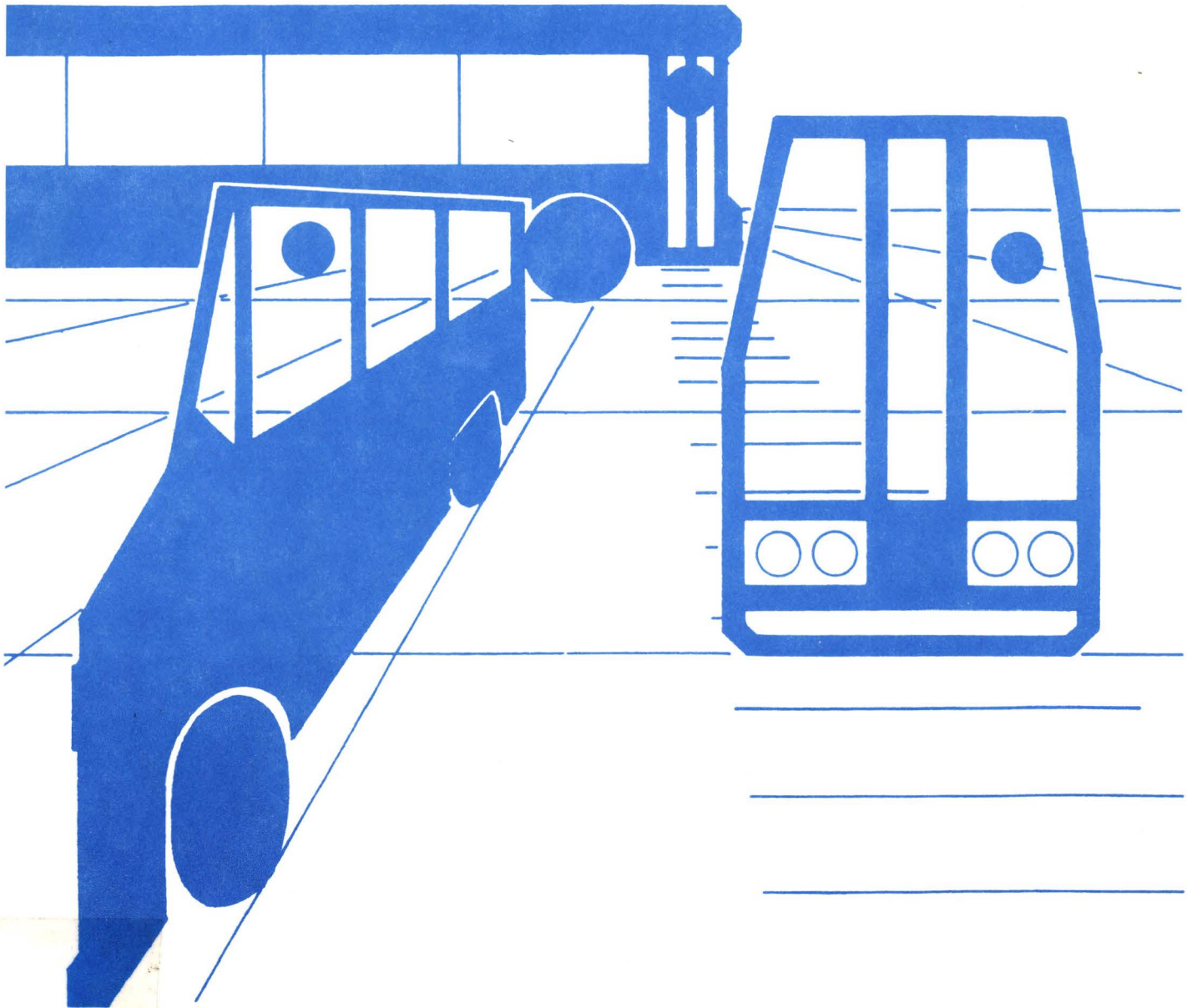




U.S. Department
of Transportation

A Self-Instructing Course in Disaggregate Mode Choice Modeling



FEDERAL TRANSIT ADMINISTRATION

5
5
7
86

HE
336
.C5
H67
1986

SP 27 33

16095

A Self-Instructing Course in Disaggregate Mode Choice Modeling

Final Report
December 1986

Prepared by

Joel L. Horowitz
University of Iowa
Iowa City, Iowa 52242

and

Frank S. Koppelman
Northwestern University
Evanston, Illinois 60201

and

Steven R. Lerman
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Prepared for

University Research and Training Program
Urban Mass Transportation Administration
(now Federal Transit Administration)
Washington, D.C. 20590

Distributed in Cooperation with

Technology Sharing Program
U.S. Department of Transportation
Washington, D.C. 20590

DOT-T-93-18

CONTENTS

Module 1: Introduction	1
1.1 The Motivation for This Course	1
1.2 Description of the Course	3
1.3 Summary	6
References	7
Module 2: Introduction to Choice Theory	8
2.1 Introduction	8
2.2 The Role of Choice in Generating Travel Demand	8
2.3 Preferences	10
2.4 Utility Theory	11
2.5 Properties of Utility Functions and Utility Models	15
2.6 Predictions of Aggregate Travel Behavior	17
2.7 Summary	19
Exercises	20
Module 3: Introduction to Probabilistic Choice Theory	23
3.1 Introduction	23
3.2 Inadequacy of Deterministic Utility Models	23
3.3 Limitations of Analysts' Information	25
3.4 Examples of Unexplained Variation in Choice Behavior	28
3.5 The Basic Formulation of Probabilistic Models	37
3.6 The Probability of Observing a Specific Choice	43
3.7 Aggregate Prediction with Probabilistic Choice	44
3.8 Summary	47
Exercises	48
Module 4: The Logit Choice Model	50
4.1 Introduction	50
4.2 The Binomial Logit Model	51
4.3 The Multinomial Logit Model	57
4.4 Application of the Multinomial Logit Model to Mode Choice Analysis	59
4.5 Incorporation of Attributes of Alternatives and Individuals	61
4.6 Alternative-Specific Constants	63
4.7 Independence from Irrelevant Alternatives	66
4.8 Summary	80
Exercises	80

Module 5: Variables of Mode Choice Models	82
5.1 Introduction	82
5.2 Classes of Variables That Must Be Included in Models	83
5.3 Functional forms and Disaggregation of Variables	94
5.4 Other Variables	107
5.5 The Selection of Choice Sets	109
5.6 Summary	112
Exercises	117
Module 6: Estimating the Utility Functions of Choice Models	119
6.1 Introduction	119
6.2 Acquisition of Data	120
6.3 Specifying the Model	123
6.4 Estimation -- The Maximum Likelihood Method	124
6.5 Interpreting the Estimation Results -- Testing the Model	128
6.6 Some Estimation Problems and How to Avoid Them	142
6.7 Conclusions	145
Exercises	146
Module 7: Prediction with Disaggregate Models	150
7.1 Introduction	150
7.2 Reasons for Estimating Disaggregate Models	152
7.3 Relation between Individual Choices and Mode Shares	165
7.4 Methods for Aggregate Prediction with Disaggregate Choice Models	168
7.5 Summary	175
Solutions to Exercises	177

MODULE 1

INTRODUCTION

1.1 The Motivation for This Course

Many practical transportation policy issues are concerned with mode choice. For example, the gain or loss in transit revenues caused by a fare increase depends on how travelers' mode choices are affected by the increase. If few current transit riders switch to other modes because of the fare increase, transit revenues will increase proportionally to the increase in fare. But if many riders switch to other modes, revenues will increase less than proportionally to the fare increase and may decrease. Similarly, the effects of changes in transit routes and schedules on ridership, revenues, and traffic congestion all depend on how the changes affect individual travelers' mode choices. The effectiveness of programs to encourage ridesharing -- for example, preferential parking or preferential access to freeways for carpools -- also depends on how the programs affect mode choice. In most situations, planners must choose among a variety of fare schedules and service designs. An understanding of the separate and combined effects of these decisions on travel mode choice is essential to selection of the best plan to meet specific transportation objectives.

The importance of mode choice in transportation policy analysis and decision making has led to a variety of methods for predicting the effects of policy measures on travelers' mode choices. Two well-known and frequently used prediction methods are the method of elasticities and aggregate mode split modeling. Both of these methods have serious defects that greatly restrict their practical usefulness. For example, the method of elasticities cannot predict accurately the effects of making several

changes in transit service simultaneously (e.g., of increasing both the fare and the schedule frequency or of adding a new route to the system). Aggregate mode split models can be exceedingly costly and cumbersome to develop. Moreover, they are subject to serious biases and prediction errors owing to their reliance on aggregate travel data rather than records of individual trips. The range of policy questions that can be treated with aggregate models is quite limited. For example, it usually is not possible to carry out multimodal analyses with these models (e.g., analyses in which it is necessary to predict the use of several different modes such as bus transit, rail transit, carpool, and single-occupant automobile).

This course is concerned with a third class of mode choice models, called disaggregate models, that have substantial practical advantages over both elasticity methods and aggregate mode split models. Disaggregate models achieve a higher degree of policy sensitivity than either elasticity or aggregate mode split models. Disaggregate models can represent a wider range of policy variables than can either elasticity or aggregate models, and they can treat multimodal problems without difficulty. Moreover, disaggregate models avoid the biases inherent in aggregate models, and they are much more efficient than aggregate models in terms of data and computational requirements. Disaggregate models can be developed using data from only 1000-3000 households -- less than one tenth the number required by aggregate models -- and they can be implemented on microcomputers. In fact, as the examples given later in this course will show, many useful applications of disaggregate models can be made by hand with the aid of a desk calculator.

