip record;
- The number of mode switches within one-way trips;
- The number of mode switches between one-way trips;
- The number of links in each round trip record;
- The occurrence of a trip that did not return to the original origin;
- The occurrence of an alternative round trip that did not originate at the same location as the detailed round trip (ATRS only); and
- The question number that identifies the alternative trip type (ATRS only).*

---

*Alternative trip records only.

The INDEX file was processed to create a new file, INDEX.EDITED. It contained only observations with at least one alternative trip. Cases that met the following criteria were excluded from further analysis:

- No alternative trip;
- Detailed trip mode was not auto driver, auto passenger, public transit or walk;
- All the alternative trips were by modes other than auto driver, auto passengers, public transit or walk; and
- The alternative type was ineligible (alternative types: 1, 2, 3, 4, 6, or 7 -- see Detailed Link File description in previous section).

The INDEX file contained 833 records, which fell into the following categories:

- No alternative trip (N=444)
- Detailed trip mode not applicable (N=21);
- All alternative trip modes not applicable (N=8);
- Ineligible alternative type (N=2); and
- Retained and copied to INDEX.EDITED (N=358).

The 358 observations in the new file were not trouble-free. However, the new file included all observations with at least one tractable alternative to a work-/school- or nonworktrip. Many of these observations had one or more of the following troublesome characteristics:

- Trips that do not return to the initial origin or one or more alternatives does not have the same origin as the detailed record;
- Chained trips;
- Trip records with mode switching within one-way trips; and
- Trip records with mode switching between one-way trips.

A code was developed and appended to each observation in the INDEX.EDITED file which identified the absence or occurrence of each condition listed above. The joint frequencies for these conditions are presented in Table D-31.

On the other hand, records with the above problems are not necessarily useless for mode choice models. For instance, the problem may occur in only one of several alternative trips, with remaining alternatives suitable for modeling. Therefore, CRA identified these observations with a suitable DTR and at least one simple alternative trip. Such records would constitute the households to be used in modeling.

First, CRA determined which of these cases are flawed due to irregularities in the DTR and eliminated them from the data set for mode choice modeling. Second, CRA identified those records with a problem condition in each of the alternative trip reports (ATRS). Trips with no useful ATRS cannot be used.

Table D-31

PROBLEMS WITH DETAILED AND ALTERNATIVE TRIPS
BY TRIP TYPE: DETAILED TRIP FILE (BDDS)

	Trip Type				
Problem	Work/Shopping Mode Alternative	Nonwork Mode Alternative	Nonwork Mode and Destination Alternative	Nonwork Destination Alternative	Total
Simple trips	38	22	44	63	167
Between trip mode switch	19	6	3	3	31
Within trip mode switch	5			1	6
Chained trip	12	20	34	39	105
Does not return home	1	--	1	1	3
Between and within trip mode switch	--	--	2	1	3
Chained trip and between trip mode switch	6	9	4	12	31
Chained trip and within trip mode switch	1	--	1	--	2
Does not return home and chained	--	--	1	2	3

Table continued on following page.

Table D-31

PROBLEMS WITH DETAILED AND ALTERNATIVE TRIPS
BY TRIP TYPE: DETAILED TRIP FILE (BODS)

Problem	Trip Type: Work/Shopping Mode Alternative	Nonwork Mode Alternative	Nonwork Mode and Destination Alternative	Nonwork Destination Alternative	Total
Chained trip, within and between trip mode switch	4	1	1	--	6
Does not return, chained, between trip mode switch	--	--	--	1	1
TOTAL	86	58	91	123	358

SOURCE: Charles River Associates, 1980.

Salvageable observations were those with a simple detailed trip and at least one alternative trip with no complicating factors.

In the final analysis, only 30 cases had adequate data and sufficient alternatives available to afford a model of worktrip mode choice with auto, transit, and walk alternatives. No meaningful results could be generated from such a small data set. Consequently, CRA used an alternative approach to constructing alternative LOS data for simple worktrip DTRs. This approach is described below.

The Alternative Approach. The second data set was created using all simple work/school detailed trip records regardless of whether there was an alternative mode ATR in the DETLINK file. The second data set initially contained 229 observations. After correcting for missing data points, 175 observations remained in the data set. LOS data were not available for modal alternatives, so CRA generated the LOS data for appended FHWA network skims. All LOS data are at the zone-to-zone level of detail.

CRA calculated LOS for three modes: auto, transit, and walk.

When walk was the chosen mode, the LOS data were hand coded by FHWA from a map. When walk was not the chosen mode, CRA assumed that the coded highway zonal distance approximated the walk distance. All walk times were calculated from these distances assuming an average constant walk speed of three miles per hour.

CRA combined auto driver and auto passenger into one uniform auto mode. The generated data would not reflect a difference in LOS between auto driver and auto passenger.

The LOS data calculated for each mode are distance, cost, and travel time. Other LOS data included out-of-vehicle time for transit, a variable for whether the trip was too short to use transit, and whether transit was readily available or required automobile access.

Other variables included in the data set are household person identification numbers, mode of trip, and the dependent variable representing mode choice. The socioeconomic variables are described in the following section.

The socioeconomic variables were drawn from both the household file and the person file. The data items included employment related variables, length of residence at present

address, previous residence, demographic data such as age, race, marital status, educational attainment, whether the respondent had a current driver's license, when he last drove, both household and personal incomes, the existence of physical disabilities affecting walking, driving, or use of public transportation, tripmaking behavior, residence data (dwelling type, own/rent, etc.), number of people residing in the house, traveling in the recording day, the number of vehicles in the household, trip purpose, whether the respondent carried any items that affected the travel method, whether transit is a possible alternative, and if not would person consider driving to a transit line.

Using this data, CRA was able to estimate models of worktrip mode choice. CRA's preliminary results with these modeling efforts are described below.

Conceptual Model

There are several considerations that theoretically influence an individual's choice among automobile, bus, and walk as modes for worktrips. Each of these considerations is briefly described below.

Level of Service. The automobile, bus, and walking modes are very different ways of getting to and from work. The automobile offers the greatest level of comfort, door-to-door service, and great flexibility and versatility for changing travel plans. It is also relatively costly, with a substantial initial fixed cost and moderately high operating costs. There also may be costs associated with storing the automobile at origins and destinations when not in use. Motorists cannot always park at their doorstep. Consequently, there may be access time costs associated with getting to the car. In relation to other modal alternatives, the auto is often the speediest way of getting to work as well as the most comfortable. Finally, the automobile allows travelers to carry substantially more baggage than would be possible with other modes.

The bus is an automotive mode for pedestrians. In relation to the automobile, it offers less comfort and convenience at a lower total cost. Unlike the auto, the transit patron does not have to cover the costs of storing the vehicle at home and work and usually only pays a portion of the total costs of providing service. On the other hand, there can be substantial time spent waiting for transit and transferring between vehicles and routes. Once in the vehicle, the transit

bus is seldom as speedy as an auto. Transit routes are often less direct than the route an individual would take with the car. Transit service is seldom door-to-door, so there are often substantial access times to transit from the ultimate origin and destination. It is generally difficult to carry more than a minimum of baggage on public transit.

Walking is the oldest form of human locomotion. For short distances it is often the easiest and most convenient way to travel. However, as travel distances increase, walking becomes too slow and tiring to effectively compete with mechanized modes. Even small amounts of baggage can be very burdensome on a long walk.

Household Characteristics. Several characteristics of the traveler's household may affect his mode to work. The overall level of wealth and/or household income would influence the traveler's evaluation of time spent traveling. It also would have an indirect impact on his travel options. On average, wealthier homes would have more and better automobiles available. Competition in the household for automobiles would increase with ratio of licensed drivers to autos available.

Individual Characteristics. There are several characteristics of the individual that help determine his choice of auto, transit, and walk to work. With respect to automobile travel, a driver's license is helpful, but not essential, because one can be driven by a friend or family member. Personal income can also determine one's access to a household vehicle. We would expect "breadwinners" to fare well in competition for scarce family cars and for teenagers with part-time jobs to be more likely to walk to work. Work conditions can also determine mode choice. With respect to transit, other personal characteristics may also influence mode choice. These may include attitudinal factors, i.e., that transit is not a practical alternative. They may also include the overall level of individual mobility. Transit may not be practical for travelers who make many trips during the day.

In considering the pedestrian alternative, age and physical condition are obvious considerations. Younger, more vigorous persons should find walking less onerous. Persons who work and live in the central city would also be more likely to walk since the costs of travel by other modes would be more unfavorable.

Model Specification

There are variables in the BDDS estimation data set that measure each of the factors described above. These variables are briefly described below, along with a discussion of how each variable should be specified in the representative utility function for each alternative in the logit model.

Automobile Level of Service. Out-of-pocket costs for auto travel are measured by zone-to-zone automobile travel distance from the 1977 coded Baltimore zonal network times a factor representing average city mileage per gallon of the 1977 Baltimore fleet (11.98 mpg) and the 1977 average Baltimore fuel cost (62.5 cents per gallon).

Travel time is measured for in-vehicle time from the 1977 highway network zone-to-zone travel times.

Auto access time and parking costs are discussed under household characteristics.

Other factors affecting the utility of the automobiles for worktrips are unobserved in the estimation data set. These include comfort, convenience, and capital costs. The impact of these factors is absorbed in a mode-specific constant term.

Transit Level of Service. Out-of-pocket costs for a transit trip are measured by transit fare derived from the 1977 zonal network.

Travel time for transit can be broken into two components: in-vehicle travel time (IVTT) and out-of-vehicle travel time (OVTT), both of which are derived from the 1977 network.

Distance to the nearest transit stop was measured in the BDDS using respondents' perceptions. This perceived distance can be included in model specification as an alternative measure of OVTT.

A mode-specific constant is included to pick up the effects of unobserved attributes of transit such as comfort and convenience.

Pedestrian Level of Service. The financial costs of the walking are virtually zero and are entered as such into the evaluation of the cost coefficient for the three modes.

Walking distance was measured from maps for chosen alternatives and imputed from network highway distances for rejected alternatives. An average speed of three miles per hour was assumed for all pedestrian trips to measure walking travel time. Walk time is included in the utility function for the pedestrian alternative as IVTT. (An alternative specification might include walk time as equivalent to transit OVTT.)

Household Characteristics. Household wealth can be measured by a proxy. This proxy is the number of rooms in the household. Household income is a measure of wealth and current liquid household assets.

The ratio of total household vehicles to total household travelers represents the competition for automobiles within the household. One would expect that as this ratio increased, the

probability that a walker would use an automobile for his worktrip would also increase.

Auto access time and parking costs can be measured by a proxy measuring residential densities. This proxy, dwelling type, describes the household on a continuum from single family home to high rise apartment. One would expect home-end costs of parking and automobile access to increase with higher residential densities.

Individual Characteristics. Variables in the estimation data set measure all the theoretically pertinent characteristics that determine an individual's evaluation of his worktrip mode alternatives. A driver's license dummy variable is included in the data set to partially reflect availability of the automobile mode. The ratio of personal income to household income should reflect individual status in competition for a limited fleet of household autos. As this ratio increases one would expect the individual's priority in auto competition also to increase.

Respondents were asked if they carried any packages or tools that influence their choice of mode. This baggage variable helps determine mode choice.

Workers who make many trips during the day may attach a higher utility to the automobile in order to keep their busy schedule. The number of trips in the respondents' day can be included in model specifications to measure the impact of frequent travel on worktrip mode choice.

Respondents were asked if transit was a possible alternative for their worktrip. Transit must be perceived as feasible to be used. We would expect negative responses to this question to have a large negative impact on the individual's utility for transit.

Age would have a profound effect on the evaluation of the walking alternative. One would expect older respondents to have a greater disutility for walking.

Summary. The specified model for worktrip mode choice (automobile, bus, and walk) includes the following LOS variables that are specified generically: (estimated coefficients do not vary among alternatives) cost and in-vehicle travel time. Transit out-of-vehicle travel time and distance to the nearest transit stop are only entered in the transit utility function.

A large number of households and individual characteristics are included as socioeconomic variables. The values of these variables do not vary across alternatives. Consequently, separate coefficients are estimated for automobile and transit. (No socioeconomic coefficients are

estimated for walk. It is the base case. The auto and bus coefficients represent variations from the base case.) The socioeconomic variables include: rooms in the home (wealth), household income, possession of drivers license, personal income as a proportion of household income, whether baggage is carried, total trips on travel day, work schedule flexibility, perception of transit as a practical mode, dwelling type, and age.

Age is specified as a polynomial since it is not imagined that there is a direct linear relation between age and the utility of different modes such as walking and transit. Rather, in the case of walking, for instance, one would expect the utility of walking to be relatively high over a large age span and then decrease markedly with the onset of old age. Age is thus entered twice in the utility equations; once as a simple age value and once squared to allow for a parabolic function.

Findings

The results of the worktrip mode choice model are shown in Table D-32. The model was estimated with 175 worktrips (120 automobile, 28 transit, and 27 walk). The model properly predicts mode choice for 91.4 percent for all observations.

With respect to the estimated LOS coefficients, the following observations should be noted. First, the cost

Table D-32

BDDS WORKTRIP MODE CHOICE MODEL PRELIMINARY ESTIMATION RESULTS
(T-statistics in Parentheses)

V_A = 16.43 - 0.09 AUTOIVTT + 9.07 AGASCOST + 0.56 ROOMS
(1.13)(-3.48) (3.09) (0.99)

+ 4.92 DRIVLIC - 0.22 PIPERHHI + 9.09 CARSPERT
(1.04) (-0.05) (2.15)

- 0.30 HHINC - 2.51 AGE + 0.05 AGESQ
(1.42) (1.87) (2.12)

V_T = 15.44 - 0.09 TRANIVTT + 9.07 TRANFARE + 14.46 MIPEROVTT
(1.07) (3.48) (3.09) (3.64)

+ 0.26 ROOMS + 3.35 DRIVLIC - 0.64 PIPERHHI
(0.45) (0.72) (0.14)

7.41 CARSPERT - 0.33 HHINC - 2.61 AGE
(1.73) (1.55) (1.95)

+ 0.05 AGESQ
(2.12)

V_W = 0.09 WALKTIME
(3.48)

NOBS = 175

ρ^2 = 0.81

$\rho_c^{\ 2}$ = 0.76

L(0) = 177.61 L(C) = 135.81 L($\hat{\theta}$) = 33.23

Percent Estimated Correctly = 91.4

Notes and source on following page.