Disaggregate mode choice models have been available for use in transportation planning and policy analysis for nearly 15 years. Many

transportation agencies now use these models for practical policy analysis. This makes it important for transportation professionals to understand the principles underlying the development and use of disaggregate models, since failure to understand these principles can lead to the development of seriously erroneous models and to serious prediction errors.

Unfortunately, materials that explain how to use disaggregate models are not readily available. Most descriptions of disaggregate modeling techniques are written for members of the research community or for graduate students. People in both groups have extensive backgrounds in mathematics and statistics, and graduate students may be able to spend several months learning to use the techniques. Consequently, the available descriptions emphasize the mathematical and statistical details of the techniques and, thereby, convey the impression that the techniques are useful mainly to researchers and can be used only by people with considerable mathematical training. This is a false impression. The main concepts and methods of disaggregate mode choice modeling can be understood and applied by anybody who has mastered high-school algebra. The purpose of this course is to explain what disaggregate mode choice models are, how they work, and how they can be applied to practical problems, and to do this with a minimum of mathematics and jargon.

1.2 Description of the Course

This is a complete, self-instructing course in disaggregate mode choice modeling. It includes a text, worked examples, problems for readers to solve, and solutions to the problems. The course is designed for readers who are familiar with urban transportation planning issues and methods and have knowledge of mathematics at the level of high school algebra. No prior

familiarity with statistics or computer programming is needed. The only equipment required, apart from pencil and paper, is a desk calculator. A supplement to the course provides problems to be worked on a microcomputer. This supplement may be skipped by readers without access to an IBM-compatible microcomputer. The course is divided into self-contained modules of 1-2 hours duration. It is expected that most individuals will be able to complete the entire course in 15-20 hours of work.

The purpose of the course is to familiarize readers with the basic concepts and methods of disaggregate mode choice modeling and to do so with a minimum of mathematics and technical jargon. It is designed to help readers understand how disaggregate models work and why they are useful so that readers can become informed users of these models and their outputs. The course will not make experts out of its readers. No short, self-instructing, non-mathematical course could do this. However, this course will enable readers to understand what the experts are doing (or should be doing) and how the results can be used. It also will enable readers to do some of the things that they previously may have thought require the services of an expert. Readers who wish to achieve a more detailed understanding of disaggregate models or a higher level of expertise than this course provides should consider taking a college course in travel demand modeling or reading one or more of the references listed at the end of this module.

The course consists of 7 modules, including this introduction. Modules 2-4 describe the conceptual foundations of disaggregate mode choice modeling. These modules explain the assumptions about travel behavior that underlie disaggregate mode choice models and show how these assumptions are represented in models suitable for use in practical analysis. Numerical

examples are given that illustrate the usefulness of the behavioral assumptions and the plausibility of mode choice models based on these assumptions.

Modules 5-7 are concerned with the practical development and implementation of mode choice models. Module 5 discusses the explanatory variables that typically are used in disaggregate mode choice models, the choices that analysts face in selecting variables, and the practical consequences of alternative choices. Module 6 explains how disaggregate mode choice models are estimated or calibrated. This module also describes the data requirements of these models and discusses how the models can be tested empirically. Particular emphasis is placed on practical procedures for determining whether the correct explanatory variables have been used in a model and on comparing different versions of the same model to determine which provides the best explanation of the available data. Module 7 explains how aggregate travel demand can be predicted using disaggregate mode choice models.

The modules build on one another. Each uses material from its predecessors, and none can be understood without first understanding its predecessors. Therefore, readers are strongly advised to work through the modules in sequence without skipping any. Each module contains numerical examples that illustrate the material being presented, and each includes problems for the reader to solve. Readers are urged to work through the numerical examples and to understand them fully. Readers are also urged to solve the problems. It is possible to gain a complete understanding of the ideas presented here only by working with them. The problems provide an opportunity to do such work. Solutions to the problems are given following Module 7.

The time required to work through a module will vary greatly among both modules and readers. It is likely that most readers will be able to work through the text and examples of most modules in 1-2 hours, although some modules and readers may require more or less time. Working the problems at the end of a module may require an additional hour.

1.3 Summary

Disaggregate mode choice models have important practical advantages over other available methods for predicting the consequences of transportation policy measures that affect mode choice. This course is designed to provide readers with a working knowledge of practical disaggregate mode choice modeling. It does not require extensive mathematical training or other special technical preparation. It can be worked by readers who are familiar with urban transportation planning practice and are comfortable with high school algebra. The course is suitable for individuals who must carry out mode choice analyses, use and interpret the outputs of such analyses, or supervise those who do such work.

REFERENCES

Readers wishing additional information on disaggregate mode choice modeling, beyond that provided in this course, are encouraged to consult the following references.

M. Ben-Akiva and S.R. Lerman, Discrete Choice Analysis: Theory and Application to Travel Demand, The M.I.T. Press, Cambridge, MA, 1985.

Cambridge Systematics, Inc., Analytic Procedures for Estimating Changes in Travel Demand and Fuel Consumption, report DOE/PE/8628-1, Vol. 2, U.S. Department of Energy, October 1979.

Cambridge Systematics, Inc., Case City Applications of Analysis Methodologies, report DOE/PE/8628-1, Vol. 3, U.S. Department of Energy, October, 1979.

D.A. Hensher and L.W. Johnson, Applied Discrete-Choice Modelling, Croom-Helm, London, 1981.

T.A. Domencich and D. McFadden, Urban Travel Demand: A Behavioral Analysis, North Holland/American Elsevier, New York, 1975.

D.L. McFadden, "The Theory and Practice of Disaggregate Demand Forecasting for Various Modes of Urban Transportation," in Emerging Transportation Planning Methods, report DOT-RSPA-DPB-50-78-2, U.S. Department of Transportation, August 1978.

MODULE 2

INTRODUCTION TO CHOICE THEORY

2.1 Introduction

This module introduces the behavioral theory that forms the basis of disaggregate mode choice models. The theory is presented in its simplest form in this module. The theory is expanded and made more realistic in Modules 3 and 4.

2.2 The Role of Choice in Generating Travel Demand

The basic idea underlying modern approaches to travel demand modeling is that travel is the result of choices made by individuals or collective decision-making units such as households. For example, an individual preparing to travel to work must choose whether to drive alone, take a bus, travel in a carpool, etc. The individual also must choose when to leave home and, depending on the chosen mode, may have to choose which route to use. The objective of travel demand modeling is to model and predict the outcomes of these choices by individuals (or, if appropriate, by collective decision-making units such as households). Measures of aggregate travel, such as bus ridership, are obtained by adding up the choices of individuals.

To model the outcomes of individuals' choices, it is necessary to:

1. Identify the decisions that must be made and the options, or alternative outcomes, that are available to the individual. In this course, the decision that will be considered is choice of mode, and the options are travel modes such as drive alone, carpool, and bus. However, the methods that will be discussed also

are applicable to other travel choices, including choices of trip frequencies, destinations, and routes.

2. Identify variables likely to affect the choices of interest. It is particularly important to identify policy variables -- i.e., variables whose values may be changed through deliberate policy decisions -- since much practical travel demand modeling is concerned with predicting the consequences of changing the values of these variables. Travel time and travel cost are examples of policy variables relevant to mode choice.
3. Develop a mathematical formula that describes the dependence of choices on the relevant variables.

This module is concerned primarily with item 3. The module describes a theory of human preferences and choices that is useful for guiding the development of mathematical formulas relating choices to appropriate sets of variables. The application of the theory is illustrated with examples involving the prediction of mode choice.

To minimize the complexity of the presentation, it will be assumed in this module that all of the variables relevant to individuals' choices of modes are known to the analyst. This makes it possible to develop models that predict individuals' choices of modes with certainty and without error. Of course, it is not possible in practice to achieve such a high degree of modeling perfection, and it will be necessary to modify the models discussed in this module to make them suitable for use in real-world applications. The modified models, which are explained in Modules 3 and 4, are based on the behavioral concepts described in this module. The modifications needed to achieve practical models extend, rather than replace, the concepts presented in this module.

2.3 Preferences

An individual's choice represents an expression of his preferences among the available options at the time and under the conditions in which the choice is made. For example, if an individual decides to travel to work by bus rather than by driving alone or by carpooling, this means that he prefers bus to the other two modes under the conditions that exist when the choice is made.

It is important to understand that the preferences relevant to choice are the ones that pertain to the chooser's existing circumstances, not to an ideal set of circumstances. For example, a commuter boarding a bus may think to himself that he would really rather take a taxi if he could afford it and that he is taking the bus only because he does not have much money. Such thoughts do not imply that the commuter prefers taxi to bus under the existing circumstances. He would prefer taxi to bus under ideal circumstances (e.g., having a lot of money), but under the existing circumstances (e.g., having to give up lunch if he spends money on a taxi), he prefers bus.

Since choice is an expression of preferences, modeling and predicting choices is equivalent to modeling and predicting preferences -- i.e., if one has a model that enables one to predict an individual's preferences among the available options, one also is able to predict the same individual's choices. Preferences among a set of options depend on the attributes of the options and of the individual involved. For example, attributes of travel modes that are relevant to preferences among modes include travel time, travel cost, comfort, and reliability. Attributes of individuals that affect preferences among modes include income and the number of automobiles

owned. The next section describes a way of relating attributes to preferences and choices.

2.4 Utility Theory

Virtually all operational models for predicting individuals' choices are based on a behavioral principle called "utility maximization." This principle and its relation to choice can be stated in words very simply. According to the utility maximization principle, there is a mathematical function U , called a utility function, whose numerical value depends on attributes of the available options and the individual. The utility function has the property that its value for one option exceeds its value for another if and only if the individual prefers the first option to the second. Thus, the ranking of the available options according to the individual's preferences and the ranking according to the values of the utility function are the same. The individual chooses the most preferred option, which is the one with the highest utility-function value.