Table D-32 (Continued)

BDDS WORKTRIP MODE CHOICE MODEL PRELIMINARY ESTIMATION RESULTS
(T-statistics in Parentheses)

NOTES:

V_A = Estimated representative utility of auto;

V_T = Estimated representative utility of transit;

V_W = Estimated representative utility of walk;

AGASCST = Automobile fuel cost (in dollars);

TRANFARE = Transit fare (in dollars);

AUTOIVTT = Automobile in-vehicle travel time (in minutes);

TRANIVTT = Transit in-vehicle travel time (in minutes);

MIPEROVTT = Ratio of travel distance (in miles) to transit out-of-vehicle travel time (in minutes);

WALKTIME = Walk time (in minutes);

ROOMS = Number of rooms in the household;

HHINC = Household income (in thousands of dollars);

DRIVLIC = Drivers license dummy (1 = yes);

PIPERHHI = Ratio of personal income to household income;

CARSPERT = Ratio of household automobiles to household travelers;

AGE = Respondent's age (in years); and

AGESQ = Respondent's age squared.

SOURCE: Charles River Associates, 1980.

coefficient has a counterintuitive positive sign. We hypothesize that this anomaly results from the fact that the generally inferior alternative, walk, is also costless. Second, the coefficient for in-vehicle travel time is negative and significant as would be expected. A third LOS term was included in the transit utility function representing the ratio of total travel distance to out-of-vehicle time (walking, waiting, and transferring). One would expect that a traveler's evaluation of the transit alternative would improve as this ratio grew larger. The estimated coefficient for this term is, in fact, large and significant.

With respect to household variables, wealth, as measured by the number of household rooms, is positively related to the utility of motorized modes, but not significantly so. The estimated coefficient for income has the theoretically wrong sign. It is negative, but not significant. This could represent a collinearity problem that should be investigated in later specifications and estimations. The ratio of household autos to travelers, CARSPERT, is an auto availability measure that is both positive and significant in the automobile utility function.

With respect to personal attributes, the total number of daily trips and carriage of baggage were excluded from the

model since they contributed very little to its explanatory power. Retained variables include a drivers license dummy and personal income as a proportion of household income. The drivers license variable was positive but not significant in the automobile utility function. This may reflect the fact that the automobile alternative includes auto passengers.

The ratio of personal to household income represents a power variable in competition for household transportation resources. The estimated coefficient for this variable has the theoretically wrong sign in the automobile utility function and does not differ significantly from zero. It probably should be dropped from later specifications.

The coefficients for the age variables are significant for both the automobile and transit alternatives. They are very similar in both these utility functions. This suggests that age should be more properly included in only the walk utility function. This will be tried in later specifications. Analysis of the age/age-squared expression indicates a parabolic function with a minumum in the neighborhood of age 25. This indicates that, all other things held constant, the utility of walk as a worktrip mode peaks at around age 25 when workers are still young and vigorous.

Alternative OVTT Specification. Table D-33 reports results of an alternative specification for walk time in the model. In this alternative specification, walk time is included as access time (OVTT) rather than linehaul time (IVTT). The predictive power was reduced substantially. The log likelihood at convergence increased from 33.23 to 42.92. This implies that walkers value time spent walking as productive travel time rather than onerous time spent waiting to get going.

IIA Test. Table D-34 displays the results of a binomial logit estimation run with 125 observations from the estimation data set. Only the automobile and transit alternatives were used in the choice set (100 respondents chose auto; 25 chose transit). The reduced model correctly predicted mode choice in the estimation data set for 92.8 percent of all respondents ($\rho^2 = 0.68$). There is no significant difference in coefficient estimates between this model and the full model reported in Table D-32. The cost coefficient remains largely positive and significant. Based on these results we cannot prove the hypothesis that a IIA model misspecification due to a correlation in the unobserved attributes of the motorized modes is causing the anomalous positive cost coefficient.

Comparison with SUNY Buffalo Findings. Talvitie, et al. (D11) also attempted to estimate work trip mode choice models with data from DETLINK file. They also obtained

Table D-33

BDDS WORKTRIP MODE CHOICE MODEL: ALTERNATIVE OVTT SPECIFICATION
(T-statistics in Parentheses)

$$V_A = \underset{(6.06)}{-0.57} - \underset{(2.33)}{0.05}\ \text{AUTOIVTT} + \underset{(2.07)}{2.96}\ \text{AGASCST}$$

$$+ \underset{(1.15)}{0.59}\ \text{ROOMS} + \underset{(0.64)}{1.92}\ \text{DRIVLIC} + \underset{(0.23)}{0.73}\ \text{PIPERHHI}$$

$$+ \underset{(2.29)}{6.50}\ \text{CARSPERT} - \underset{(0.95)}{0.13}\ \text{HHINC} - \underset{(1.13)}{0.99}\ \text{AGE} + \underset{(1.32)}{0.02}\ \text{AGESQ}$$

$$V_T = \underset{(0.66)}{6.65} - \underset{(2.33)}{0.05}\ \text{TRANIVTT} - \underset{(3.68)}{0.07}\ \text{TRANOVTT} + \underset{(2.06)}{2.96}\ \text{TRANFARE}$$

$$+ \underset{(0.35)}{0.17}\ \text{ROOMS} + \underset{(0.28)}{0.84}\ \text{DRIVLIC} + \underset{(0.06)}{0.20}\ \text{PIPERHHI}$$

$$+ \underset{(1.21)}{3.38}\ \text{CARSPERT} - \underset{(1.13)}{0.15}\ \text{HHINC} - \underset{(1.22)}{1.10}\ \text{AGE} + \underset{(1.41)}{0.02}\ \text{AGESQ}$$

$$V_W = \underset{(3.68)}{0.07}\ \text{WALKTIME}$$

NOBS = 175

$\rho^2 = 0.76$

$\rho_C^2 = 0.68$

$L(0) = 177.61$ $\quad L(C) = 135.81$ $\quad L(\hat{\theta}) = 42.92$

Percent Estimated Correctly = 89.4

Notes and source on following page.

Table D-33 (Continued)

BDDS WORKTRIP MODE CHOICE MODEL: ALTERNATIVE OVTT SPECIFICATION
(T-statistics in Parentheses)

NOTES:

V_A = Estimated representative utility of auto;

V_T = Estimated representative utility of transit;

V_W = Estimated representative utility of walk;

AGASCST = Automobile fuel cost (in dollars);

AUTOIVTT = Automobile in-vehicle travel time (in minutes);

TRANIVTT = Transit in-vehicle travel time (in minutes);

TRANOVTT = Transit out-of-vehicle travel time (in minutes);

WALKTIME = Walk time (in minutes);

ROOMS = Number of rooms in the household;

HHINC = Household income (in thousands of dollars);

DRIVLIC = Drivers license dummy (1=yes);

PIPERHHI = Ratio of personal income to household income

CARSPERT = Ratio of household automobiles to household travelers;

AGE = Respondent's age (in years); and

AGESQ = Respondent's age squared.

SOURCE: Charles River Associates, 1980

Table D-34

BDDS BINOMIAL WORKTRIP MODE CHOICE MODEL: IIA TEST
(T-statistics in Parentheses)

$$V_A = \underset{(1.11)}{2.95} - \underset{(3.05)}{0.09}\ \text{AUTOIVTT} + \underset{(2.85)}{8.41}\ \text{AGASCST} - \underset{(0.71)}{0.19}\ \text{ROOMS}$$

$$+ \underset{(1.66)}{1.59}\ \text{DRIVLIC} + \underset{(0.93)}{0.04}\ \text{HHINC} + \underset{(0.66)}{0.83}\ \text{PIPERHHI}$$

$$+ \underset{(1.65)}{1.58}\ \text{CARSPERT}$$

$$V_T = \underset{(3.05)}{-0.09}\ \text{TRANIVTT} + \underset{(3.42)}{13.59}\ \text{MIPEROVTT} + \underset{(2.85)}{8.41}\ \text{TRANFARE}$$

NOBS = 125

$\rho^2 = 0.68$

$L(0) = 86.64$ $L(\hat{\theta}) = 27.32$

Percent Estimated Correctly = 92.8

NOTES:

V_A = Estimated representative utility of automobile;

V_T = Estimated representative utility of transit;

AUTOIVTT = Automobile in-vehicle travel time (in minutes);

AGASCST = Automobile fuel cost (in dollars);

ROOMS = Number of rooms in the household;

DRIVLIC = Drivers license dummy (1=yes);

Table continued on following page.

Table D-34

BDDS BINOMIAL WORKTRIP MODE CHOICE MODEL: IIA TEST
(T-statistics in Parentheses)

HHINC = Household income (in thousands of dollars);

CARSPERT = Household automobiles per traveler;

TRANIVTT = Transit in-vehicle travel time (in minutes);

MIPEROVTT = Travel distance ÷ transit out-of-vehicle travel time; and

TRANFARE = Transit fare (in dollars).

SOURCE: Charles River Associates, 1980.

counterintuitive signs for the coefficients of at least one of their LOS variables (cost divided by income and time) in all specifications tested. They were, however, able to estimate models with plausible coefficients when the choice sets were constructed from the trip file (standard origin-destination data). These findings suggest that there may be problems with the DETLINK data for purposes of choice model estimation. These difficulties are in addition to the high degree of attrition in constructing choice sets (only 175 worktrips from a data set containing 966 households).

REFERENCES

D1. Charles River Associates. "A Disaggregated Behavioral Model of Urban Travel Demand." Springfield, Va., National Technical Information Service, 1972.

D2. Metropolitan Council. "A Summary of Travel in the Twin Cities Metropolitan Area." Minneapolis, Twin Cities Transportation Planning Program, April 1974.

D3. Lancaster, K. J. "A New Approach to Consumer Theory." Journal of Politcal Economy, Vol. 74, 1966, pp. 132-157.

D4. Quandt, Richard and Baumol, W. J. "The Demand for Abstract Modes: Theory and Measurement." Journal of Regional Science, Vol. 6, pp. 13-26, 1966.

D5. Adler, T. J. and Ben-Akiva, M. "Joint-Choice Model for Frequency, Destination and Travel Mode for Shopping Trips." Transportation Research Record, No. 569, 1976, pp. 136-150.

D6. McFadden, D. "The Measurement of Urban Travel Demand." Journal of Public Economics, Vol. 3, pp. 303-328, 1974.

D7. (Thiel, Henry), Principles of Econometrics. New York: John Wiley and Sons, Inc. (1971), pp. 396-397.

D8. Berechman, Joseph and Paaswell, R. E. "The Impacts of Car Availability on Urban Transportation Behavior." Urban Studies Group Working Paper, Department of Civil Engineering, State University of New York at Buffalo, August 1975.

D9. Lerman, Steven R , Manski, Charles F., and Atherton, Terry J. "Non-Random Sampling in the Calibration of Disaggregate Choice Models." Prepared for U.S. Department of Transportation. Washington, D.C., 1976.

D10. Federal Highway Administration. "User's Guide: The Baltimore Disaggregate Dataset." Washington, D.C., 1980.

D11. Talvitie, A., Dehghani, Y., Supernak, J., Morris, M., Anderson, M., De John, A., Kousheshi, B., and Feuerstein, S. "Refinement and Application of Individual Choice Models in Travel Demand Forecasting" (2 Volumes). Prepared for the Federal Highway Administration. Buffalo, N.Y., State University of New York at Buffalo, 1981.

APPENDIX E

THE INDEPENDENCE OF IRRELEVANT ALTERNATIVES PROPERTY OF THE MULTINOMIAL LOGIT MODEL

INTRODUCTION

This Appendix provides an analysis of the Independence of Irrelevant Alternatives (IIA) property of the multinomial logit model. The IIA property is perhaps the most controversial issue in the application of disaggregate demand models and has been alleged to be a substantial impediment to implementation. For this reason, this study devoted considerable attention to analyzing the property.

The basic conclusion of the analysis is that the IIA assumption is not an inherently undesirable assumption of the logit model. However, the diagnostic tests developed in Phase I (E1) should be applied when a violation of the independence assumption is suspected. If the test confirms that a violation is likely, modifications of the model should be performed to account for the violation. These modifications are discussed in Chapter 2.

This appendix has approached the problem of the IIA by accepting the basic framework of the multinomial logit model and modifying the model, the form of the data, or the estimating approach to account for violations of the model's assumptions. For a discussion of alternative modeling approaches which do not assume IIA, see Chapter 2.

This appendix proceeds in the following manner: discussion of the significance of the IIA property; examination of the reasonableness of the IIA property; and classification of possible violations into two cases. Readers interested in a more detailed discussion of the IIA property should refer to the Phase I Interim Report (E1), or McFadden, Tye, and Train (E2). The Phase I CRA report contains an extensive analysis of the consequences of a violation of the independence property, diagnositic tests for identifying a violation, remedies for a violation, a test of the assumption using actual data, and a discussion of the relationship between partitioned (sequential, conditional) models and the independence assumption.

Much of the detailed analysis of the IIA property has not been included in the final report because of space limitations. The principal conclusions of the earlier report may be summarized as follows.

1. Violation of the IIA property will generally result in biased estimates of the model coefficients and erroneous forecasts of modal shares. The direction of error depends on the specific source of error (see CRA (E1) for a detailed classification of the types of errors and direction of bias).

2. Detailed statistical tests are proposed using restricted choice sets and application of the "universal logit model" to the full data set.

3. Improved model specification to make the IIA valid is often the preferred remedy. In other cases elimination of a dependent mode in the choice estimation process may eliminate the bias. The "maximum model," the "Cascade Model," and the "dominance model," discussed in the Phase I report, have been superseded by other model specifications discussed in Chapter 2.

1. THE SIGNIFICANCE OF THE INDEPENDENCE OF IRRELEVANT ALTERNATIVES (IIA) PROPERTY

The IAA in Models of Individual and Group Behavior

The multinomial logit (MNL) model of travel demand assumes that an *individual* will choose among alternatives with probabilities based on the attributes of the choices (level of service, or LOS) offered and the socioeconomic characteristics (SEC) of the individual:

$$P_i = \frac{e^{V_i}}{\sum_{j=1}^{N} e^{V_j}} \qquad \text{(E-1)}$$

where P_i = probability of choosing ith alternative (e.g., mode);

e = base of natural logarithms;

V_i = function (usually linear) of the LOS of ith alternative and the SEC of the individual; and

N = number of alternatives in the choice set.