The utility maximization principle can be stated mathematically as follows. Let C denote the set of options available to an individual (e.g., drive alone, carpool, and bus in the case of mode choice). C is called the choice set. For each option i in C , let X_i denote the attributes of i for the individual in question. For example, if i corresponds to drive alone, X_i denotes the travel time, travel cost, and other relevant attributes of the drive alone mode for the individual in question. Let S denote the attributes of the individual that are relevant to preferences among the options in C (e.g., income, automobiles owned, etc.). Then, according to the utility maximization principle, there is a function U (the utility function) of the attributes of options and individuals that describes

individuals' preferences. U has the property that for any two options i and j in C

$$U(X_i, S) > U(X_j, S) \quad (2.1)$$

implies that the individual prefers alternative i to alternative j and will choose i if given a choice between i and j . Given a choice among many options, alternative i in C is chosen if

$$U(X_i, S) > U(X_j, S) \quad (2.2)$$

for all alternatives j (other than i) in C .

The utility function U is defined to have the following properties:

1. The function U is the same for all options. Differences among options are accounted for by differences in the numerical values of the attributes X , not by changing the function U . (Of course, the numerical value of U depends on the option, but the functional form of U is the same for all options.)
2. The utility of an alternative depends only on attributes of that alternative and of the individual. It does not depend on attributes of other alternatives. Thus, for example, the utility of driving alone does not depend on bus travel time and cost. Of course, the choice the individual makes depends on the attributes of all alternatives since the chosen mode is the one with the highest utility.

The following example illustrates the use of the utility maximization principle in mode choice analysis.

Example 2.1: A Utility Model of Mode Choice

Suppose that an individual can travel to work by driving alone, carpooling, or riding the bus. Assume that the relevant attributes of these

modes are travel time and travel cost. Assume that the relevant attribute of the individual is annual income. Let T denote door-to-door travel time in hours, C denote travel cost in dollars, and Y denote annual income in thousands of dollars per year. Let the utility function be

$$U(T,C,Y) = -T - 5C/Y. \tag{2.3}$$

Suppose the values of travel time and cost for the available modes are:

<u>Mode</u>	<u>Time (T)</u> <u>(hours)</u>	<u>Cost (C)</u> <u>(\$)</u>
Drive Alone	0.50	2.00
Carpool	0.75	1.00
Bus	1.00	0.75

Then if income is \$40,000 per year, for example, the value of U for drive alone is $-0.50 - 5(2.00)/40 = -0.75$. The following table shows the value of U corresponding to each mode for an individual whose income is \$40,000 per year ($Y = 40$) and an individual whose income is \$10,000 per year ($Y = 10$):

<u>Mode</u>	<u>U</u>	
	<u>Y = 40</u>	<u>Y = 10</u>
Drive Alone	-0.75	-1.50
Carpool	-0.88	-1.25
Bus	-1.09	-1.38

The high-income individual chooses to drive alone (because drive alone has the highest utility for this individual), and the low-income individual chooses to carpool. Note that all utilities are negative because U consists

of (generalized) costs of travel but excludes the value of reaching the destination. In this case, the highest value of U is the one that is least negative.

Now suppose that the quality of transit service is improved so that travel time for bus is 0.75 hr. Then the utilities become:

<u>Mode</u>	<u>U</u>	
	<u>Y = 40</u>	<u>Y = 10</u>
Drive Alone	-0.75	-1.50
Carpool	-0.88	-1.25
Bus	-0.84	-1.13

The high-income individual still drives alone, but the low-income individual switches to bus.

Although this example is very simple, it illustrates some important characteristics of choice models based on the utility maximization principle. First, it shows how a utility function can be used to describe the dependence of preferences and choices on attributes of individuals and options. Notice, in particular, that the same utility function describes the preferences of more than one individual. It is not necessary to have a separate utility function for each individual if differences among individuals can be accounted for by attribute variables such as income. Second, the example illustrates the use of utility theory to predict changes in preferences and choices that occur when an attribute of one of the options changes. Moreover, the utility model is able to capture differences in the responses of different individuals to the same attribute change.

Finally, the example illustrates some advantages of utility models over many traditional mode choice models. For instance, the model in the example treats choice among three modes and can easily be extended to treat more than three modes. Many traditional models are able to treat only two modes. In addition, since the utility model operates at the level of individuals, it guarantees that the percentages of individuals choosing a mode always are within the range 0-100% and always add up to 100%. Many traditional mode choice models do not have this obviously desirable property.

2.5 Properties of Utility Functions and Utility Models

It is tempting to interpret the numerical values of the utilities of a set of options as indicators of an individual's strengths of preference for the options. For example, if a certain individual's utility of driving alone is 5 and his utility of traveling by bus is 1, then it might be said that the individual's preference for driving alone is 5 times greater than his preference for traveling by bus. As it turns out, such an interpretation is both unnecessary and incorrect.

The interpretation is unnecessary because choice does not depend on strengths of preference; it depends only on preference ordering. A utility model always predicts that the option with the highest utility will be chosen, regardless of whether that option's utility is much larger or only slightly larger than the utilities of the other available options. For example, driving alone will be chosen if the utilities of driving alone, carpooling, and traveling by bus are 100.0, 2.0, and 1.0, respectively; and driving alone will also be chosen if the utilities are 2.1, 2.0, and 1.9.

That the preference strength interpretation of utility is incorrect follows from the observation that the utility function is defined only as a

function whose numerical values for the available options have the same ordering (e.g., highest to lowest) as the individual's preferences among the options. The definition of a utility function does not include any assumptions or statements about strengths of preference. Any function that reproduces preference orderings can serve as a utility function and will give the same predictions of choice, regardless of the signs or numerical values of the utilities. Thus, the utility function contains no information about strengths of preference.

In fact, there are infinitely many utility functions that can be used to describe the same preferences and that give the same predictions of choices. This nonuniqueness of utility functions is illustrated by the following example.

Example 2.2: Nonuniqueness of Utility Functions

In example 2.1, preferences among the alternatives drive alone, carpool, and bus were described with the utility function

$$U(T,C,Y) = -T - 5C/Y, \tag{2.4}$$

where T, C, and Y, respectively, are travel time in hours, travel cost in dollars, and income in thousands of dollars per year. However, exactly the same preference rankings and choice predictions would be obtained with any of the following alternative utility functions:

$$V(T,C,Y) = -TY - 5C, \tag{2.5}$$

$$W(T,C,Y) = 10 - 20T - 100C/Y, \tag{2.6}$$

$$X(T,C,Y) = -T^2 - 10CT/Y - 25C^2/Y^2. \tag{2.7}$$

To see this for the case of utility function V, suppose that $U(T_1, C_1, Y) > U(T_2, C_2, Y)$, where T_1 and C_1 denote the travel time and cost of option 1, and T_2 and C_2 denote the travel time and cost of option 2. Then

$$-T_1 - 5C_1/Y > -T_2 - 5C_2/Y. \quad (2.8)$$

Multiplying both sides of equation (2.8) by Y yields

$$-T_1Y - 5C_1 > -T_2Y - 5C_2, \quad (2.9)$$

which is equivalent to

$$V(T_1, C_1, Y) > V(T_2, C_2, Y). \quad (2.10)$$

Thus, the utility functions U and V are interchangeable: they give the same preference orderings and the same predictions of choice. Similarly, the utility functions W and X defined in equations (2.6) and (2.7) are interchangeable with U. You will be asked to show this in Exercise 2.3.

2.6 Predictions of Aggregate Travel Behavior

The utility maximization principle and utility-based choice models are methods for describing and predicting choices made by individuals. However, practical travel demand analysis rarely is concerned with the choices of individual travelers. Rather, it is concerned with the behavior of large groups or aggregates of travelers. A utility model can be used to obtain predictions of aggregate travel behavior: one simply adds up the model's predictions of the choices of the individuals in the group of interest. This process is illustrated in the following example.

Example 2.3: Aggregate Travel Behavior

As in Example 2.1, consider choice among the modes drive alone, carpool, and bus with the utility function

$$U(T, C, Y) = -T - 5C/Y, \quad (2.11)$$

where T, C, and Y, respectively, denote travel time in hours, travel cost in dollars, and income in thousands of dollars per year. Suppose that

individuals who live in a certain suburb of a city and work downtown face the following travel times and costs for trips from home to work:

<u>Mode</u>	<u>Time (T)</u> <u>(hours)</u>	<u>Cost (C)</u> <u>(\$)</u>
Drive Alone	0.50	2.00
Carpool	0.75	1.00
Bus	1.00	0.75

In addition, suppose that the incomes of these individuals are distributed as follows:

<u>Income</u>	<u>Percentage of</u> <u>Individuals</u>
17	5
19	15
27	25
33	25
37	20
40	10

Then the utility values and mode choices according to income group are:

<u>Income</u>	<u>Drive Alone</u>	<u>Carpool</u>	<u>Bus</u>	<u>Choice</u>
17	-1.09	-1.04	-1.22	Carpool
19	-1.03	-1.01	-1.20	Carpool
27	-0.87	-0.94	-1.14	Drive Alone
33	-0.80	-0.90	-1.11	Drive Alone
37	-0.77	-0.89	-1.10	Drive Alone
40	-0.75	-0.88	-1.09	Drive Alone

Since 20% of the individuals belong to income groups in which carpool is chosen and 80% belong to income groups in which drive alone is chosen, the aggregate mode shares are 20% for carpool and 80% for drive alone. No individuals in the population under consideration choose bus.