It is clear that this choice model assumes that the ratio of probabilities of choosing any two alternatives is invariant to the attributes (or even existence) of a third alternative:

$$\frac{P_i}{P_j} = \frac{e^{V_i}}{e^{V_j}} .$$

This is the IIA property of the MNL model, and it is one of the most important assumptions made by the model.

The IIA property states that if two modes are available and a new mode is introduced, the ratio of the probabilities of the two old modes will be unchanged regardless of the probability of choice for the new mode. For example, if the new mode will be chosen with probability of 0.10, and each old mode had 0.50 probability before the introduction of the new mode, the probability of each of the old modes will be

0.45 after the new mode is introduced, thus preserving the one-to-one ratio of probabilities.

Translated to the behavior of homogeneous groups, the property greatly facilitates the forecasting problem for a new mode. If 100 persons have the same levels of SEC and LOS which are included in Equation E-1, the demand for a new mode can be calculated (under certain conditions) by adding another term to the denominator of Equation E-1 and recomputing all choice probabilities. The new probabilities can then be multiplied by 100 to estimate the demand for each mode. If the old modes formerly shared the market equally and the new mode's probability was 0.10 for each individual in the market segment, the predicted mode demands would be 45, 45, and 10.

The IIA "Problem"

The example that has troubled many critics of the IIA property is the classic "blue auto/red auto" case, i.e., the "new mode problem." This example has been modified from the "blue bus/red bus" example familiar to students of the MNL model to permit a new irrelevant auto mode rather than a new irrelevant bus mode. The familiar "blue bus/red bus" example appears to increase bus LOS when the new mode is introduced.

Suppose, for example, that the blue auto and bus mode each capture 50 percent of a given travel market as shown in Table E-1. Assume that a new auto mode is introduced with exactly the same service attributes as the old auto mode except that the auto is painted a different color, red (to which the patrons are indifferent). We assume that the auto is leased for this trip only, to abstract away from auto ownership questions and competing demands for the auto. We expect that the true modal shares will now be 1/2, 1/4, 1/4 for bus, the first auto alternative, and the second auto alternative respectively.* However, the ordinary MNL model will forecast that each of the three modes captures one-third of the market (see column 2), which is clearly a poor forecast. The Independence of Irrelevant Alternatives property of the MNL model is clearly the reason; the property requires that the ratio of the bus share to the first auto mode share should be unaffected by the introduction of a new mode. In this example the ratio is 1.0. When we add a new auto mode, the IIA property requires that we reduce the bus share from 1/2 to 1/3 to keep constant the ratio of bus mode share to the first auto mode share.

---

*No bus users will switch to the new mode and auto users will split evenly between the two auto modes.

Table E-1

THE BLUE AUTO/RED AUTO PROBLEM

(Percent of Market)

	1 True and MNL Mode Choice (Binary Choice)	2 Predicted MNL Mode Choice (3 Modes)	3 True Mode Choice (3 Modes)
Bus	50	33	50
Blue Auto	50	33	25
Red Auto	0	33	25
Total Percent	100.0	100.0	100.0

If the problem were confined to this simple example, there would be no problem. The new auto mode is clearly "irrelevant" and should not be listed as a mode. But what if we were to introduce "express bus" or the "auto passenger" mode? How "independent" are they? Clearly, the extreme case in the example points to a "gray area" where the demand forecast for a "new mode" or forecasting the effect of a change in service *could be* seriously imperiled by incorrectly applying the IIA property.

What the IIA Property Does Not Say

It is extremely important to note that the IIA property translates from the probability of the individual to predicted market shares in a group of people only when applied to an aggregation of "homogeneous" individuals with identical specification of the choice structure (i.e., Equation E-1) and identical observed socioeconomic attributes and LOS. (The differences between the behavior of individuals in this market are due only to random independent unobserved LOS and socioeconomic attributes not included in the model.) The IIA property will not apply after the homogeneous market segments have been aggregated to represent the market as a whole.

The IIA property, like the MNL model itself, is not preserved through aggregation in a heterogeneous population. Much of the criticism of the IIA property is based on the incorrect assumption that the MNL specification permits the

IIA property to predict the demand for a new mode for a heterogeneous population, composed of individuals with unequal choice probabilities, LOS, and SEC. The IIA property of the MNL model does not hold for mode shares in such a heterogeneous population.

Although "mode shares" and "probability" will be used interchangeably in the following discussion of the IIA, the property does not apply to aggregate mode shares in general. When the term "mode shares" is used to illustrate the IIA property, it is assumed that the market is composed of persons with identical choice probabilities and measured socioeconomic attributes, LOS, and specification of Equation E-1.

To take a specific example, the MNL model does not in general predict that if a new mode is introduced to a population composed of different market segments, with different socioeconomic characteristics and different level-of-service attributes, that equal percentages will be drawn from both auto and transit users to the new mode. Rather, the model predicts, when summed over market segments, as we expect, that if the new mode resembles transit in its measured attributes, the percentage of transit users in the heterogeneous population switching to the new mode will be greater than the percentage of auto users.

This principle may be illustrated by an example. Table E-2 illustrates the case where a population of 200 is composed of two homogeneous market segments of 100 persons each. For the purpose of illustration, we assume the choice environment of observed attributes is identical for all persons within each segment but differs substantially between the two market segments. Segment 1 is "auto oriented," splitting 90/10 in favor of auto. Segment 2 is "transit oriented," splitting 90/10 in favor of transit.

A new mode is introduced, "dial-a-bus," and it is predicted by the MNL model to capture 5 percent of the first market share and 15 percent of the second market share. The ratio of auto market share to bus market share is preserved *within each homogeneous market segment*, and the MNL predictions emerge in columns 4, 5, and 6. We note that the overall ratio of bus mode share to auto mode share is not constant after the new bus mode is introduced, but falls from 1.0 to 0.91 for the entire population (86÷94 = 0.91).

Although the same percentage diversions occur from bus and auto to dial-a-bus *within each homogeneous market segment* (e.g., in segment 1, 5 percent of both bus and auto patrons switch), the predicted diversions from auto and bus are not the same for the population as a whole. Of the 100 total bus patrons in the binary choice situation,

Table E-2

EFFECT OF THE INDEPENDENCE OF IRRELEVANT ALTERNATIVES PROPERTY ON A FORECAST OF BEHAVIOR IN A POPULATION OF HETEROGENEOUS MARKET SEGMENTS

(Persons)

Mode	MNL Mode Share (Binary Choice)			Predicted MNL Mode Share (3 Modes)		
	1	2	3	4	5	6
	Market Segment I	Market Segment II	Total Market	Market Segment I	Market Segment II	Total Market
Bus	10	90	100	9.5	76.5	86.0
Auto Driver	90	10	100	85.5	8.5	94.0
Dial-a-Bus	0	0	0	5.0	15.0	20.0
Total Persons	100	100	200	100.0	100.0	200.0

14 percent (100-86) were predicted to switch to dial-a-bus. However, only 6 percent of total auto users were predicted to switch to the dial-a-bus.

It is important to note that if the IIA property is violated for each of the homogeneous market segments, the MNL model will make poor forecasts for each of the market segments and for the market as a whole. Although the IIA property is not preserved through aggregation over homogeneous market segments to represent a heterogeneous market (i.e., the ratio of probabilities will not remain the same after aggregation over the homogeneous market segments), forecasts of the entire market will be erroneous if the property is violated for each market segment. Put differently, aggregation will not generally offset the effects of a violation of the IIA property for individual market segments.

Importance of the IIA Issue

The IIA property of the MNL model is obviously a key assumption of the model.

1. The urban transportation planning community has been particularly interested in the problems of forecasting demand for "new modes" -- such as jitneys, carpooling, dial-a-ride, subscription carpooling, and other varieties of "demand responsive" and "paratransit" systems. However, the applicability of the MNL model to forecasting demand for a "new mode" depends critically on the validity of the IIA property of the

model. As shown above, the IIA property can be questionable in hypothetical problems involving new modes. Clarification of this issue is obviously critical to applying the MNL model to this planning problem.

2. The controversy concerning the Independence of Irrelevant Alternatives property of the MNL model of disaggregate travel demand extends to issues beyond the model's applicability to new modes. Progress in resolving the questions raised by this issue is essential to securing general acceptance of the disaggregate method by practitioners of demand modeling and transportation systems planning. In fact, some researchers consider this property of the model so undesirable that they seriously question the usefulness and applicability of the MNL model to *any* travel demand forecasting problem.

3. The IIA property is a critical factor in resolving other controversies relating to the use of the MNL model. In particular, it is a key factor in designing data collection methods and estimation procedures that avoid bias in model calibration that may otherwise result in using cross-section data. The relationship between the IIA and the other properties is discussed below.

2. PRINCIPAL CONCLUSIONS REGARDING THE INDEPENDENCE OF IRRELEVANT ALTERNATIVES PROPERTY

The validity of the IIA property of the MNL model has been identified as one of the most important issues to be addressed in disaggregate demand modeling (E3). The present study has concentrated on identifying (1) the assumptions of the MNL model that cause it to have the IIA property, (2) the basic underlying reasons why the IIA property could be violated in a particular application, (3) the consequences of a violation of the IIA property, (4) tests to diagnose a violation, and (5) corrective measures for modifying the MNL model if the IIA property has been violated. Detailed discussion of the latter three items is presented in the Phase I Interim Report (E1). The research has produced the following conclusions.

1. The IIA property of the MNL model is not an inherent drawback of the MNL model, but is an assumption that is frequently reasonable and desirable.

Comment

The IIA property is not preserved through aggregation to describe the behavior of aggregates of heterogeneous market segments. The IIA property of the constancy of the ratio of the market shares of two modes with the introduction of a new mode does not hold for a large population of heterogeneous people, with different alternatives and different socioeconomic variables and tastes.

2. The IIA property is based on the assumption that the attributes of the transportation system and individuals that are not included in the analysis (the "omitted attributes") do not have systematic effects which distort the model results.

Comment

Intuitively, the IIA property is analogous to the assumption of "independent error terms" or disturbances employed in regression analysis. Unless the researcher has *a priori* or empirical confirmation that the omitted attributes have a systematic effect on the results, the agnostic assumption of the IIA may be perfectly reasonable in many applications.

3. Systematic effects of unobserved attributes are an important possible source of a violation of the independence assumption.

The IIA assumption can be violated in two basic ways: the unobserved attributes have common values in two or more alternatives (the "blue auto/red auto" problem), or unobserved attributes have a systematic relationship with observed attributes.

Comment

A "Case 1" violation of the independence assumption occurs when two alternatives have common unobserved attributes, such as when a mode has unobserved comfort or safety levels that are similar to those of another mode and the independent unobserved attributes are relatively insignificant. A "Case 2" violation of independence occurs when an unobserved attribute is correlated with an observed attribute. For example, it may be the policy of the transit authority to place the most comfortable (unobserved attribute) buses on the routes serving the longest trips (observed attribute), passengers may locate their residences to take advantage of preferable levels of observed attributes in modes that are preferred for unobserved taste reasons, and passengers may have substantial differences in their valuation of the observed attributes for unobserved taste reasons. These systematic effects of unobserved attributes are inconsistent with the independence assumption. These two violations will be referred to as Case 1 dependence and Case 2 dependence in the remainder of the appendix.

4. When the IIA assumption is unreasonable, the MNL model cannot be applied without error. The error may be large or small depending on the circumstances.

Comment

If the MNL model is calibrated on data that do not satisfy the IIA property, the results generally will be biased estimates. This bias will not be overcome by increased sample size. If the MNL model is used to forecast demand for a new mode when the IIA property is violated, demand for the new mode will generally be overstated. In examples of a violation of the independence assumption resulting from dependence between observed and unobserved attributes, the calibration procedure will impute to the observed attribute both its independent effect and the effect of the unobserved attribute with which it is correlated. Frequently, the effect will be to estimate elasticities of observed attributes that are too large.

5. Because the reasonableness of the IIA property is dependent on the circumstances, *diagnostic tests* that identify when the assumption is reasonable and unreasonable should be applied.

Comment

There are costs to assuming the independence property when it is false and there are costs to rejecting the independence assumption when it is true. CRA has designed diagnostic tests to test the reasonableness of the independence assumption. Based on research in this project (see the Phase I Report (E1)) the most powerful tests were performed on restricted data sets, e.g., calibrating the MNL model for a two-mode choice setting and comparing the results with the model calibrated in the three-mode choice setting. Research performed by Horowitz (E4) expanded on this finding and should be consulted by the practitioner desiring to apply these tests. Readers should also see McFadden, Tye, and Train (E2).

6. If the diagnostic test indicates that the IIA assumption is invalid, corrective measures are available to take the dependence into account.

Comment

If the independence assumption is invalid, the alternative corrective measures are to improve the model specification so that the dependence is eliminated, change the model calibration procedure to account for the dependence, or to change the model to allow for dependence. Improved model specification to make the IIA assumption reasonable is the preferred remedy when it is feasible. Research indicates that the "Fully Competitive Model" and the "Cascade Model" do not correct for a violation of the IIA. Research also indicates that arbitrary nonbehavioral modifications of the MNL model to eliminate the IIA property on *a priori* grounds will frequently make matters worse. Preferred approaches, such as multinomial probit, are discussed in Chapter 2.

3. AN EXAMINATION OF THE REASONABLENESS OF THE INDEPENDENCE OF IRRELEVANT ALTERNATIVES PROPERTY

Introduction

The MNL model is founded on a well-defined set of assumptions concerning the choice behavior of the individual and concerning the method by which the behavior of the individual can be extrapolated to predict the behavior of groups. The approach of this study is to investigate the IIA property by returning to the fundamental assumptions of individual behavior inherent in the model. If the assumptions of the individual's behavior inherent in the MNL model are violated, the IIA property will generally be invalid. To correct for the invalid assumptions, the preferred approach is to modify those assumptions of individual behavior as necessary to account for the violation, and infer what new models of group behavior follow from the amended assumptions of individual behavior. This approach is distinctly different from unsuccessful attempts to deal with a violation of the IIA property that arbitrarily modify the MNL model of group behavior without identifying the assumption of the model of individual choice that has been violated.

Intuitively, the IIA assumption plays a role in the MNL model which is analogous to the assumption of "independent error terms" (or "disturbances") in least squares regression.

The IIA assumption implies that factors that are omitted from the analysis ("unobserved attributes" in the MNL terminology) do not systematically affect the outcome, but are independent random factors.

One possibly important source of a violation of the IIA property is systematic variations in unobserved attributes across alternatives. This source is correlation between alternatives on important "left out" variables resulting from model misspecification. An example might be the fact that "local bus" and "express bus" modes have precisely the same level (perhaps) of comfort and other unobserved attributes and differ only in terms of travel time. If so, the assumption that the unobserved attributes of these two alternatives are uncorrelated is clearly violated. The result is biased and inaccurate model predictions.

In the section below, we develop examples that illustrate intuitively why the assumption of independent unobserved attributes is critical to the MNL model. We will formally link the IIA assumption to the theoretical model of individual behavior inherent in the MNL model and discuss reasons why the IIA property is likely to be satisfied in many applications.

The Relationship Between the Role of the Unobserved Attributes of Alternatives and the Independence of Irrelevant Alternatives Property

Chance, unobserved determinants of behavior, and the laws of probability play an important part in the formulation of the MNL model. The MNL model assumes that actual traveler choices are based on two types of factors:

1. "Observed" attributes of the traveler and the attributes of the choices he is offered, (these are identified and included in the model); and
2. "Unobserved" attributes, which may also influence decision making (these are not identified, measured, and included in the model).

The MNL approach specifically considers that behavior of individuals is random and cannot be predicted with certainty. Only the average behavior of groups with similar observed attributes and facing similar choices can be predicted with any confidence. If we observe an individual with certain attributes and facing certain choices, we can only say that he will choose the first alternative with probability of (say) 0.75 and the second with probability 0.25. We cannot predict the actual choice, which is random.

Translating these assumptions to group behavior, the MNL model says that of 100 subjects with identical observed attributes relevant to the choice, 75 will (on average) choose the first alternative and 25 will choose the second. A very important implication of this example is that, as all 100 subjects had the same observed attributes and faced the same choices, *the split of this population between the two alternatives was determined exclusively by the "unobserved" attributes* (e.g., personal idiosyncracies). Since all subjects had the same level of observed attributes, the differences in behavior were determined solely by differences in unobserved attributes.

The MNL model explicitly assumes that unobserved attributes of choice alternatives affect choice in a nonsystematic manner. When behavior does not satisfy the IIA property, it may be because *unobserved attributes* of two alternatives are similar (or, more generally, statistically correlated). By itself, similarity of observed attributes will not cause a violation of the IIA property.

All models of behavior must make some assumption as to how behavior is affected by factors that are not included in the analysis -- the "unobserved" attributes. Most models of behavior, whether in transportation or elsewhere, make the simplifying assumption that unobserved attributes are random and unsystematic. In general, this assumption is reasonable: if the analyst thought that an unobserved attribute was important and systematic in its influence, he would, if feasible, include it in his analysis as an observed attribute.

The previously cited example can be used to illustrate why the similarity of unobserved attributes is an important cause of a violation of the IIA property. Table E-1 illustrates a mode choice problem where we confront the traveler with three modes -- a bus, a blue auto, and a red auto. We assume that all three modes have the same *observed* levels of service (e.g., time and cost), that all three are "independent" alternatives (an incorrect assumption), and that the traveler is indifferent to the two colors. We also assume that there is no "auto bias" or "transit bias"; if the service levels are equal, the population would truly split evenly between the auto mode and the transit mode (this assumption is not essential to the problem and will be relaxed below).

In column 1 we consider a binary choice between a bus and a "blue auto." The MNL model correctly predicts that the population will split evenly. *Because the observed attributes were identical for the two modes, the choice of individuals was determined exclusively by the unobserved attributes.*

In the second column, the MNL model is used to predict the effect of the new mode, "red auto." Since all three

modes had the same level of observed attributes, and there was no "auto bias" or "transit bias," the MNL model assigned each one third of the market. Note that the ratio of mode shares of bus and "blue auto" do not change as required by the IIA property. A key assumption of this forecast is the incorrect assumption that the new auto mode was "independent."

Column 3 illustrates that we expect that a blue auto is not "independent" of a red auto and that the only effect of introducing the new mode is to shift half the blue auto drivers to red auto drivers (i.e., a random choice among colors) and leave bus demand unchanged.

What reason do we have for assuming that the two auto modes are not "independent"? Clearly, the reason has nothing to do with the observed attributes -- they are the same for all three modes. The observed attributes of the "red auto" are no more similar to the "blue auto" than to the bus. However, it is obvious that the "blue auto" and "red auto" are identical in their unobserved attributes (comfort, status, convenience, ubiquity, etc.) and both are very dissimilar from bus in their unobserved attributes. The lack of independence of the new auto mode is a lack of independence of the unobserved attributes.

The above example assumed a 50/50 mode split and identical levels of observed attributes of choice to direct attention away from the observed attributes and to concentrate on the unobserved attributes. Table E-3 illustrates that if the 50/50 assumption is relaxed, the forecasting error due to a violation of the IIA is still present. If we assume that the auto mode share is 80 percent due to a service differential or "mode bias," the MNL model will incorrectly predict a decline from 20 to 11.11 percent of the market share for the bus mode so that the blue auto to bus share ratio remains at 4 to 1. Again, the problem is that the unobserved attributes of the two auto modes are not independent (although it is not as obvious in this example because the two auto modes have observed attributes different from the bus mode).

Assumptions of the Multinomial Logit Model and the Independence of Irrelevant Alternatives Property

Discussion of Basic Approach. The IIA property of the MNL model may be expressed in two ways:

1. For the individual, the *ratio of probabilities* of choosing any two alternatives is invariant to the attributes (or existence) of other alternatives; and
2. For a "homogeneous market segment" (i.e., composed of persons facing identical alternatives and having identical socioeconomic attributes and an identical

TABLE E-3
THE BLUE AUTO/RED AUTO PROBLEM
(Percent of Market)

	True and MNL Mode Choice (Binary Choice)	Predicted MNL Mode Choice (3 modes)	True Mode Choice (3 modes)
Bus	20	11.11	20
Blue Auto	80	44.44	40
Red Auto	0	44.44	40
Total Percent	100.0	100.0	100.0

structure of choice frequencies (e.g., Equation E-1) *the ratio of market shares* of two alternatives is invariant to the attributes (or existence) of other modes.

The IIA property ordinarily will be violated whenever the assumptions of the MNL model are violated. The approach will be to specify exactly what the MNL model assumes regarding the choices of individuals and relate that individual choice model to the MNL forecast of the mode shares for a population. If the MNL model incorporates an unwarranted assumption regarding the choices of an individual, poor forecasts of population market shares may result. Conversely, incorrect forecasts of market shares create a presumption that the model of individual choice was invalid. Accordingly, in the next section we consider in detail the relationship between individual choice behavior and the observed choice frequencies in the population.

Assumptions of the Binary Logit Model. The binary logit model hypothesizes that the probability that an individual will choose a given alternative in a choice situation is given by

$$P_i = \frac{e^{V(X_i,S)}}{\sum_{j=1}^{2} e^{V(X_j,S)}} = \frac{1}{1+e^{V(X_j,S)-V(X_i,S)}} \qquad \text{(E-2)}$$

where:

P_i = probability of choosing i^{th} alternative;

e = base of natural logarithm;

X_j = vector of attributes of the j^{th} alternative (such as time and cost of a mode);

S = vector of socioeconomic variables; and

$V(X_j,S)$ = a function (usually linear) of the attributes of the alternatives and socioeconomic variables.

A model such as Equation E-2 is a *specification of selection probabilities*. Researchers using models of this type have used basically two approaches. The first is to deal directly with the selection probabilities and to develop models that produce a good fit to the observed choice frequencies among the alternatives in a sample. The other alternative, characterized especially by Block and Marshak (E5) and further developed by McFadden (E6) and others, is to hypothesize that the selection probabilities are based on an explicit model of *individual choice behavior*. Following the latter approach, it will be shown that any undesirable properties of the model of selection probabilities result from violations of the implicit assumptions of the individual choice behavior model. Modification of the assumptions concerning the individual choice behavior leads to models of selection probabilities which may have more desirable properties. This approach suggests methods for testing the reasonableness of the assumptions of the MNL model.

The chief findings of the McFadden (E6, E7) approach that relate the selection probabilities to the underlying theory of individual choice, are as follows:

1. There is in general a direct relationship between the assumed model of individual choice and the properties of the model of selection probabilities;
2. In particular, the binary logit model of population selection probabilities can be shown to be consistent with a theory of sampling from a population of utility-maximizing consumers, where the utility of an alternative is decomposed into a component based on attributes observed by the sampler and a component based on attributes unobserved by the sampler; and
3. If we assume (a) the utility of an alternative is decomposed into a nonrandom component which is a linear function of observed attributes of the choice set and the individual, and a random utility component which is a function of unobserved attributes; (b) the unobserved random utility components are independently and identically distributed; and (c) the subject is assumed to choose the alternative

with the highest utility; a necessary and sufficient condition for the logit model is that the unobserved random utility components of the alternatives be Weibull distributed.

In short, the logit model may be derived by assuming that consumers maximize utility, that a component of the utility of each alternative cannot be observed because it depends on unobserved attributes, and that the random components of utility are independent Weibull distributions. If the observed selection probabilities are not consistent with those predicted by the model, then the key MNL assumption of independent unobserved attributes of choice across the alternatives is an obvious place to look for an explanation.

The critical role played by the unobserved attributes of the choice alternatives may be illustrated in the case of the binary logit model which is linear in parameters. The utility of the individual for the i^{th} alternative is assumed to be a function of the observed attributes of the alternative, his observed socioeconomic characteristics, and an unobserved component. The unobserved component represents the effect of random taste variations, omitted choice attributes, and omitted socioeconomic variables:

$$U_i = U_i(X_i, S, \varepsilon_i) \qquad \text{(E-3)}$$

where:

U_i = utility of i^{th} alternative;

X_i = vector of observed attributes of i^{th} alternative, $(X_{i1},\ldots X_{iN})$;

S = vector of observed socioeconomic characteristics; and

ε = utility component due to random and/or omitted influences.

As a special case assume the utility is linear:

$$U_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \alpha_{i1} S_1 + \varepsilon_i \qquad \text{(E-4)}$$

where:

X_{ij} = j^{th} attribute of the i^{th} alternative;

β_j = coefficient of j^{th} attribute;

S_1 = first socioeconomic variable;

$\alpha_{i1}S_1$ = contribution of first socioeconomic variable to utility of i^{th} alternative; and

ε_i = unobserved utility component of i^{th} alternative.

We assume that U_i is chosen if and only if it has greater utility than any other alternative. In the case of binary choice, the first alternative is chosen if and only if $U_1 > U_2$. However, as long as there are random components, ε_1 and ε_2, there is no certainty that an alternative will be chosen even if its observed utility component is known to be greater than the observed component of the other alternative; the size of the unobserved components could be sufficient to produce a choice that is inconsistent with the

ranking based on the observed component. We define V_1 and V_2 to be the utility due to the effect of observed components:

$$V_1 = \beta_1 X_{11} + \beta_2 X_{12} + \alpha_{11} S_1 \qquad \text{(E-5)}$$

$$V_2 = \beta_1 X_{21} + \beta_2 X_{22} + \alpha_{21} S_1$$

Because of the unobserved determinants of choice, the choice of the individual cannot be known with certainty based on the observed components. However, we do know that alternative number one will be chosen if the observed utility component of number one dominates that of number two and this is not offset by the differences in unobserved components, or if the unobserved components of number one sufficiently dominate those of number two so that the lesser observed utility of number one is compensated by its greater unobserved utility. That is

$$U_1 > U_2$$

implies $V_1 + \varepsilon_1 > V_2 + \varepsilon_2$,

or $\quad V_1 - V_2 > \varepsilon_2 - \varepsilon_1$.

Because ε_1 and ε_2 are random variables, the event $U_1 > U_2$ is also random. The probability that it occurs is given by:

$$P(U_1>U_2) = P(V_1-V_2>\varepsilon_2-\varepsilon_1) \tag{E-6}$$
$$= P(\varepsilon_2-\varepsilon_1<V)$$

where $V=V_1-V_2$.

To determine the probability that U_1 is selected over U_2, we must know the probability distribution of the ε_i's. We assume that ε_i has the reciprocal exponential, or Weibull distribution:

$$P(\varepsilon_i \leq t) = e^{-e^{-t}}$$

McFadden (E7) has proved the Weibull distribution is both necessary and sufficient to produce a logit cumulative probability distribution for the population selection probabilities when the model of individual choice is given by Equation E-6, where the ε_i are *independent:*

$$P(U_1>U_2) = P(\varepsilon_2-\varepsilon_1<V) = \frac{1}{1+e^{-V}} \quad . \tag{E-7}$$

The conventional notation for the binary logit specification in odds ratio form may be derived by letting

$$P_1 = P(U_1>U_2)$$

$$P_1 = \frac{1}{1+e^{-V}}$$

$$P_1 + P_1 e^{-V} = 1$$

$$e^{-V} = \frac{1-P_1}{P_1}$$

$$\frac{P_1}{1-P_1} = e^V$$

$$log \left(\frac{P_1}{1-P_1}\right) = V=V_1-V_2=\beta_1(X_{11}-X_{21})+\beta_2(X_{12}-X_{22})$$
$$+ (\alpha_{11}-\alpha_{21})S_1 \tag{E-8}$$

The proof of the relationship between the assumed models of individual choice and resulting models of selection probabilities is extremely important to our investigation of the IIA property. It suggests that an important source of a violation of the IIA property could occur when the unobserved components of the utilities are not distributed by the independent Weibull distribution. Probing the reasonableness of the assumption of the independence of the unobserved components is the key to understanding the effects of violating that assumption. From this understanding we can develop tests to determine whether the independent Weibull assumption is reasonable and develop remedies for cases where the assumption is violated.

A theory of the selection probabilities of a sampled population has been developed from a theory of individual choice. In the theoretical model, the individual always chooses the alternative with the highest utility. The observed frequency distribution of choices for a sampled population is determined by the random, unobserved components that affect each alternative's utility for that individual. The randomness is only perceived by the observer and is due to inadequate knowledge of the determinants of behavior. Our model does not state that an individual assesses the utility of each alternative according to the observed attributes of the choice and then proceeds to choose an alternative randomly with probability of the choices determined by the V_i.