Notice that aggregate behavior cannot be predicted correctly by averaging the utility values over individuals and predicting aggregate behavior using the average utilities. The average utility of driving alone in this example is -0.86 [i.e., $0.05(-1.09) + 0.15(-1.03) + \dots + 0.10(-0.75)$], and the average utilities of carpooling and traveling by bus are -0.93 and -1.13 , respectively. Thus, use of the average utility values would result in the erroneous prediction that all of the travelers in the population under consideration drive alone.

2.7 Summary

Modern approaches to travel demand modeling are based on a behavioral principle called utility maximization. This module has explained the utility maximization principle and illustrated its use in predicting individuals' mode choices and aggregate mode shares.

EXERCISES

2.1 Refer to Example 2.1. Suppose the utility function in this example were

$$U(W,T,C,Y) = W - T - 5C/Y,$$

where W is the value of arriving at work. Let $W = 3.0$, and let the values of T and C be as in Example 2.1. Compute the utilities of drive alone, carpool, and bus for $Y = 40$ and $Y = 10$ using this new utility function. Are there now any negative utilities? Are any of the predicted choices different from those in Example 2.1?

2.2 Suppose, as in Example 2.1, that an individual can travel to work by driving alone, carpooling, or riding the bus. Assume that the relevant attributes of these modes are travel time and travel cost. Assume that the relevant attribute of the individual is annual income. Using the notation of Example 2.1, let T denote door-to-door travel time in hours, C denote travel cost in dollars, and Y denote annual income in thousands of dollars per year. Let the utility function be

$$U(T,C,Y) = -3T - 8C/Y.$$

Let the values of travel time and cost for the available modes be

<u>Mode</u>	<u>Time (T)</u> <u>(hours)</u>	<u>Cost (C)</u> <u>(\$)</u>
Drive Alone	0.35	2.25
Carpool	0.60	0.95
Bus	0.75	0.60

Compute the utility values for an individual with an income of \$40,000 per year and for an individual with an income of \$10,000 per year.

What mode would each individual choose?

2.3 Show that the utility functions W and X in Example 2.2 are interchangeable with U. Evaluate the utility functions U, V, W, and X for some representative values of the attributes of the drive alone, carpool, and bus modes. Use the results to illustrate why utility values should not be interpreted as strengths of preference.

2.4 Consider choice among the modes drive alone, carpool, and bus with the utility function:

$$U(T,C,Y) = -T - 5C/Y,$$

where T, C, and Y, respectively, denote travel time in hours, travel cost in dollars, and income in thousands of dollars per year. Suppose that individuals who live in a certain suburb of a city and work downtown face the following travel times and costs for trips from home to work:

<u>Mode</u>	<u>Time (T)</u> <u>(hours)</u>	<u>Cost (C)</u> <u>(\$)</u>
Drive Alone	0.50	2.50
Carpool	0.75	1.25
Bus	1.00	0.50

In addition, suppose that the incomes of these individuals are distributed as follows:

<u>Income</u>	<u>Percentage of Individuals</u>
14	5
18	15
22	25
26	25
30	20
34	10

Determine individuals' mode choices according to income, and use these to compute the aggregate shares of each mode.

Now suppose that bus travel time is reduced to 0.95 hr. Compute the new aggregate shares and the percentage changes in the shares resulting from the improvement in bus service. Also, compute the aggregate shares by first averaging the utilities over individuals and then predicting mode choices based on the average utilities. Do you obtain the same aggregate shares as when you determine the mode choices before averaging? If not, which method is correct?

MODULE 3

INTRODUCTION TO PROBABILISTIC CHOICE THEORY

3.1 Introduction

Module 2 introduced the theory of behavior that forms the basis of disaggregate mode choice models. This module continues developing the theory in ways that make it more realistic and useful for practical applications.

3.2 Inadequacy of Deterministic Utility Models

The theory of travel choice described in Module 2 yields a simple model of decision making that makes deterministic predictions of travel choices. In other words, according to this theory there is no uncertainty in the predicted choices. An individual is predicted to choose the alternative with the highest utility, and according to the model, there is no possibility that any other alternative will be chosen. Models based on utility maximization that yield deterministic predictions of choice are called deterministic utility models.

If deterministic utility models describe travel behavior correctly, then similar individuals would be expected to make the same travel choices when faced with the same sets of alternatives. In practice, however, it is not unusual for apparently similar individuals to make different choices when faced with similar or even identical alternatives. In fact, the same individual may make different choices when faced with the same alternatives on different occasions. For example, in studies of work trip mode choice it is frequently found that individuals who have identical personal characteristics according to the available data and who face similar sets of

travel alternatives choose different modes of travel to work. Some of these individuals may vary their choices from day to day for no apparent reason.

Deterministic utility models cannot treat such "unexplained" variations in travel behavior. Thus, deterministic utility models provide inadequate descriptions of travel behavior. The purpose of this module is to explore the reasons for this inadequacy and to lay the groundwork for a family of models, also based on utility theory, that do take account of unexplained variations in travel choices. This family of improved models is presented in detail in Module 4.

It is easy to understand the basic sources of the inadequacy of deterministic utility models. First, analysts and the individuals making the travel choices being modeled are unlikely to have the same information about the available alternatives. For example, the analyst may not have data on the reliability of a particular bus line or the likelihood that a particular individual will get a seat on the bus. But the individual in question is likely either to know these things (if he has had experience with the bus line) or to have opinions about them that are unknown to the analyst. Second, the analyst is unlikely to know all the characteristics of each individual that are relevant to mode choice. For example, an individual's choice of mode for the work trip may depend on whether other family members want to use the car that will be driven to work if automobile is the chosen mode. However, it is unlikely that an analyst will have detailed information on the activities of family members.

If analysts had data on all of the variables relevant to mode choice, it would be reasonable to expect that mode choice could be described and predicted satisfactorily by deterministic utility models. Experience has shown, however, that analysts do not have such data and have no realistic

possibility of obtaining them. Therefore, mode choice models should take a form that recognizes and accommodates analysts' lack of information. In this module and the next, it is shown how deterministic utility models can be modified to achieve this objective. The resulting models are called "random utility models" or "probabilistic choice models" because they describe preferences and choice in terms of probabilities. Instead of predicting that an individual will choose a particular mode with certainty, these models give probabilities that each of the available modes will be chosen. Thus, the analyst's lack of complete information about the attributes of alternatives and individuals is accommodated in the modeling process by predicting the probabilities with which choices will be made instead of predicting that a specific choice will be made with certainty.

The remainder of this module describes some specific limitations of information that affect mode choice modeling and explores their consequences. Examples are presented to show how these limitations lead to "unexplained" variations in choice and, therefore, the appearance of probabilistic choice behavior (i.e., choices that are made according to a deterministic utility model but that appear probabilistic to an analyst who has only partial knowledge of the relevant variables). These examples suggest the usefulness of probabilistic models to describe this behavior.

3.3 Limitations of Analysts' Information

Two types of limitations of analysts' information make deterministic utility models inadequate for practical mode choice analysis. First, travelers may not have exact knowledge of the attributes of the available alternatives. For example, a traveler choosing between bus and car for a trip to a downtown shopping location is unlikely to know the exact travel

time by car or bus, the exact waiting time for the bus, whether he will get a seat on the bus, or the exact likelihood of finding a free or low-cost parking space near his destination. Consequently, the traveler's opinions or perceptions of these attributes are likely to differ from the objectively measured values of the attributes. An analyst often has no way of obtaining exact information about an individual's opinions and perceptions. However, without this information, the analyst will not be able to predict precisely the individual's choices.

Second, the analyst may not know the true values of the travel attributes important to the individual, and he may not know the individual's utility function. These limitations of knowledge further restrict the ability of analysts to predict the travel behavior of individuals accurately. In the remainder of this section, five specific limitations of analysts' knowledge are discussed. Each of these limitations leads to the occurrence of unexplained variation in travel choices. The five limitations are:

1. Omission of relevant variables from models: A model of mode choice may omit one or more variables that are important to the traveler. This may happen either because such omission achieves a useful simplification of the model or because the analyst does not have data on the omitted variables. For example, the utility function used in Module 2 includes only travel time and cost as attributes of modes. If travelers consider other factors, such as comfort, reliability, and privacy, in addition to travel time and cost, their mode choices will vary, even if travel time and cost do not, according to their perceptions of these other factors.

2. Measurement error: Analysts' information about service quality may be subject to measurement errors. For example, data on travel time may be obtained from network models that yield estimates of travel times between zone centroids. These estimates may be erroneous due to network coding errors, errors in the assumed volume-delay functions, or because the trips in question originate or terminate at locations other than zone centroids. Thus, the analyst's estimates of travel time may be substantially different from the travel times actually experienced by travelers.
3. Proxy variables: It is often necessary in practical modeling to use variables that are different from the ones that are theoretically appropriate. For example, employment density may be used as a proxy for carpooling opportunities. If individuals' mode choices depend on carpooling opportunities, these choices will not be predicted precisely by a model that uses employment density in place of more precise indicators of carpooling opportunities. Another proxy variable commonly used in mode choice models is income as a proxy for automobile ownership.
4. Differences between individuals may be ignored: Different individuals may evaluate alternatives differently. For example, differences in costs among alternatives may be less important to wealthy people than to poor people. Some models attempt to capture this difference by, for example, dividing the cost of travel by the individual's income or wage rate, as was done in Module 2. However, other differences are more difficult to represent in a model. For example, physical characteristics of the individual that may affect the importance of seating availability, walk

distance, or wait time may not be known to the analyst and, therefore, not included in a model.