The Multinomial Logit Model. The multinomial logit (MNL) model is a generalization of the binary logit case and in general is distinguished mainly by the number of alternatives. As before,

$$U_i = V_i(X_i,S) + \varepsilon_i$$

$$P_i = P[\varepsilon_j-\varepsilon_i<V(X_i,S)-V(X_j,S)] \quad j\neq i$$

$$P_i = \frac{e^{V_i(X_i,S)}}{\sum_{j=1}^{N} e^{V_j(X_j,S)}} \tag{E-9}$$

The odds, or ratio of probabilities of any two choices, are given by the IIA property:

$$\frac{P_i}{P_j} = \frac{e^{V_i(X_i,S)}}{e^{V_j(X_j,S)}} \; . \tag{E-10}$$

In the case where the V are linear

$$\log\left(\frac{P_i}{P_j}\right) = \beta_1(X_{i1}-X_{j1})+\beta_2(X_{i2}-X_{j2})+\ldots$$

$$+ (\alpha_{11}-\alpha_{21})S_1. \tag{E-11}$$

Forecasting Properties of the Model Due to the IIA Property. Two very important features of the model follow from the independence property, Equation E-10.

1. "Equality of the cross elasticities." Suppose a third mode improves in desirability relative to the first two modes. The percentage diversions from the first two modes must be identical to preserve the constancy of the ratio of the selection probabilities for the first two as required by the IIA. Consequently, the cross elasticity of all other mode shares with respect to a change in the attributes of a given mode must be equal. (Cross elasticity here is defined to be the percentage change in a mode share, using the prior share as base, resulting from a 1 percent change in the attribute of another mode.)

This property is illustrated by the formula for cross-elasticity in the MNL model. (See Ben-Akiva (E8), p. 184.)

$$E_{X_{ik}}^{P(j)} = -P_i \; \beta_k \; X_{ik} \tag{E-12}$$

where $E_{X_{ik}}^{P(j)}$ = elasticity of the j^{th} probability with respect to the k^{th} attribute of the i^{th} alternative.

It is clear from the expression that the cross-elasticity coefficient is the same for all alternatives; it is independent of the index j. If an improvement in dial-a-ride service reduces the probability *of an individual* choosing auto by 10 percent, it is also assumed according to the IIA property to reduce the probability of choosing transit by 10 percent.

2. The "new mode property". If a new mode enters the market, its probability of choice is given by the logit specification. Since all probabilities add to 1 after the new mode is introduced, and since the IIA property requires the odds ratio of the old modes to remain unchanged, the percentage reduction in probability of choice for each old mode must be the same, and must be exactly equal to the probability of choosing the new mode.

The A Priori Reasonableness of the IIA Property

One's intuitive reaction to the two forecasting consequences of the IIA property is to assume that the IIA property is inherently undesirable. After all, it is not difficult to devise examples where the model nonsensically predicts the ratio of mode shares to be unchanged when common sense says they should change. The most frequently cited example is the red auto/blue auto case cited above.

To understand why the MNL model forecasts poorly in the red auto/blue auto example, we must first establish what assumption of the MNL model was invalid. Clearly the key assumption was that the unobserved components of utility were *independently* distributed with the *Weibull* distribution. Intuition tells us that the assumption of independence of unobserved components is critical. For example, the problem in the red auto/blue auto case is clearly the assumption of *independence*: the unobserved components of utility for the blue auto and red auto were perfectly and positively correlated (as were the observed components), since they were viewed as equivalent alternatives.

If the source of the problem is dependence (or correlation) in the unobserved attributes of the alternatives, several consequences are immediate. First, the problem may be most obvious in the pure case of the blue auto/red auto example, but it could exist in a lesser degree in other, more-difficult-to-diagnose examples. Second, the two forecasting consequences of the IIA property, the "new mode problem," and the "equality of cross elasticities," appear superficially to be *per se* undesirable properties in any forecasting situation. Since they derive from the assumption of independent ε_i, it might be inferred that a fundamental assumption of the logit model, independence of the ε_i, must be rejected *a priori* as arbitrary and unreasonable. If so, the logit model itself may be an arbitrary and unreasonable model of behavior.

However, research on the IIA property and its relationship to possible systematic effects of unobserved attributes indicates that the IIA property probably can be shown to be reasonable and desirable in many applications. The two forecasting properties of the model (the equality of cross-elasticities and the new mode property) will often be desirable and reasonable properties. Indeed, as shown below, it is desirable that models be specified so that the IIA property is reasonable rather than to reject the property as inherently undesirable.

On the other hand, the practitioner cannot always cavalierly assume the property to be satisfied. Nonindependence is not confined

to the simple case of the red auto/blue auto example, and may be manifested in a number of subtle ways.

Considering the IIA property as reasonable and desirable in many circumstances is not merely making a virtue of a necessity. To see that the IIA assumption may be reasonable, we note first that the IIA property applies to the probabilities of choice of an individual and can be extended to the mode shares of a population only when that population is "homogeneous"; each individual has the same probability of choosing each alternative because all possess identical socioeconomic attributes and attributes of choice. Within a "homogeneous" market segment, the actual split of individuals into two choices in a population facing the same observed choice attributes and observed socioeconomic variables is dictated totally by the unobserved components. Since all individuals possess the same observed attributes, the actual split of the population is dictated by the random unobserved components.

In such circumstances, it would be more appropriate to question how a departure from the IIA assumption could be justified. Within such homogeneous market segments where the IIA may apply, the researcher has ordinarily, as it were, "used up" his knowledge of the determinants of choice in specifying the observed attributes, which make no contribution to explaining the differences in choice behavior within this homogeneous group. If a new mode is introduced to this homogeneous market segment, the researcher has no basis for predicting in advance that more patrons will be diverted from one mode than another. Just as the researcher had no knowledge of why particular individuals selected the old modes, neither can he say which individuals or group of individuals will be more likely to switch to the new alternative.

To help clarify these points, we take an example where mode C is introduced to a homogeneous population (with common observed variables) split between modes A and B. We pose the following two questions.

Question One. In applying the MNL model to the entire homogeneous population, we get the same results as if we had first split the population into those who chose A in the binary choice situation and those who chose B in the binary choice situation and applied a binary model to the choices between A and C and between B and C separately. Is this reasonable?

The assumption may well be reasonable. The assumption of the IIA is tantamount to saying that within a homogeneous population, the researcher can infer no knowledge from an individual's selection of A over B or B over A that is relevant to predicting his probability of switching to C. The IIA assumes that the initial choice between A and B for an individual is dictated by unobserved attributes. If the researcher has knowledge of some attribute or taste which the individual who chose A in the binary

case possesses which was not possessed by the individual who chose B, which could be used to predict a differential response of the two groups to C, the researcher would have included that determinant of choice in the observed attributes. However, if he had included that distinguishing characteristic of the two groups as a measured attribute, the two groups would no longer be homogeneous with respect to the observed attributes, and the IIA property would not hold. The very fact that an attribute affecting choice is not included as an observed attribute creates a presumption that its affect is random. If a differential response of the two groups (those who choose A and those who choose B in the binary choice) can be predicted on the basis of some differential attribute of the two groups, the two groups are no longer homogeneous and the IIA no longer applies.

Question Two. The MNL model also predicts the same percentage diversion to the C alternative from both those who chose A in the binary choice situation and those who chose B (because of the "new mode" property). Is that reasonable?

The "new mode" property is arbitrary, but any other assumption often cannot be justified. The original split of individuals between A and B resulted from differing unobserved attributes among a population with identical observed attributes of alternatives and socioeconomic variables. If the observer has no knowledge of why particular individuals chose A and others chose B in the binary choice situation, he certainly has no basis for determining that one individual who originally chose A is more likely than another who originally chose B to be diverted to the third mode.

The assumption of identical percentage reduction in both old mode selection probabilities is arbitrary. But it is based on the fact that knowledge of the choice process for a homogeneous group generally permits no alternative. The researcher may have "used up" all his information on factors systematically affecting choice when he specified the model and this is incorporated in the estimated mode shares for the population. Where all identically situated individuals in a group face the same observed attributes, any attempt to depart from the IIA assumption requires knowledge of the dependence structure of the random utility components which may not be available.

"Belling the Cat"

The assumptions of the MNL model reveal an inherent paradox in asserting axiomatically that the IIA assumption is unreasonable. The model assumes that the actual mode splits are not systematicall affected by unobserved factors and the random components of utility are therefore independent. The paradox of asserting

a violation of the IIA property due to systematic unobserved attributes (in the absence of a statistical test of dependence) is that the researcher must know enough about the unobserved attributes to know that they are systematically affecting choice but not enough to specify the attributes and include them as observed attributes. Frequently, if the observer knows that the random components of utility are not independent because of some systematic unobserved effect, he also may know enough to devise a method to make the systematic unobserved component into an observed component and the offending element of dependence would be eliminated from the unobserved component.

The IIA property is not inherently unreasonable: it should be the goal of the researcher to quantify and include every known systematic effect so that the random effects are independent. If the unobserved utility components are not independent, one solution is to use the knowledge of dependence to improve the model specification so that unobserved components will be random. If the IIA is unreasonable because of the systematic effects of omitted factors, the researcher should modify his behavioral model to consider the systematic omitted factor and thereby make the assumption reasonable. Making the IIA property a reasonable assumption should be an important objective of travel demand modeling. But solving the problem by specifying more systematic influences does not eliminate the IIA property; it merely makes it reasonable.

The IIA property is frequently reasonable because it reflects the researcher's inability to know how omitted factors influence choice. There are, however, cases where a violation of the independence assumption is possible and the dependent omitted factors cannot be measured and included in the model. The blue auto/red auto problem was one such example. The existence of taste variations in the population that are not captured by socioeconomic attributes is another. Generally, these cases must be identified empirically and not on the basis of the "inherent unreasonableness of the IIA assumption." For these cases, the model should be generalized to consider how the structure of dependence among unobserved attributes affects choice. In the next section the cause of such dependence is identified.

4. VIOLATIONS OF THE INDEPENDENCE ASSUMPTION

The MNL model assumes that the unobserved attributes of choice do not vary systematically, i.e., they are independent. Violation of the independence assumption may arise for two important reasons:

Case 1: Correlation of the unobserved components of utility among the alternatives (such as common unobserved attributes among choices); and

Case 2: Observed and unobserved attributes of utility are not independent of one another.

Case 1: Correlation of Unobserved Attributes Across Alternatives

One type of violation of the independence of unobserved attributes is clear. As Figure E-1 illustrates, the correlation of the unobserved utility components is suspected when there is a strong correlation of the observed utility components. It was the correlation of the observed components that made the correlation of the unobserved components easy to detect.

It is clear that we can extend the example to the case where the unobserved attributes are correlated but the observed attributes are not. For example, an "express bus" may have different observed attributes from a "local bus," but similar unobserved attributes. This type of correlation is more difficult to detect than the classic case in Figure E-1.

Table E-1 (above) illustrates that a violation of the IIA assumption in the red auto/blue auto case causes the MNL to underpredict transit ridership. Figure E-2 indicates that the effect of positive correlation of unobserved attributes for auto is to generate observed mode shares for transit that are higher than those predicted by the MNL response curve (using the correct parameter for the attribute) for every level of the bus attribute. The systematic tendency for the MNL model to underestimate the demand for transit in the example directly follows from the specification of the MNL model: adding a new irrelevant

Figure E-1

CORRELATION OF BOTH OBSERVED AND UNOBSERVED COMPONENTS OF UTILITY ACROSS ALTERNATIVES FOR RED AUTO AND BLUE AUTO

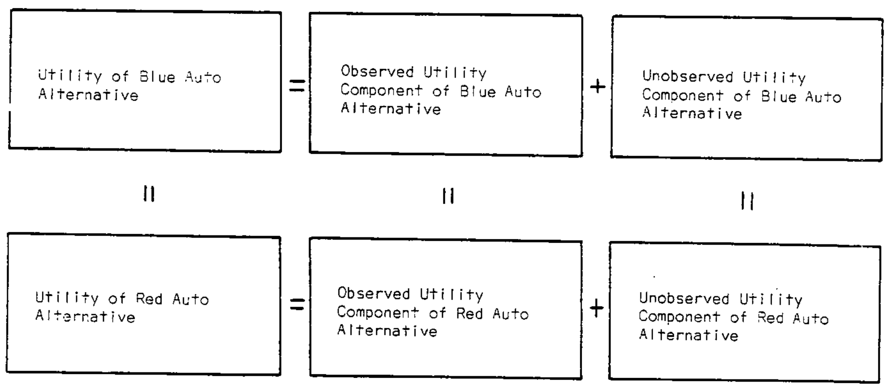
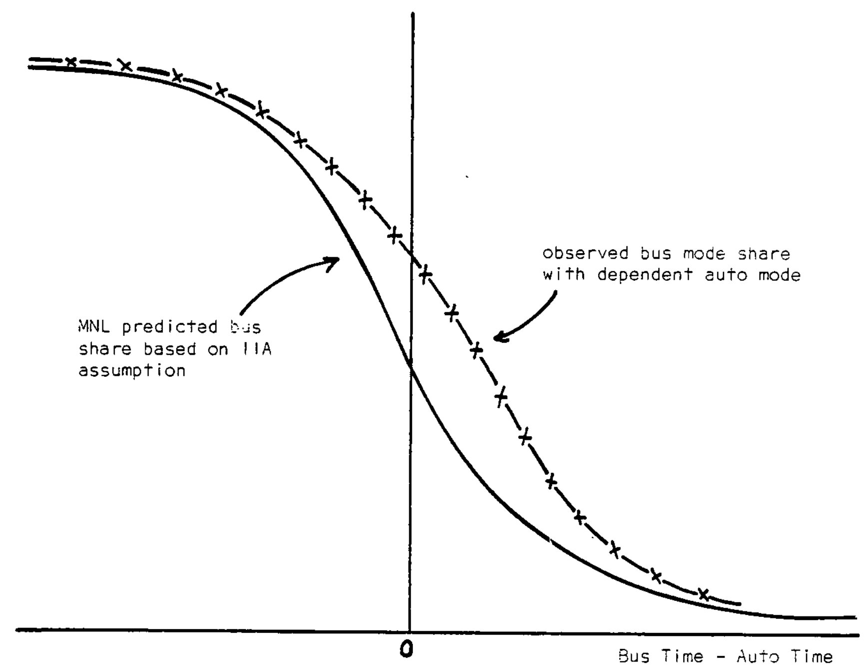


Figure E-2

OBSERVED AND MNL BUS MODE SHARES IN THE RED AUTO/BLUE AUTO EXAMPLE



auto mode must reduce the estimated bus mode share for all levels of bus attributes.

Case 2: Dependence Between Observed and Unobserved Attributes

The MNL model assumes that the observed attributes of the choice alternatives are exogenously determined for each individual (who takes them as given), and that the unobserved attributes (omitted LOS, SEC, and taste variations are independently distributed over alternatives and identically distributed across individuals. This assumption that unobserved attributes are independent of observed attributes can be violated in three ways.

Correlation* Between Observed and Unobserved LOS: Competition Among Modes and Supply Responses. Suppose that in an auto-bus mode split, the transit agency tends for competitive reasons to put newer and more comfortable buses in corridors where the auto has the greatest relative time advantage. In this example, the difference in comfort is dependent on time

*"Correlation" in this context means "empirical dependence" and is meant to apply to both statistical correlation of random variables and functional dependence of a random variable with a nonrandom variable. The usual interpretation of the observed attributes is that they are nonrandom.

differences between the modes. If comfort were an omitted attribute in the logit model and if the true relationship between time differences and mode choice were used for forecasting, the tendency would be to underestimate the demand for transit in the market segment with relatively good transit travel times (see Figure E-3).

On the other hand, if the most comfortable buses were put in the areas where the bus time advantage was best, the results would look like Figure E-4.

Correlation Between Unobserved Tastes and Observed Attributes: "Endogenous Determination" of SEC and LOS. The MNL model assumes that the observed LOS for modes is distributed randomly across the population -- the independent variables are "exogenous." However, consumers generally do not take the travel choice environment as given, but mold it according to their tastes. For example, automobile ownership is ordinarily not taken as a given socioeconomic attribute by the consumer, but is chosen in a manner that is highly correlated with unobserved taste variations that affect mode choice. High levels of auto ownership are statistically related to unobserved tastes which are predisposed to the choice of auto; these tastes affect mode choice in addition to their effect through auto ownership.

The classic case of this "cross-sectional bias" is consumers who make home and workplace location decisions based on the level-of-service offered by their preferred mode. Thus, persons who prefer transit may be more likely to live near transit stops. Conversely, automobile aficionados may choose to live in a place that minimizes highway access time.

Figure E-3

OBSERVED AND MNL BUS MODE SHARES IN THE EXAMPLE WHERE BUS TIME (OBSERVED) AND BUS COMFORT (UNOBSERVED) ARE POSITIVELY CORRELATED

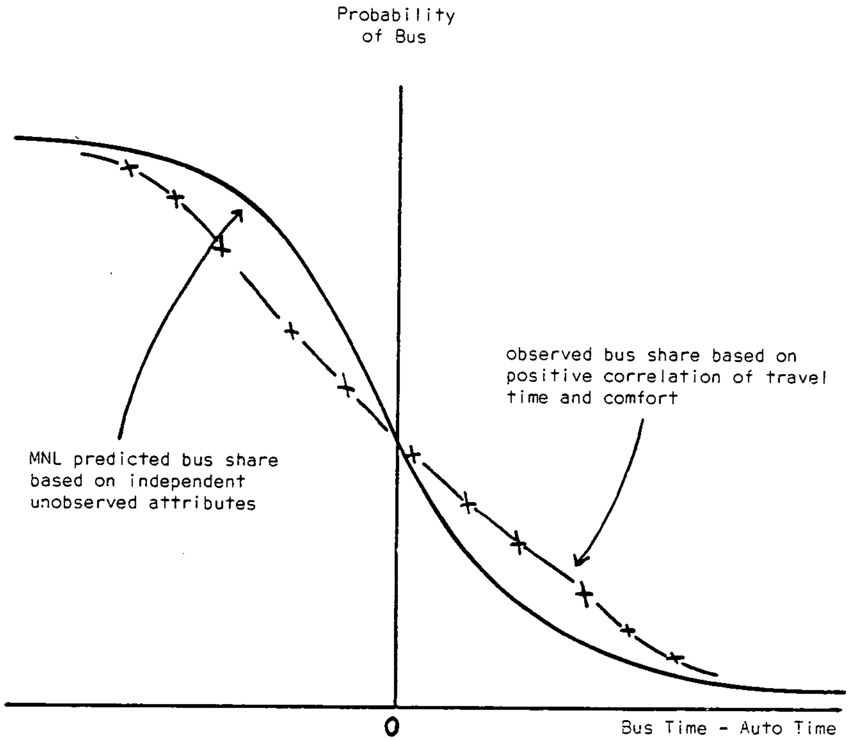


Figure E-4

OBSERVED AND MNL BUS MODE SHARES IN THE EXAMPLE WHERE BUS TIME (OBSERVED) AND BUS COMFORT (UNOBSERVED) ARE NEGATIVELY CORRELATED

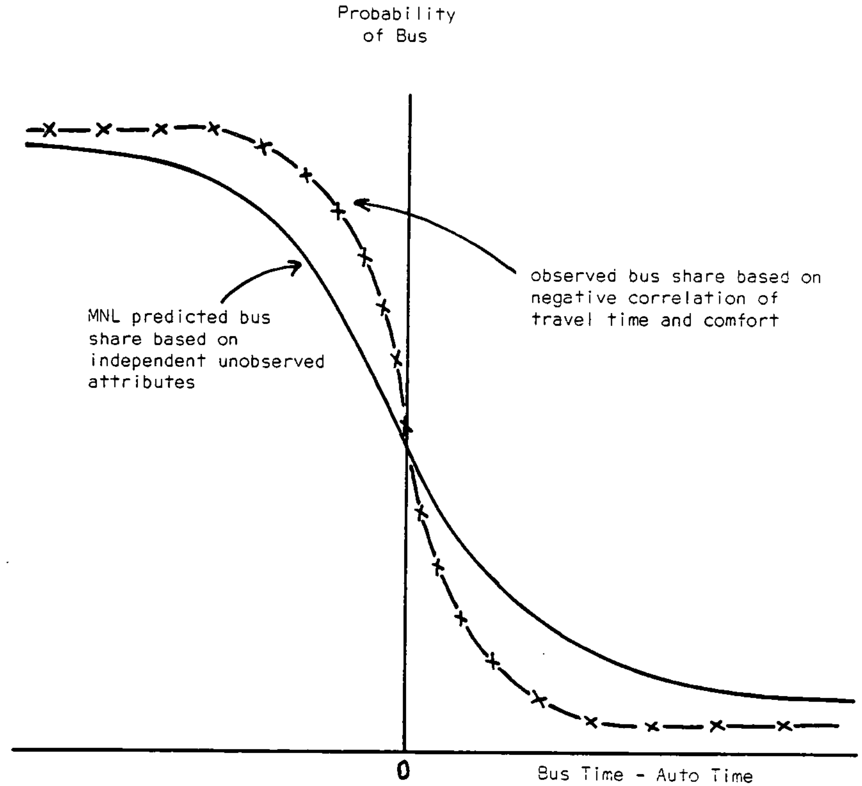


Figure E-5 illustrates an example of dependence of observed and unobserved determinants of choice. As indicated in the example above, consumers may have substantial unobserved variations in taste that lead them to choose travel origins and destinations that maximize LOS on their preferred mode choice alternative. For instance, consumers who strongly prefer the auto for unobserved taste-related reasons embody desirable levels of the observed attributes relative to the other mode choice alternatives, e.g., consumers who strongly prefer the auto for unobserved taste reasons may locate their residences with disregard for the observed attributes of transit service; "transit-prone" consumers will do just the opposite. Persons who face choice alternatives favorable to auto and transit will therefore be "biased" toward those respective choices for reasons other than the level of the observed attributes.

The result of such location decisions is to create dependence between the observed attributes and the unobserved attributes of choice; the observed attributes are "endogeneous." When this happens due to location decisions, the result will be a systematic tendency to underforecast the demand for modes with superior attributes in a given location.*

---

*Assuming that the model was calibrated on data that did not incorporate the correlation of observed and unobserved attributes. Conversely, in model calibration there is a tendency to overestimate consumers' responsiveness to changes in LOS under these circumstances.

An example of the effects of an endogeneous LOS variable is provided in Figure E-6. Assume that persons who have preferences for auto due to unobserved taste variations tend to have relatively high auto LOS (the "auto biased," probably suburban, population) and those who have preferences for transit due to unobserved taste variations tend to have relatively low auto LOS (the "transit biased," probably urban population). The heavy line indicates the mode shares that would be predicted by the MNL model if the unobserved components were independent. However, observed selection frequencies of the "auto biased" population will tend to fall above the MNL response curve in the area of the attribute levels favorable to auto choice. These observations are marked by X's to the right. Similarly, the "transit prone" population will tend to generate observed choice frequencies below the MNL response curve in the region favorable to transit choice.

Incidentally, the figure also illustrates that the MNL calibration procedure will fit a curve (dashed line in figure) with too high a coefficient (too much slope) if taste variations are excluded from the model specification. The population's response to changes in the observed attributes will be exaggerated. In effect, the estimation procedure will be biased upward because the variations in choice due to unobserved taste variations are imputed to the variations in observed LOS with which they are correlated.

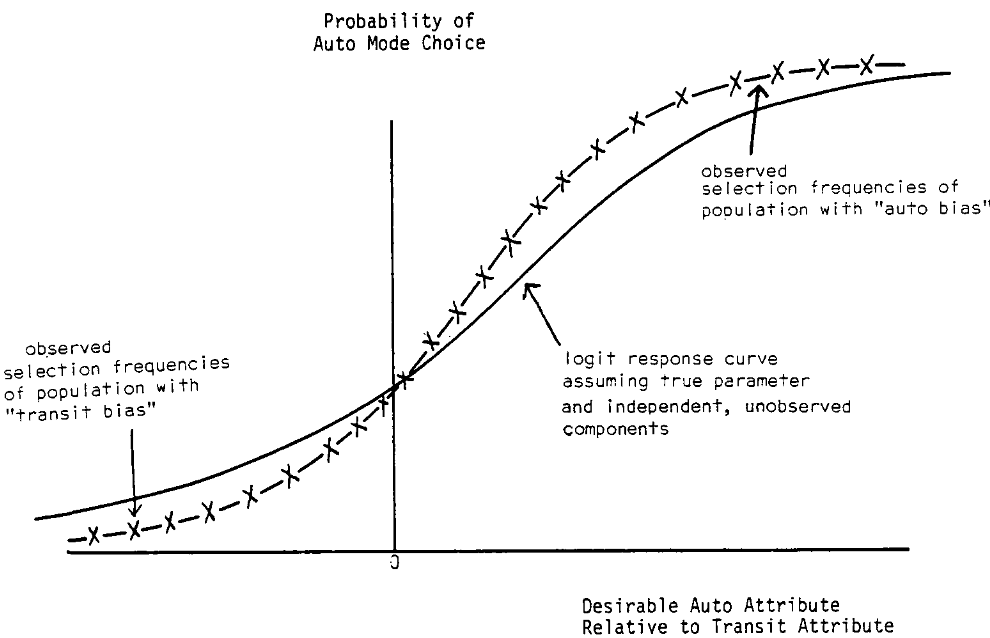
Figure E-5

STRUCTURE OF CORRELATION BETWEEN OBSERVED AND UNOBSERVED UTILITY COMPONENTS

Utility of Auto Alternative	=	Observed Utility Component Auto Alternative Due to Cost	+	Observed Utility Component Auto Alternative Due to Travel Time	+	Unobserved Utility Componen Due to "Taste" for Auto
Utility of Bus Alternative	=	Observed Utility Component Bus Alternative Due to Fares	+	Observed Utility Component Bus Alternative Due to Travel Time	+	Unobserved Utility Componen Due to "Taste" for Transit

Figure E-6

THE EFFECT OF CORRELATION OF OBSERVED AND UNOBSERVED UTILITY COMPONENTS ON OBSERVED MODE CHOICE FOR AUTO



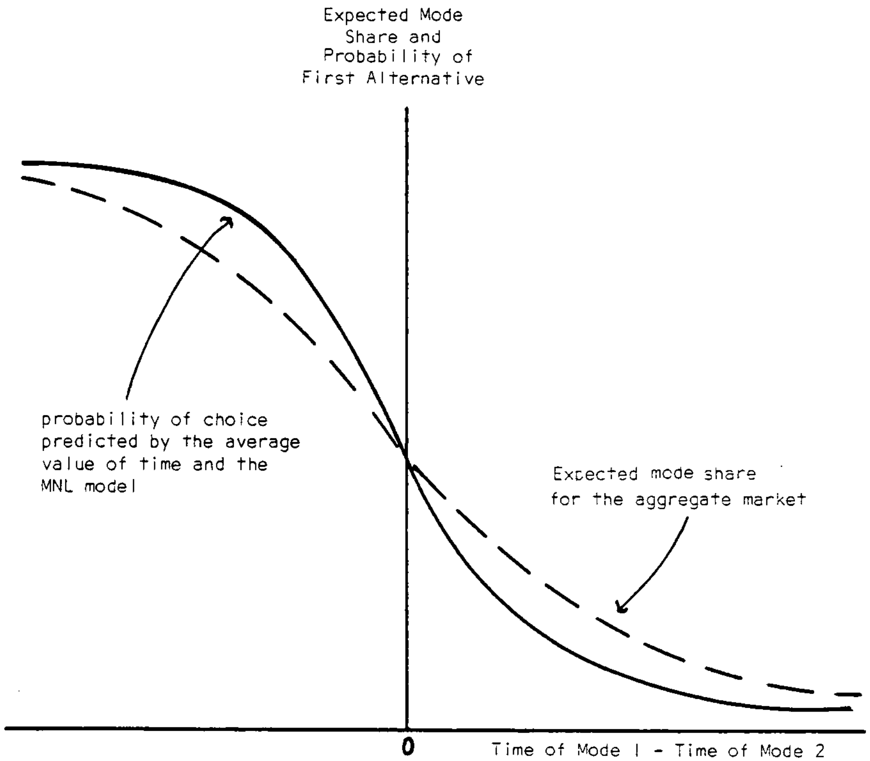
Aggregation Error. It has been demonstrated in Table E-2 that the IIA property is not preserved through aggregation. A special case is where two market segments, e.g., with differing values of time, are combined and treated as a homogeneous group. The result is a violation of the IIA property, and the consequences are formally equivalent to the omission of an attribute which is correlated with included variables, as shown in Figure E-7. In this case the omitted attribute may be thought of as the variation in the value of time for the individual from the average for the group.