5. Day to day variations in the choice context may be ignored: The data customarily used in mode choice modeling do not include information on day-to-day variations in the choice context that may affect mode choice. Examples of such variations include short-term unavailability of a car due to repair needs or variations in the needs of another family member, the need to carry heavy packages on a particular day, or variations in the activities planned to be undertaken after work.

Each of the foregoing limitations of analysts' knowledge can cause variations in mode choices, either among individuals or by the same individual on different occasions, that cannot be explained by the observed (or measured) attributes of travelers or modes. The next section presents several examples that illustrate how such unexplained variation in choices arise and show how it creates the appearance of probabilistic travel behavior.

3.4 Examples of Unexplained Variation in Choice Behavior

The examples in this section illustrate how the limitations of analysts' knowledge described in Section 3.3 lead to unexplained variation in choices and the appearance of probabilistic behavior.

Example 3.1: Missing Variables

Suppose, as in Example 2.1, that an individual can travel to work by driving alone, carpooling, or riding the bus. Suppose that the relevant attributes of these modes are travel time (T) and cost (C). However,

extending Example 2.1, let the relevant characteristics of the individual include both annual income (Y) and the number of automobiles owned by his household (A). The effect of automobile ownership is to increase or decrease the utility of the drive alone and carpool modes, depending on the number of automobiles owned. This effect is represented by the following equations for the utility function:

$$U_{DA} = -T_{DA} - 5C_{DA}/Y + 0.4(A - 1) \quad (3.1a)$$

$$U_{CP} = -T_{CP} - 5C_{CP}/Y + 0.2(A - 1) \quad (3.1b)$$

$$U_B = -T_B - 5C_B/Y. \quad (3.1c)$$

As in Example 2.1, suppose the values of travel time and cost are

<u>Mode</u>	<u>Time (T)</u> <u>(Hours)</u>	<u>Cost (C)</u> <u>(\$)</u>
Drive Alone	0.50	2.00
Carpool	0.75	1.00
Bus	1.00	0.75

If income is \$15,000 per year (Y = 15), then the utilities of the three alternatives for households with no automobiles (A = 0) are:

$$U_{DA} = -0.5 - 5(2.0/15) + 0.4(0 - 1) = -1.57$$

$$U_{CP} = -0.75 - 5(1.0/15) + 0.2(0 - 1) = -1.28$$

$$U_B = -1.0 - 5(0.75/15) = -1.25.$$

Using the same equations, the utility values of the three alternatives for three different levels of automobile ownership are:

<u>Mode</u>	<u>Zero Cars</u>	<u>One Car</u>	<u>Two Cars</u>
Drive Alone	-1.57	-1.17	-0.77
Carpool	-1.28	-1.08	-0.88
<u>Bus</u>	<u>-1.25</u>	<u>-1.25</u>	<u>-1.25</u>
Mode Chosen	Bus	Carpool	Drive Alone

According to the principle of utility maximization, individuals in households without cars use bus, those in households with one car carpool, and those in households with two cars drive alone.

Now consider what happens if the automobile ownership variable is not included in the data set and the analyst predicts choice with the utility function of Example 2.1. All individuals would be predicted to have utilities equal to -1.17 for drive alone, -1.08 for carpool, and -1.25 for bus, so all individuals would be predicted to travel by carpool. However, if there were 20 zero-car households, 50 one-car households, and 30 two-car households, it would be observed that, in fact, 20% of the individuals choose bus, 50% choose carpool, and 30% drive alone. Thus, the omission of the automobile ownership variable from the utility function causes variations in travel choices that are not explained by the model. These variations in choices among alternatives give the appearance of probabilistic choice behavior because they can be described by a probability distribution in which the probabilities that bus, carpool, and drive alone are chosen are 0.20, 0.50, and 0.30, respectively.

Example 3.2: Measurement Error

Now consider the zero-car households in Example 3.1, but assume that different individuals have different travel times for the automobile modes. Specifically, assume that the true drive alone and carpool travel times for individuals are distributed according to the following relative frequencies:

<u>Percentage of Individuals</u>	<u>20%</u>	<u>50%</u>	<u>20%</u>	<u>10%</u>
Drive Alone Travel Time (hr.)	0.40	0.50	0.60	0.70
Carpool Travel Time (hr.)	0.65	0.75	0.85	0.95

In this table, the travel times in the second column are the same as those in Example 3.1. The travel times in the first column are lower than those in Example 3.1, and the travel times in the third and fourth columns are higher than in Example 3.1. Thus, the travel times in Example 3.1 are measured with error. Fifty percent of the individuals have travel times that are given correctly by Example 3.1, but 20% have travel times that are lower than those of Example 3.1 and 30% have travel times that are higher. Assume that the travel costs are the same as in Example 3.1.

The utilities of the three modes can be obtained from the utility function of equations (3.1). For zero-car households with incomes of \$15,000 per year ($Y = 15$), the utilities are:

Percentage of Individuals	Utilities Based on the Travel Time Distributions in the Previous Table				Based on Ex. 3.1
	20%	50%	20%	10%	100%
Drive Alone	-1.47	-1.57	-1.67	-1.77	-1.57
Carpool	-1.18	-1.28	-1.38	-1.48	-1.28
Bus	-1.25	-1.25	-1.25	-1.25	-1.25
Chosen Mode	Carpool	Bus	Bus	Bus	Bus

In this table, the first column gives the utilities for travelers whose drive alone and carpool travel times are less than those in Example 3.1, the second column gives the utilities for travelers with travel times equal to those in Example 3.1, and the third and fourth columns give the utilities for travelers with drive alone and carpool travel times greater than those in Example 3.1. It can be seen that 20% of the travelers (those with "lower" drive alone and carpool travel times) choose carpool, and the remaining 80% choose bus. However, if the distributions of travel time were ignored as in Example 3.1, 100% of these travelers would be predicted to choose bus. The choices of 20% of the travelers would be predicted erroneously because erroneous travel time data had been used.

If the same travel time distributions were applied to one-car households with incomes of \$15,000 per year, the utilities and mode choices would be:

Percentage of Individuals	Utilities Based on the Travel Time Distributions				Based on Ex. 3.1
	20%	50%	20%	10%	100%
Drive Alone	-1.07	-1.17	-1.27	-1.37	-1.17
Carpool	-0.98	-1.08	-1.18	-1.28	-1.08
Bus	-1.25	-1.25	-1.25	-1.25	-1.25
Chosen Mode	Carpool	Carpool	Carpool	Bus	Carpool

In this case, 90% of the travelers would choose carpool and 10% would choose bus. However, if the distributions of travel time were ignored, all of these travelers would be predicted to choose carpool. In this case, the use of erroneous travel time data would cause erroneous predictions of the choices of 10% of the travelers.

In summary, ignoring the distributions of travel times of zero- and one-car households results in predictions that do not reflect the true variations in mode choices. In other words, the actual choices vary in ways not explained by the model used to make the predictions. As in Example 3.1, the variations can be described by a probability distribution.

In Exercise 3.2, you will be asked to determine the percentages of two-car households choosing each mode when travel time is distributed as in this example.

Example 3.3: Differences in Preferences among Individuals

Examples 3.1 and 3.2 assume that the same utility function applies to every individual. However, it is possible that, for reasons not known to the analyst, different individuals have different preferences among the same sets of alternatives. When this happens, the preferences of different individuals are described by different utility functions. For example,

consider a population consisting of two groups in which time is valued differently. Suppose the preferences of one group are described by the utility functions:

$$U_{DA}^{(1)} = -0.75T_{DA} - 5C_{DA}/Y + 0.4(A - 1) \quad (3.2a)$$

$$U_{CP}^{(1)} = -0.75T_{CP} - 5C_{CP}/Y + 0.2(A - 1) \quad (3.2b)$$

$$U_B^{(1)} = -0.75T_B - 5C_B/Y. \quad (3.2c)$$

Suppose the preferences of the second group are described by the utility functions:

$$U_{DA}^{(2)} = -1.5T_{DA} - 5C_{DA}/Y + 0.4(A - 1) \quad (3.3a)$$

$$U_{CP}^{(2)} = -1.5T_{CP} - 5C_{CP}/Y + 0.2(A - 1) \quad (3.3b)$$

$$U_B^{(2)} = -1.5T_B - 5C_B/Y. \quad (3.3c)$$

Then the members of group two consider time to be twice as valuable as do the members of group 1. As a result, the utilities of the available modes will be different for members of the two groups and the choices of mode will be different.

To illustrate this, suppose that the travel times and costs of Example 3.1 are correct, and consider individuals whose incomes are \$15,000 per year. The utility values for individuals in group 1 and owning zero cars are:

$$U_{DA}^{(1)} = -0.75(0.5) - 5(2.0/15) + 0.4(0 - 1) = -1.44$$

$$U_{CP}^{(1)} = -0.75(0.75) - 5(1.0/15) + 0.2(0 - 1) = -1.10$$

$$U_B^{(1)} = -0.75(1.0) - 5(0.75/15) = -1.00.$$

The utilities for members of both groups and owning zero, one, or two cars are:

<u>Mode</u>	Zero Cars		One Car		Two Cars	
	Group 1	Group 2	Group 1	Group 2	Group 1	Group 2
Drive Alone	-1.44	-1.82	-1.04	-1.42	-0.64	-1.02
Carpool	-1.10	-1.66	-0.90	-1.46	-0.70	-1.26
<u>Bus</u>	<u>-1.00</u>	<u>-1.75</u>	<u>-1.00</u>	<u>-1.75</u>	<u>-1.00</u>	<u>-1.75</u>
Chosen Mode	Bus	Car-pool	Car-pool	Drive Alone	Drive Alone	Drive Alone

Notice that for zero- and one-car households, the mode chosen by members of group 1 is different from the mode chosen by members of group 2. Thus, the proportions of individuals owning zero, one, and two cars choosing each mode depend on the proportions that belong to each preference group. However, if the analyst does not know that individuals belong to different preference groups, it will appear that "identical" individuals (i.e., individuals who have the same incomes and levels of automobile ownership) make different choices when faced with identical alternatives. This variation in choices will be unexplained by the information available to the analyst and will give the appearance of probabilistic choice.