This observation is an extension of the "left out variable" problem common to all modeling techniques. That is, an important explanatory variable left out of a model equation will always reduce the explanatory value of the model. However, the results of the model will not be biased unless the omitted variable is correlated with an explanatory variable included in the model. In this case the included variable "picks up" some of the impact of the omitted variable on the dependent variable.

Case 2 dependence should not be confused with the simple case of "mode bias." A "pure mode bias" is the result of an omission of an attribute (or attributes) that systematically varies across the alternatives but is not highly correlated with an observed attribute. For example, "comfort" may be systematically rated higher in the auto mode and not be a quantified element of utility. As long as comfort is not correlated with the observed attributes (say time and cost), the comfort effect is picked up in the mode-specific constant and not imputed to the observed attribute.

Figure E-7

MODE SHARES FOR AGGREGATE OF TWO MARKET SEGMENTS AND THE PREDICTED PROBABILITY FOR INDIVIDUAL WITH AVERAGE VALUE OF TIME



5. CONCLUSIONS ON THE REASONABLENESS OF THE INDEPENDENCE ASSUMPTION

This section briefly describes circumstances that tend to reduce the problem of dependence among the observed attributes of choices and choosers and thereby reduce the severity of IIA violation problems. The existence of at least some degree of dependence among the unobserved components appears almost certain on *a priori* grounds:

(1) Groups of alternatives will have important common omitted variables which vary systematically (Case 1); and

(2) Consumers will locate themselves so that the observed attributes of choices which are preferred for nonobserved taste reasons will rate high; important unobserved attributes will be correlated with observed attributes; the sample will not have identical tastes toward observed attributes; and consumers will not take as given certain socioeconomic variables such as auto ownership (Case 2).

The above comments suggest that independence is likely to be a common occurrence. Fortunately, a number of factors operate to help reduce the correlation in many situations.

Multinomial Choice Situations May Tend to Break Up the Correlation

Posing the problem of IIA as a "new mode" problem suggests that it can occur because adding new modes may create correlation between the unobserved components. The commonly cited example leads to the presumption that the situation is most likely to occur in situations of many options. However, adding further choice alternatives may tend to break up the dependence. Thus, in many cases a large number of independent alternatives will often tend to minimize the errors due to systematic correlation of unobserved attributes in two alternatives.

This very important point contravenes the intuitive impression that modeling a choice among many alternatives is more likely to create IIA problems than modeling a few alternatives. Modeling a binary choice is no guarantee that the independence assumption will not be violated. Neither can one assume that increasing the number of alternatives necessarily exacerbates the problem. In fact, we may speculate that more alternatives may help break up the structure of dependence.

Correlation of Observed Attributes Does Not Necessarily Imply Correlation of Unobserved Attributes

The presence of correlation in the observed components of two alternatives by no means is sufficient to guarantee correlation of unobserved components. In fact, the independence of the unobserved components may be the very reason for the existence of the two alternatives. For example, two shopping destinations may have similar observed components (such as distance and size), but be completely independent in unobserved utility components.

Correlated Unobserved Attributes May be Unimportant Relative to Independent Unobserved Attributes

The presumption that a dominant unobserved attribute (e.g., comfort) will emerge to violate the independence assumption often will not be true. Positively correlated unobserved attributes often tend to be balanced out by negatively correlated unobserved attributes. An attribute which is positively evaluated by some subjects will be negatively evaluated by others.

The Distribution of the Attributes Among the Population May Mitigate the Effects of Dependence

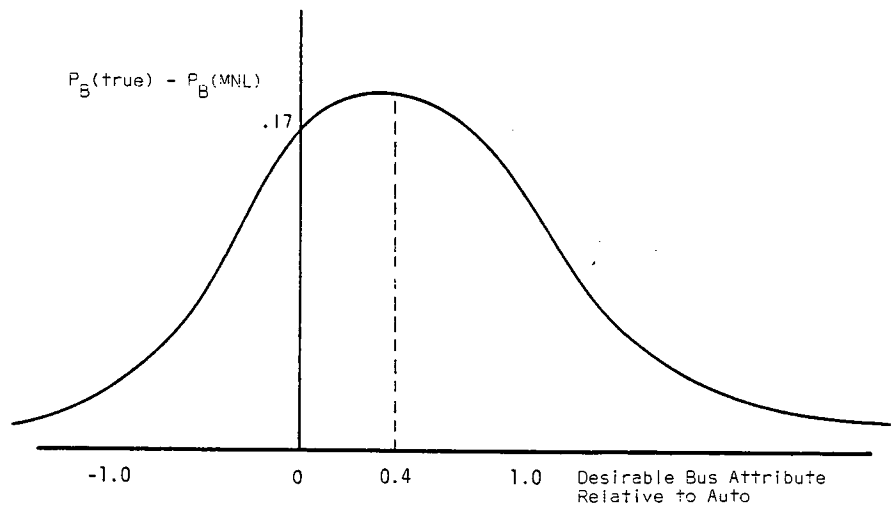
The forecasting error due to biased parameter estimates greatly depends on where on the logit response curve the forecast is required. Figure E-8 illustrates that the absolute error in forecasting a mode share for Case 1 is low in the tails. If the mode share is large, the percentage error will be small as well.

<u>Much of the Bias Due to the Omission of Systematic Attributes will be Absorbed by the "Mode-Specific Constant"</u>

In the Case 1 violation, the mode-specific constant is the most "dominant" trait among the modes and tends to absorb much of the bias. Therefore, the behavioral coefficients on attributes will be less affected.

Figure E-8

UNDERESTIMATE OF THE MNL RESPONSE CURVE
IN THE BLUE AUTO/RED AUTO EXAMPLE



SOURCE: Charles River Associates, (<u>E1</u>).

REFERENCES

E1. Charles River Associates. "Disaggregate Travel Demand Models, Project 8-13: Phase I Report." Two volumes. Prepared for National Cooperative Highway Research Program, National Research Council, February 1976.

E2. McFadden, D. L., Tye, W., and Train, K. "Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model." Working Paper No. 7616, Travel Demand Forecasting Project, Institute of Transportation Studies, University of California at B rkeley, 1976.

E3. McFadden, D. L. "The Mathematical Theory of Demand Models." In P. R. Stopher and A. H. Meyburg, Eds. Behavioral Travel-Demand Models. Lexington, MA, D.C. Heath, 1976.

E4. Horowitz, Joel. "Sources of Error and Uncertainty in Behavioral Travel Demand Models." In P. R. Stopher, A. H. Meyburg, and W. Brog, Eds., New Horizons in Travel-Behavior Research. Lexington, MA, D.C. Heath, 1981.

E5. Block, H. and Marschak, J. "Random Orderings and Stochastic Theories of Response." In I. Olkin, Ed., Contributions to Probabilities and Statistics. Stanford, Cal., Stanford University Press, 1960.

E6. McFadden, Daniel. "Quantal Choice Analysis: A Survey." Annals of Economic and Social Measurement, Vol. 5, pp. 363-390, 1976.

E7. McFadden, Daniel. "Conditional Logit Analysis of Qualitative Choice Behavior." In P. Zarembka, Ed., Frontiers in Econometrics. New York, N.Y., Academic Press, 1974.

E8. Ben-Akiva, Moshe E. "Structure of Passenger Travel Demand Models." Ph.D. Dissertation, Department of Civil Engineering, Massachusetts Institute of Technology, 1973.

APPENDIX F

A MULTINOMIAL LOGIT MODEL WHICH PERMITS VARIATIONS IN TASTES ACROSS INDIVIDUALS

Note: This appendix appeared earlier as a CRA working paper by N. Scott Cardell and Bernard J. Reddy.

INTRODUCTION

The multinomial logit model has proved to be quite useful in analyzing a number of transportation problems, such as the effects of introducing new travel alternatives or of altering the characteristics (such as price or frequency of service) of existing alternatives. The multinomial logit model does, however, have some drawbacks, most of which can be traced back to the independence of irrelevant alternatives assumption.

A discrete choice model developed recently by researchers at Charles River Associates (the CRA hedonics model) extends the logit model to permit variations in tastes across individuals. This model generalizes the work of Quandt (F1), extends the logit work of McFadden (F2), Domencich and McFadden (F3), and others, and is closely related to recent work on the multinomial probit model by Hausman and Wise (F4), Albright et al., (F5), and Daganzo, Bouthelier and Sheffi (F6).

The disaggregate logit, probit, and CRA hedonics models are based on the assumption that an individual faced with a choice among competing alternatives selects the one that yields the highest level of utility. All three models further assume that the individual utility function is linear in the attributes of

the alternatives, that these attributes can be level of service variables or interactions between level of service variables and household characteristics, and that the utility function also contains an additive error term. That is, the utility an individual receives from alternative i can be written as:

$$U_i = \sum_{k=1}^{K} X_{ik} \beta_k + \varepsilon_i, \qquad \text{(F-1)}$$

where:

U_i = the utility of alternative i;

X_{ik} = the attributes;

β_k = the unknown parameters; and

ε_i = the error term for alternative i.

Unless otherwise stated in this paper, the superscript denoting a given individual will be suppressed.

The three models diverge at this point. The logit model assumes that the β_k are constant across individuals and that the ε_i are independent Weibull random variables. This last assumption, as McFadden has demonstrated, is equivalent to the independence of irrelevant alternatives assumption. The probit model assumes that the ε_i are multivariate normal random variables, and the versions we will discuss also assume that the β_k are normal random variables that are independent of the ε_i. This implies for a given individual that the β_k can still be considered fixed across alternatives. The multivariate normal assumption for the ε_i implies that these error terms need not be independent. The CRA hedonic model assumes that the ε_i are independent Weibull random variables, just as the logit model does, but it further assumes that the β_k are random variables with any specified, well behaved distribution.

IMPLICATIONS OF RANDOM PARAMETER MODELS

Before examining the complexities that the random parameter models introduce, a discussion of their relevance is in order. What advantage does a model that explicitly allows for variations in tastes offer over the logit model, which has proved so valuable in applied transportation research?

A good argument can be made a priori for the proposition that tastes do vary across a population. Some people may weight their transit times much more heavily than their transit costs, while others may not. To the extent that this behavior is related to factors such as household income, the specification of the logit model may capture some of these effects. However, it may not capture them all.

These variations in tastes have important implications for the logit model, because they imply a failure of the independence assumption. For example, suppose that a given individual places a higher than average weight on out-of-vehicle travel time (OVTT), and that this individual is faced with a choice of two transit modes and one auto mode for travel to work. Suppose further that the two transit modes have similar OVTTs that are higher

than that for the auto mode. The higher than average OVTTs and the higher than average weight on OVTT would combine to create a large common component in the error terms for the two transit modes. Statistically this translates into correlation between the error terms, causing the independence assumption to be violated.

This can be seen algebraically quite easily. Write β_k as $\bar{\beta}_k + \tilde{\beta}_k$, where $\bar{\beta}_k$ is the expected value of β_k and $\tilde{\beta}_k$ is its deviation from the mean. Then the individual utility from alternative i can be written as:

$$U_i = \sum_{k=1}^{K} X_{ik}\,\bar{\beta}_k + \left(\sum_{k=1}^{K} X_{ik}\tilde{\beta}_k + \varepsilon_i\right), \qquad \text{(F-2)}$$

where the term in parentheses is the total error term for alternative i. Call this term e_i. Assuming that the $\tilde{\beta}_k$ are independent of the ε_i, the covariance between e_i and e_j can be written as:

$$\begin{aligned} Cov(e_i, e_j) &= E\left(\sum_{k=1}^{K} X_{ik}\,\tilde{\beta}_k + \varepsilon_i\right)\left(\sum_{k=1}^{K} X_{jk}\,\tilde{\beta}_k + \varepsilon_j\right) \\ &= \sum_{k=1}^{K}\sum_{m=1}^{K} X_{ik}\,X_{jm}E(\tilde{\beta}_k\tilde{\beta}_m) + E(\varepsilon_i\varepsilon_j). \end{aligned} \qquad \text{(F-3)}$$

If ε_i is independent of ε_j and the $\tilde{\beta}_k$ are independent, this simplifies to:

$$Cov(e_i, e_j) = \sum_{k=1}^{K} X_{ik}X_{jk}\, E(\tilde{\beta}_k^{\,2}). \qquad \text{(F-4)}$$

The term $E(\tilde{\beta}_k^{\,2})$ is simply the variance of β_k. If the parameters of the individual utility function are not fixed, these variances will be positive; so the covariance between the total error terms for any two alternatives generally will be nonzero. If the β_k's are not independent, the covariance expression will have more terms but still will be generally nonzero.

How important are variations in taste likely to be in practice? Hausman and Wise (F4) estimated a simple transportation demand model in which a statistically significant variation in parameters is measured, particularly for OVTT and for travel cost relative to income. CRA (F7, F8) has estimated a demand model of choice among automobile models in which variations in the marginal utilities are substantial, particularly those for automobile turning radius. In addition, both CRA and Hausman and Wise present examples in which the random parameter models produce forecasts that are not only very different from those of the logit model, but also more believable.

A summary of some forecast results obtained with the CRA hedonics model is presented in Table F-1. An aggregate model of the market shares of automobile models was estimated and an artificial aggregate forecast was made, based loosely on the results of the estimation. The automobile buying public was assumed to be faced originally with a choice among 20 automobile

Table F-1

SUMMARY OF MARKET SHARES WITH A NEW ALTERNATIVE

(Percent of Total)

Alternative	Original Share	New Share: Logit	New Share: CRA
1	-	19.68	13.02
2	24.50	19.68	13.02
Others	75.50	60.64	73.96

SOURCE: Charles River Associates. "Methodology for Predicting the Demand for New Electricity Using Goods." Cambridge, MA, 1977, p. 26.

models. When a new alternative (Number 1) identical to the second alternative is introduced, the market shares of these two alternatives are the same, in both the logit and the CRA hedonics models. However, the total market share of the first two alternatives increases considerably in the logit model but only slightly in the CRA hedonics model. This indicates that in some cases the CRA hedonics model can cope well with the red auto-blue auto problem. The reason that the combined market shares of the models in question increase at all is that the CRA hedonics model still makes the independence of irrelevant alternatives assumption at the individual level. That is, the CRA hedonics model assumes that the characteristics of the new good are identical with those of the existing good, but that the error terms for the two are independent. This is likely to be a reasonable assumption in some instances but not in others. The probit model can assume that the error terms are identical, independent, or correlated.