Example 3.4: Multiple Sources of Unexplained Variations in Choices

The sources of unexplained variations in mode choices that have been illustrated in the Examples 3.1-3.3 can occur simultaneously in practice. That is, a choice model may be used that omits the automobile ownership variable even though automobile ownership affects mode choice, does not account for differences in preferences among individuals, and is based on data that include errors in measured travel time. As an example of this, suppose that the utility function used by the analyst is:

$$U(T,C,Y) = -T - 5C/Y . \quad (3.4)$$

Suppose, also, that the measured values of T and C are as in Example 3.1. Then for an individual whose income is \$15,000 per year, the analyst would estimate the utilities of drive alone, carpool, and bus to be:

<u>Mode</u>	<u>Utility</u>
Drive Alone	-1.17
Carpool	-1.08
Bus	-1.25

The analyst would predict that all of these individuals will choose carpool. However, the actual choices of a specific individual will depend on which preference group he is in, the number of automobiles owned by his household, and his true travel times. Therefore, the modes chosen will vary among individuals in ways that are not explained by the analyst's specification of the utility function (Equation 3.4). Choices will appear to be probabilistic to the analyst.

The foregoing examples illustrate some of the ways in which unexplained variation in travel behavior can arise. As has been discussed, this variation causes travel behavior to appear to be probabilistic. The next two sections discuss in more detail how unexplained variation in travel behavior can be described in probabilistic terms. The use of a probabilistic representation of travel behavior enables models to reflect both the effects of variables that are included in the analyst's specification of the utility function and the effects of errors in the analyst's specification.

3.5 The Basic Formulation of Probabilistic Models

In each of the examples of the preceding section, the correct utility function differs from that used by the analyst due to the omission of a variable that influences mode choice (Example 3.1), measurement error (Example 3.2), variations in preferences among individuals (Example 3.3), or all of these (Example 3.4). In each case, the correct utility function, U , can be written as the sum of the utility function specified by the analyst, V , and an error term, e . That is:

$$U = V + e. \tag{3.5}$$

The specified utility functions (V) and error terms (e) for Example 3.1 are:

Components of the Utility Function -- Example 3.1

<u>Mode</u>	<u>Specified Utility</u>	<u>Error Term</u>
Drive Alone	$-T_{DA} - 5C_{DA}/Y$	$0.4(A - 1)$
Carpool	$-T_{CP} - 5C_{CP}/Y$	$0.2(A - 1)$
Bus	$-T_B - 5C_B/Y$	0

The utilities of drive alone and carpool include an error term due to the omission of automobile ownership from the specified utilities.

The specified utility functions and error terms for Example 3.2 are given in the following table. In this table T^* denotes measured travel time, and T denotes true travel time.

Components of the Utility Function -- Example 3.2

<u>Mode</u>	<u>Specified Utility</u>	<u>Error Term</u>
Drive Alone	$-T_{DA}^* - 5C_{DA}/Y$	$T_{DA} - T_{DA}^*$
Carpool	$-T_{CP}^* - 5C_{CP}/Y$	$T_{CP} - T_{CP}^*$
Bus	$-T_B - 5C_B/Y$	0

In this case, the utilities of drive alone and carpool include an error term because the drive alone and carpool travel times of some individuals are measured with error. There is no error term in the utility of bus because bus travel time is measured without error.

The components of the utility function for Example 3.3 are:

Components of the Utility Function -- Example 3.3

<u>Mode</u>	<u>Specified Utility</u>	<u>Error Term</u>
Drive Alone	$-T_{DA} - 5C_{DA}/Y + 0.4(A - 1)$	$0.25T_{DA}$ (group 1)
		$-0.50T_{DA}$ (group 2)
Carpool	$-T_{CP} - 5C_{CP}/Y + 0.2(A - 1)$	$0.25T_{CP}$ (group 1)
		$-0.50T_{CP}$ (group 2)
Bus	$-T_B - 5C_B/Y$	$0.25T_B$ (group 1)
		$-0.50T_B$ (group 2)

In this case, errors are present in all the utilities because the specified utilities ignore the differences between the two population groups.

Of course, these examples are artificial. The true utility functions are known, and there is no need to use a representation such as Equation (3.5) that replaces part of the known utility function with an error term. In practice, however, an analyst never knows the true utility function. An analyst can hope to know the utility function only up to an error term. In effect, the analyst always measures or estimates utility with error, and an error term of unknown size is always present in the analyst's specification of the utility function. This error term accounts for variables that the analyst knows influence travel behavior but that are not included in his data set or that he chooses to omit from his model (e.g., because he cannot

forecast them well). It also accounts for any variables that influence travel behavior but are completely unknown to the analyst.

It is customary to think of the error term as a random variable whose values are described by a probability distribution. The true utility function, U , is then also a random variable consisting of the sum of a deterministic component, V , and a random component, e . The deterministic component of the utility function is what the analyst can measure or estimate, and the random component is the difference between the true utility function and the deterministic component.

When the true utilities of the alternatives are random variables, it is not possible to state with certainty which alternative has the greatest utility or which alternative is chosen. This is because utility and choice depend on the random components of the utilities of the available alternatives, and these components cannot be measured. The most an analyst can do is predict the probability that an alternative has the maximum utility and, therefore, the probability that the alternative is chosen. Accordingly, the analyst must represent travel behavior as being probabilistic. The need for a probabilistic representation is a consequence of the analyst's inability to make the utility measurements required to predict behavior with certainty.

Of course, the probability that a particular alternative is chosen depends on the values of the deterministic components of the utilities, and the values of these components can be measured (or estimated). In general, the probability that a particular alternative is chosen either increases or stays the same when the deterministic component of its utility increases and either decreases or stays the same when the deterministic component decreases. This relation is illustrated by the following example.

Example 3.5: Dependence of Choice on the Deterministic Component of Utility

Let the true utilities be given by equations (3.2) and (3.3), and let the deterministic (or specified) component of the utility function be given by equation (3.4). Thus, the effects of differences in automobile ownership and differences in preferences among population groups are not represented in the deterministic component of the utility function and are accounted for by the error term (i.e., the difference between true utility and the deterministic component of utility). Let the population be distributed over automobile ownership classes and preference groups according to the following percentages:

	Automobiles Owned		
	<u>0</u>	<u>1</u>	<u>2</u>
Preference Group 1	15%	25%	5%
Preference Group 2	5	30	20

Let income be \$15,000 per year ($Y = 15$), and let the travel times and costs according to mode be as in Example 3.1. Then, the values of the deterministic component of the utility function are -1.17 for drive alone, -1.08 for carpool, and -1.25 for bus. The values of true utility and the chosen modes according to automobile ownership and preference group are the same as in Example 3.3 (see the table on p. 34). The percentages of the population choosing each mode can be computed from the following table:

<u>Autos Owned</u>	<u>Preference Group</u>	<u>Percent of Population</u>	<u>Mode Chosen</u>
0	1	15	Bus
	2	5	Carpool
1	1	25	Carpool
	2	30	Drive Alone
2	1	5	Drive Alone
	2	20	Drive Alone

The percentages of the entire population choosing each mode are 55% for drive alone, 30% for carpool, and 15% for bus. However, the deterministic component of the utility function for each mode has the same values for all preference groups and automobile ownership levels. Thus, the deterministic component of utility cannot account for the variations in choice within the population. Since an analyst knows only the deterministic components of utility, the population will appear to be choosing among the modes with probabilities 0.55, 0.30, and 0.15 for drive alone, carpool, and bus, respectively. This appearance of probabilistic choice is due to the fact that the analyst does not know the variations in preferences and automobile ownership levels that cause different individuals to choose different modes.

Now, suppose that in an effort to shift mode choice to high-occupancy modes, a parking tax of \$0.50 is imposed on each single-occupant vehicle. Then, the cost of driving alone increases to \$2.50 while the costs of carpooling and traveling by bus remain unchanged at \$1.00 and \$0.75, respectively. The deterministic component of drive alone utility decreases to -1.34 from its former value of -1.17. The values of the deterministic

components of the utilities of carpool and bus remain unchanged. The new values of the true utilities are:

<u>Mode</u>	Zero Cars		One Car		Two Cars	
	<u>Group 1</u>	<u>Group 2</u>	<u>Group 1</u>	<u>Group 2</u>	<u>Group 1</u>	<u>Group 2</u>
Drive Alone	-1.61	-1.99	-1.21	-1.59	-0.81	-1.19
Carpool	-1.10	-1.66	-0.90	-1.46	-0.70	-1.26
<u>Bus</u>	<u>-1.00</u>	<u>-1.75</u>	<u>-1.00</u>	<u>-1.75</u>	<u>-1.00</u>	<u>-1.75</u>
Chosen Mode	Bus	Car-pool	Car-pool	Car-pool	Car-pool	Drive Alone

The percentages of the population choosing each mode are now 20% for drive alone, 65% for carpool, and 15% for bus. Again, since the analyst knows only the values of the deterministic component of utility, the population will appear to be choosing among the modes with probabilities of 0.20, 0.65, and 0.15 for drive alone, carpool, and bus, respectively.