Random parameter models seem best suited for applications to problems in which an individual is faced with a number of goods with widely differing, easily quantified characteristics. Automobile market shares illustrate this point well. Purchasers of automobiles in the United States are faced with a choice among over 100 models, including major options like engine size. Many automobile characteristics can be quantified easily, such as gas mileage, acceleration, turning radius, and interior

room. Others cannot, such as "luxury" and amount of soundproofing. In all, however, choice among automobile models seems to be an excellent application for random parameter models.

Many problems in travel demand also seem well-suited for random parameter models such as the mode choice for worktrip model estimated by Hausman and Wise (F4). Travel modes offer individuals tradeoffs among such variables as in-vehicle travel time, out-of-vehicle travel time, and travel costs. It seems reasonable a priori that different individuals might weigh the importance of these variables in different ways, and the estimates presented by Hausman and Wise support this view.

COMPLEXITY OF LOGIT AND RANDOM PARAMETER MODELS

Now that the value of the random utility models has been demonstrated, our attention will turn to the computational differences between these models and the logit model.

In all three models the probability that an individual will choose alternative i is the probability that the utility from alternative i is greater than the utility from the other alternatives. That is,

$$P_i = P_r(U_i > U_j, \ \forall j \neq i). \qquad \text{(F-5)}$$

All three models use these choice probabilities to estimate the parameters with maximum likelihood techniques. Some of the computational advantages and disadvantages of the three models can be observed immediately by examining the expressions for the choice probabilities. For the logit model the probability that alternative i will be chosen can be written simply as:

$$P_i = \frac{e^{\sum_{k=1}^{K} X_{ik}\beta_k}}{\sum_{j=1}^{n} e^{\sum_{k=1}^{K} X_{jk}\beta_k}} \qquad \text{(F-6)}$$

The corresponding expression for the probit model is conceptually straightforward, although much more difficult to evaluate:

$$P_i = \int_{-\infty}^{\infty} de_i \int_{-\infty}^{X_i\bar{\beta}-X_j\bar{\beta}+e_i} de_j \ldots \int_{-\infty}^{X_i\bar{\beta}-X_n\bar{\beta}+e_i} \phi(e_i,\ldots,e_n;\Sigma)d\, e_n \qquad \text{(F-7)}$$

where ϕ is the multivariate normal density function with mean zero and variance-covariance matrix Σ. Although this expression can be simplified somewhat, it requires the evaluation of a multivariate normal integral of order N-1 (where N is the number of alternatives). Hausman and Wise (F4) evaluate this integral directly, while Albright et al. (F5) have suggested a Monte Carlo approach, and Daganzo, Bouthelier, and Sheffi (F6) have suggested an approximation method first studied by Clarke (F9) in another context.

The CRA hedonics model results in a form with elements of both of these expressions. First define:

$$P_i(\beta) = \frac{e^{\sum_{k=1}^{K} X_{ik}\beta_k}}{\sum_{j=1}^{n} e^{\sum_{k=1}^{K} X_{jk}\beta_k}} \qquad \text{(F-8)}$$

The variable $P_i(\beta)$ is simply the logit choice probability, given that the parameter vector is β. The choice probability for the CRA model then can be written as:

$$P_i = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdot\cdot\cdot \int_{-\infty}^{\infty} P_i(\beta)\, f(\beta_1, \ldots, \beta_K)d\beta_1 \ldots d\beta_K, \qquad \text{(F-9)}$$

where $f(\beta_1,\ldots,\beta_K)$ is the probability density function of the parameters of the individual utility function. This expression implies that the choice probability in the CRA hedonics model is simply the expected value of the choice probability of the logit model, where the expectation is made over the parameters. As a result, the logit model is a special case of the CRA hedonics model. The expression above requires the evaluation of a multivariate integral of order K, the number of parameters. If the parameters of the utility function are constant, the choice probabilities in the CRA hedonics and logit models are identical.

The Monte Carlo methods used to estimate the CRA model will be described briefly. The researcher specifies a distribution function for the parameter vector β, such as normal, log-normal, or exponential. Any distribution that is sufficiently well-behaved so that the cumulative distribution function can be calculated easily can be used. Many otherwise ill-behaved distributions, such as the Cauchy, could be specified. After the distribution is specified, a number of drawings from that distribution are made, based on the values of the means, variances, and covariances of the β's in the current iteration. The $P_i(\beta)$ are used to approximate the choice probabilities P_i. This approach is conceptually and computationally straightforward, although it is somewhat time consuming.

PROPERTIES OF THE CRA HEDONICS MODEL

Although the CRA hedonics model makes the independence of irrelevant alternatives assumption at the individual level, the fact that the coefficients of the attributes can vary across individuals enables the CRA hedonics model to avoid several of the problems of the logit model. The red auto-blue auto problem already has been mentioned and will not be discussed further. An implication of the logit model is that at the individual

level the cross-elasticities of the choice probabilities are equal. That is,

$$\frac{\partial \log P_i}{\partial \log X_{jk}} = \frac{\partial \log P_\ell}{\partial \log X_{jk}} = -\beta_k P_j X_{jk} \qquad \text{(F-10)}$$

$$\text{for } j \neq i, \ell$$

This is not true of the CRA model, in which the cross elasticity is:

$$\frac{\partial \log P_i}{\partial \log X_{jk}} = \frac{-\beta_k X_{jk} \int_{-\infty}^{\infty} P_i(\beta)\, P_j(\beta)\, f(\beta) d\beta}{\int_{-\infty}^{\infty} P_i(\beta)\, f(\beta) d\beta} \qquad \text{(F-11)}$$

where the integral is the K dimensional integral described earlier. This expresision is not independent of alternative i, so the cross elasticities generally are not equal.

Because the probit and CRA hedonics models both permit variations in tastes across individuals, a comparison of some of their properties is in order. First, the CRA hedonics model assumes that the ε_i, the component of the error term which does not result from the randomness of the parameters, is independently distributed across alternatives. The probit model permits the ε_i to be correlated, which means that in theory the probit model can handle virtually all causes of failure in the independence assumption. It is difficult to assess the practical significance of this greater generality of the probit model. The random parameter feature of the CRA hedonics model can account for an important type of correlation across error terms, but it cannot account for all. For example, suppose all transit modes are uncomfortable, but no measure of comfort appears as a level of service variable. Then this unobserved variable problem will cause the error terms for the transit modes to be correlated, but the CRA model will not capture this correlation and therefore its use will be inappropriate. In theory the probit model can be used to calculate this correlation. If such correlations do exist, then the probit model or different generalizations of the logit model should be used. A version of the logit model which permits such correlations across alternatives has been developed by one of the authors of this appendix. This version is essentially an error components method for the Weibull distribution and is compatible with the CRA hedonics model.

This problem, however, may be more theoretical than practical. If the probit model is estimated with nonzero correlations between the ε_i, forecasts of the model when a new alternative is introduced will have to specify the correlation of the ε for the new alternative with those for the existing alternatives. In practice, many users of the probit model are

likely to assume that the ε_i are uncorrelated, thereby avoiding this problem. If the ε_i are assumed to be uncorrelated, then the probit model offers no more generality in this regard than does the CRA hedonics model. If, on the other hand, there is reason to believe that the ε_i are correlated, then the CRA hedonics model is not applicable.

Another area of interest is the specification of the distribution of the parameters of the individual utility function. The probit model requires that they be normal random variables, while the CRA model permits them to have any well-behaved distribution. This is a clearcut advantage of the CRA model, for a variety of reasons. A normal random variable can be either positive or negative, so the probit assumption implies that parameters will be positive for some individuals and negative for others. This may not be unreasonable in some cases, but it is difficult to imagine situations in which an individual enjoys spending money on transportation to work. In some cases this problem will be minimal. The estimated variance of the distribution of a parameter might be very small relative to the estimated mean, which would imply that the parameter in question would have the "wrong" sign for only an insignificant fraction of the sample. In other cases this problem will be more severe, particularly those in which parameter distributions are highly skewed with relatively small means.

It appears that the parameter assumptions of the CRA hedonics model are much more flexible than those of the probit model.

The assumption in the CRA hedonics model that the ε_i are independent Weibull variables rather than independent normals, as they are in the probit models most likely to be estimated, should matter little. The normal and Weibull distributions are quite similar, except in the extreme tails, so there is little reason on theoretical or empirical grounds to select one distribution over the other. Hausman and Wise (F4) compare mode share forecasts based on a logit model and a probit model in which the total error terms are independent, that is, in which the β_k are fixed and the ε_i are independent. Both before and after the introduction of a new mode, the modal split forecasts of the two models were virtually identical. This indicates that if the ε_i are independent there may be little reason to assume that they have one distribution rather than the other.

Large-scale comparisons of the costs of estimation with the probit and CRA models have not been made, but some evidence is available. The two most promising versions of probit are those of Hausman and Wise (F4) and of Daganzo *et al.* (F6). The Hausman-Wise approach is highly accurate but somewhat limited, because computation costs increase rapidly with the number of alternatives. The Daganzo *et al.* approach reduces the computation costs considerably, but the accuracy may decline with a large

number of alternatives, because the quality of the successive approximations deteriorates as the number of alternatives increases. The computation costs of the CRA model do not depend greatly on the number of alternatives, but they increase rapidly as the number of parameters increases. The CRA model would have advantages over the Hausman-Wise model in problems with a large number of alternatives (more than five) and a small number of parameters (less than seven) while the Hausman-Wise model would have advantages in problems with a large number of parameters but a small number of alternatives. Both models would be quite expensive to estimate on problems with a large number of both alternatives and parameters.

POSSIBLE FUTURE RESEARCH

At present the CRA hedonics model has been programmed only for application to aggregate logit problems (i.e., market shares), with the parameters of the utility function estimated as independent normal or log-normal variables. Current work on the model includes reprogramming it for application to disaggregate data and expanding the options for specifying the distributions of the parameters. This work is relatively straightforward and should meet no major difficulties.

One of the most important uses of the CRA hedonics model might be for problems in which the individual utility function is nonlinear in its parameters. For example, economic theory might imply that household income and the monetary cost associated with alternative i should appear in the individual utility function as a composite variable of the form $\beta_1 \log \frac{\beta_2 Y - P_i}{\beta_2 Y}$. The parameters β_1 and β_2 could be specified as log-normal variables (or whatever) and estimation would take place exactly as described earlier, with only a few minor differences.

In general the individual utility function can be written as

$$U_i = V(X_i, \beta) + \varepsilon_i, \tag{F-12}$$

in which case the choice probability conditional on the parameters then becomes

$$P_i(\beta) = \frac{e^{V(X_i,\beta)}}{\sum_{j=1}^{n} e^{V(X_j,\beta)}} . \tag{F-13}$$

The expression for the choice probability does not change, remaining

$$P_i = \int_{-\infty}^{\infty} P_i(\beta)\, f(\beta) d\beta, \tag{F-14}$$

except for the different construction of the $P_i(\beta)$. The only other changes required in the model are those concerning the derivatives of the likelihood function with respect to the parameter means and variances. These must take explicit account of any nonlinearities in the individual utility function.

Another area for further work would be the combination of the CRA hedonics model with a generalized version of the

multinomial logit model which permits correlations between the ε_i. Such a model would offer as much or more generality in estimation than would the probit model. Its computation costs might exceed those of the approximation method for estimating the multinomial probit model, but its accuracy is likely to be greater as well.

The development of models as sophisticated as the multinomial probit and CRA hedonics models highlights the importance of collecting data sets of sufficiently high quality to take advantage of the sophistication of these models. Similarly, dynamic adjustment or diffusion processes should be studied in more detail, particularly as they concern the choices of residence location, automobile ownership, and travel mode.

REFERENCES

F1. Quandt, R. E. "Estimation of Modal Splits." Transportation Research, Vol. 2, 1968, pp. 41-50.

F2. McFadden, Daniel. "Conditional Logit Analysis of Qualitative Choice Behavior." In P. Zarembka, Ed., Frontiers in Econometrics. New York, Academic Press (1974).

F3. (Domencich, T. and McFadden, D.), Urban Travel Demand: A Behavioral Analysis. Amsterdam, North-Holland Publishing Co. (1975).

F4. Hausman, J. and Wise, D. "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." Econometrica, Vol. 46, pp. 403-426, 1978.

F5. Albright, R. L., Lerman, S. R., and Manski, C. F. "Report on the Development of an Estimation Procedure for the Multinomial Probit Model." Prepared for the Federal Highway Administration, Cambridge, MA, Cambridge Systematics, Inc., 1977.

F6. Daganzo, C. F., Bouthelier, F., and Sheffi, Y. "Multinomial Probit and Qualitative Choice: A Computationally Efficient Algorithm." Transportation Science, Vol. 11, pp. 338-358, 1977.

F7. Charles River Associates. "Impact of Trade Policies on the U.S. Automobile Market." Cambridge, MA, 1976.

F8. Charles River Associates. "Methodology for Predicting the Demand for New Electricity-Using Goods." Cambridge, MA, 1977.

F9. Clarke, C. "The Greatest of a Finite Set of Random Variables." Operations Research, Vol. 9, pp. 145-162, 1961.

THE TRANSPORTATION RESEARCH BOARD is an agency of the National Research Council, which serves the National Academy of Sciences and the National Academy of Engineering. The Board's purpose is to stimulate research concerning the nature and performance of transportation systems, to disseminate information that the research produces, and to encourage the application of appropriate research findings. The Board's program is carried out by more than 270 committees, task forces, and panels composed of more than 3,300 administrators, engineers, social scientists, attorneys, educators, and others concerned with transportation; they serve without compensation. The program is supported by state transportation and highway departments, the modal administrations of the U.S. Department of Transportation, the Association of American Railroads, and other organizations and individuals interested in the development of transportation.

The Transportation Research Board operates within the National Research Council. The National Research Council was established by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and of advising the Federal Government. The Council operates in accordance with general policies determined by the Academy under the authority of its congressional charter of 1863, which establishes the Academy as a private, nonprofit, self-governing membership corporation. The Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in the conduct of their services to the government, the public, and the scientific and engineering communities. It is administered jointly by both Academies and the Institute of Medicine.

The National Academy of Sciences was established in 1863 by Act of Congress as a private, nonprofit, self-governing membership corporation for the furtherance of science and technology, required to advise the Federal Government upon request within its fields of competence. Under its corporate charter the Academy established the National Research Council in 1916, the National Academy of Engineering in 1964, and the Institute of Medicine in 1970.

TRANSPORTATION RESEARCH BOARD
National Research Council
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

ADDRESS CORRECTION REQUESTED