The next table shows the value of the deterministic component of utility for drive alone and the percentages of the population choosing each mode for three different values of the parking tax on single-occupant vehicles:

<u>Tax</u>	Deterministic Component of Utility for <u>Drive Alone</u>	Percentage of Population Choosing		
		<u>Drive Alone</u>	<u>Carpool</u>	<u>Bus</u>
\$0.00	-1.17	55	30	15
0.25	-1.25	15	65	20
0.50	-1.34	15	65	20

Notice that as the parking tax increases, the deterministic component of the utility of drive alone decreases and the percentage of the population choosing to drive alone either decreases or stays unchanged. Since the analyst knows only the deterministic components of utility, mode choice will appear to be probabilistic. The probability that a given mode is chosen will be the percentage of the population choosing that mode divided by 100. Therefore, such an analyst will observe that as the deterministic component of the utility of drive alone decreases, the probability that drive alone is chosen either decreases or stays the same.

3.6 The Probability of Observing a Specific Choice

An important interpretation of probabilistic travel behavior can be obtained by considering a common procedure for collecting mode choice data. Consider a group of apparently similar travelers who make different choices when faced with the same alternatives. If observations of mode choice are obtained by sampling this group randomly, the probabilities of selecting travelers who choose drive alone, carpool, and bus are the same as the probabilities that individual travelers make these choices. In other words, the sampling probabilities and the individual choice probabilities are the same.

To illustrate this, consider Example 3.1, in which the utility function specified by the analyst includes travel time, travel cost, and income but not automobile ownership. Assume that all travelers have incomes of \$15,000 per year and that the travel times and costs of drive alone, carpool, and bus for all travelers are as in Example 3.1. If 20% of the travelers' households own 0 cars, 50% own 1 car, and 30% own 2 cars, the probability that a randomly sampled traveler chooses bus is 0.20, the probability that

he chooses carpool is 0.50, and the probability that he chooses drive alone is 0.30. These sampling probabilities are the same as the choice probabilities obtained in Example 3.1 by considering a group of 20 zero-car, 50 one-car, and 30 two-car travelers. Thus, one can think of the probability of a given choice as being the probability of sampling an individual who makes that choice. In fact, as will be discussed in Module 6, this interpretation forms the basis of methods for calibrating or estimating probabilistic models of choice.

3.7 Aggregate Prediction with Probabilistic Choice

In Module 2, where the choices of individuals were predicted deterministically, predictions of aggregate travel behavior were obtained by summing the choices of the individuals comprising the aggregate group. Aggregate predictions when choices by individuals are predicted probabilistically are obtained in an analogous manner: the probabilities of choices by the individuals in the aggregate group are summed.

Consider Example 3.1, where choice is influenced by automobile ownership but the analyst's specification of the utility function includes only travel time, travel cost, and income. Assume that the distribution of automobile ownership in different income classes is as follows:

<u>Income</u> <u>(\$000)</u>	<u>% Zero-</u> <u>Car Households</u>	<u>% One-</u> <u>Car Households</u>	<u>% Two-</u> <u>Car Households</u>
17.5	40	50	10
22.5	20	60	20
27.5	15	60	25
32.5	10	60	30
37.5	5	55	40
42.5	0	50	50

The choice of each traveler can be obtained by substituting the values of the travel time, travel cost, income, and automobile ownership variables into Equation (3.1). The chosen modes for each income and automobile ownership class are:

<u>Income</u> <u>(\$000)</u>	<u>Automobile Ownership</u>		
	<u>Zero Cars</u>	<u>One Car</u>	<u>Two Cars</u>
17.5	Bus	Carpool	Drive Alone
22.5	Carpool	Drive Alone	Drive Alone
27.5	Carpool	Drive Alone	Drive Alone
32.5	Carpool	Drive Alone	Drive Alone
37.5	Carpool	Drive Alone	Drive Alone
42.5	Carpool	Drive Alone	Drive Alone

Since the percentages of travelers in each income class owning 0, 1, and 2 cars are known, the percentages of travelers choosing each mode according to income class can be computed. The results are:

<u>Income</u> <u>(\$000)</u>	<u>Percent Choosing</u>		
	<u>Drive</u> <u>Alone</u>	<u>Carpool</u>	<u>Bus</u>
17.5	10	50	40
22.5	80	20	0
27.5	85	15	0
32.5	90	10	0
37.5	95	5	0
42.5	100	0	0

These percentages constitute the analyst's probabilistic predictions of mode choice for each income class. (Recall that the analyst's utility function does not include automobile ownership, so the analyst cannot predict choice deterministically.) Using these predictions of individual choice, predictions of aggregate choice are obtained in the following way. For each income class, the probability that an individual chooses a given mode is multiplied by the number of individuals in the class, thereby obtaining the number of individuals in the class who are predicted to choose the given mode. These predictions are then summed over all income classes to obtain the total number of individuals who are predicted to choose the given mode.

As an example of this, suppose the numbers of individuals in the income classes are known, so the analyst has the following information available:

<u>Income</u> <u>(\$000)</u>	<u>Number of</u> <u>Individuals</u>	<u>Percent Choosing</u>		
		<u>Drive</u> <u>Alone</u>	<u>Carpool</u>	<u>Bus</u>
17.5	20	10	50	40
22.5	60	80	20	0
27.5	100	85	15	0
32.5	100	90	10	0
37.5	80	95	5	0
42.5	40	100	0	0

Then, the number of individuals predicted to choose drive alone is $0.10(20) + 0.80(60) + 0.85(100) + 0.90(100) + 0.95(80) + 1.0(40) = 341$. Similarly, the numbers of individuals predicted to choose carpool and bus are 51 and 8, respectively.

3.8 Summary

The utility maximization principle provides a valuable framework for the analysis of travel choice behavior. However, deterministic utility models are inadequate due to the inability of the analyst to know the exact utility function of an individual and to measure accurately all of the variables relevant to travel choice. This module has explained how limitations of the analyst's knowledge create unexplained variations in travel choices and the appearance of probabilistic behavior. In addition, the module has provided a basis for modeling travel choice probabilistically and has showed how aggregate travel behavior can be predicted when choices by individuals are predicted probabilistically. The next module describes a specific probabilistic choice model that can be used to predict individual choice probabilities.

EXERCISES

- 3.1 In Example 3.1, show how the utility values for each mode are obtained for individuals in households with one or two cars. What are the corresponding set of utilities for individuals in households with three cars?
- 3.2 Using Example 3.2, determine the mode choice percentages for two-car households?
- 3.3 Combine Examples 3.1 and 3.2. Assume that the proportions of zero-, one-, and two-car households are 0.20, 0.50, and 0.30, respectively. Also assume that the distribution of travel times for each automobile ownership class is as shown in Example 3.2. Compute the mode choice for each automobile ownership-travel time group. Also compute the overall probability that each mode is chosen.
- 3.4 Suppose that the distribution of individuals according to income class is:

<u>Income</u> <u>(\$000)</u>	<u>Number of</u> <u>Individuals</u>
17.5	10
22.5	20
27.5	20
32.5	30
37.5	20
42.5	3

Repeat the aggregate prediction of Section 3.7 using the above income distribution, thereby obtaining the total numbers of individuals predicted to choose each mode.

MODULE 4

THE LOGIT CHOICE MODEL

4.1 Introduction

The preceding module showed that limitations of analysts' knowledge of the variables that influence individuals' mode choices make it necessary to predict these choices in terms of probabilities. This module introduces the most widely used mathematical model for making probabilistic predictions of mode choices.

Before the model is introduced, it is worthwhile to identify some of the properties that a probabilistic choice model should have. These properties include:

1. The probability of choosing a particular alternative should depend on the deterministic components of the utilities of all available alternatives. The chosen alternative is the one with the highest total utility. Therefore, it depends on the relative values of the total utilities of all alternatives. It depends on the deterministic components of the utilities of all alternatives because the total utilities depend on the deterministic components of utility.
2. The probability that an alternative is chosen should increase when the deterministic component of its utility increases. It should decrease when the deterministic component of the utility of any other alternative increases. This property follows from the fact that increasing the deterministic component of an alternative's utility increases the probability that that alternative has the

highest utility. Increasing the deterministic component of another alternative's utility decreases the probability that the first alternative has the highest utility.

3. The model should accommodate choice sets containing any number of alternatives so that it can be applied regardless of the number of alternatives involved and can be used to predict the effects of changing the number of alternatives (e.g., as would happen when transit is initiated in an area that previously had none).
4. The model should be easy to understand and to use in practice.

4.2 The Binomial Logit Model

Before considering how choices among arbitrary numbers of alternatives can be modeled, it is useful to consider a model of choice among only two alternatives. The most frequently used model of probabilistic choice among two alternatives is the binomial logit model. In this model, the probability that alternative 1 is chosen when the choice set consists of alternatives 1 and 2 is given by the following formula:

$$\text{Pr}(1) = \frac{\exp(V_1)}{\exp(V_1) + \exp(V_2)}, \quad (4.1)$$

where $\text{Pr}(1)$ is the probability that an individual chooses alternative 1, $\exp()$ is the exponential function, and V_1 and V_2 are the deterministic components of the utilities of alternatives 1 and 2, respectively. The exponential function transforms its argument (the expression in parentheses) as shown in Figure 4.1 and as tabulated in Table 4.1. It can be seen from the figure and the table that the exponential function is monotonic (i.e., its value increases when the value of its argument increases) and that its value is always positive.

Figure 4.1: The Exponential Function

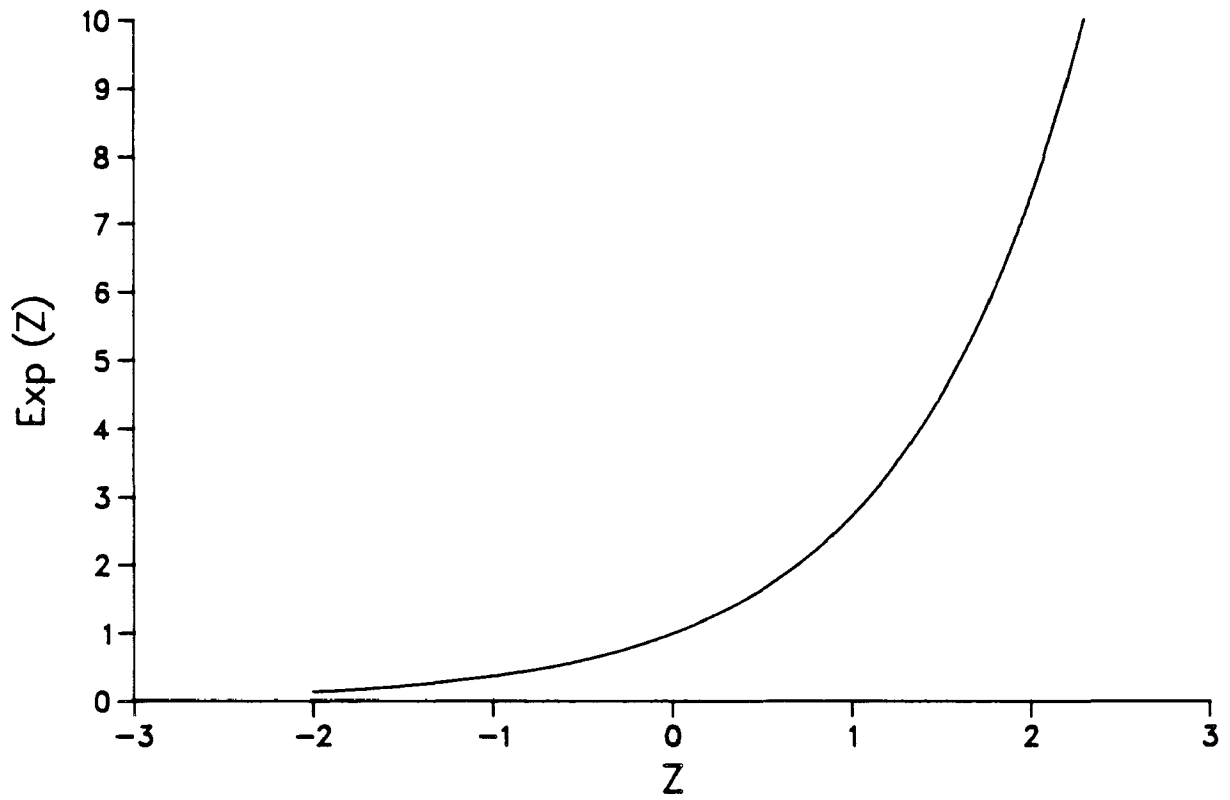


TABLE 4.1 -- THE EXPONENTIAL FUNCTION

<u>Z</u>	<u>EXP(Z)</u>
-3.0	0.050
-2.5	0.082
-2.0	0.135
-1.5	0.223
-1.0	0.368
-0.5	0.607
0.0	1.000
0.5	1.649
1.0	2.718
1.5	4.482
2.0	7.389
2.5	12.182
3.0	20.086

As was discussed in Section 3.6 of Module 3, $Pr(1)$ is the probability that a randomly selected individual with deterministic utility components V_1 and V_2 chooses alternative 1. Equation (4.1) implies that in the binomial logit model, this probability increases monotonically with the deterministic component of the utility of alternative 1 and decreases monotonically with the deterministic component of the utility of alternative 2.

Since there are only two choices available to the individual, the probability that alternative 2 is chosen is one minus the probability that alternative 1 is chosen. Thus,

$$\begin{aligned}
\text{Pr}(2) &= 1 - \frac{\exp(V_1)}{\exp(V_1) + \exp(V_2)} \\
&= \frac{\exp(V_1) + \exp(V_2)}{\exp(V_1) + \exp(V_2)} - \frac{\exp(V_1)}{\exp(V_1) + \exp(V_2)} \\
\text{Pr}(2) &= \frac{\exp(V_2)}{\exp(V_1) + \exp(V_2)} . \tag{4.2}
\end{aligned}$$

This probability increases monotonically with the deterministic component of the utility of alternative 2 and decreases monotonically with the deterministic component of the utility of alternative 1.

The binomial logit model has three of the four desirable properties of probabilistic choice models that were listed in Section 4.1. The binomial logit choice probabilities depend on the deterministic components of the utilities of all alternatives (property 1); the probability of choosing a particular alternative increases when the deterministic component of the utility of that alternative increases, and the probability decreases when the deterministic component of the utility of the other alternative increases (property 2); and the model is easy to understand and apply (property 4). The binomial logit model cannot treat choice among more than two alternatives, so it does not have property 3.

In the binomial logit model, the probabilities of choosing alternatives 1 and 2 are equal when the deterministic components of the two alternatives' utilities are equal. Moreover, the choice probabilities are most sensitive to changes in the deterministic components of the utilities when these components are approximately equal and the choice probabilities are close to 0.5. To see this, divide the numerator and denominator of Equation (4.1) by $\exp(V_1)$ to obtain

$$\text{Pr}(1) = \frac{1}{1 + \exp[-(V_1 - V_2)]} . \tag{4.3}$$

Equation (4.3) shows that $Pr(1)$ depends only on the difference between V_1 and V_2 . Figure 4.2 shows a graph of $Pr(1)$ as a function of this difference, and Table 4.2 tabulates $Pr(1)$ for selected values of V_1 and V_2 .

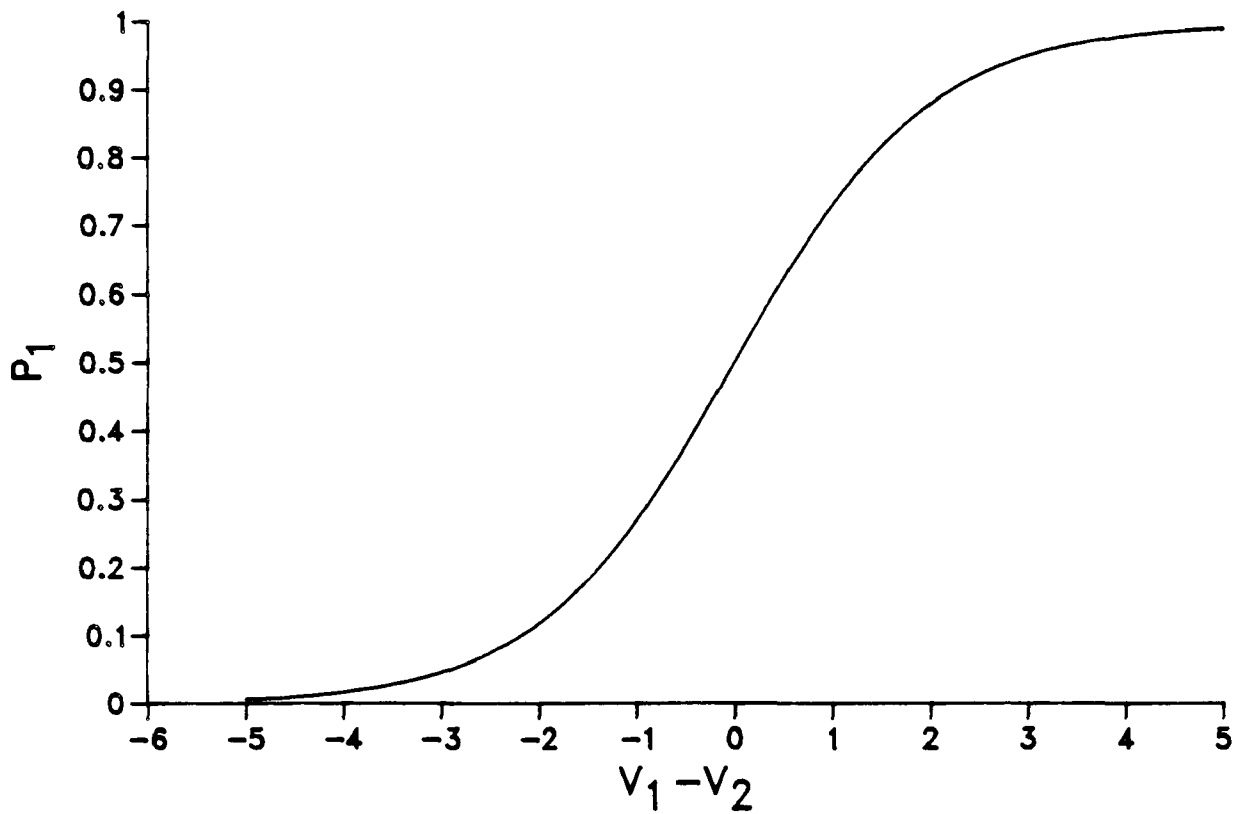
TABLE 4.2 -- VALUES OF $Pr(1)$ IN THE BINOMIAL LOGIT MODEL

<u>Case</u>	<u>V_1</u>	<u>V_2</u>	<u>$V_1 - V_2$</u>	<u>$Pr(1)$</u>
1	0.0	0.0	0.0	0.50
2	0.5	0.0	0.5	0.62
3	2.0	0.0	2.0	0.88
4	2.5	0.0	2.5	0.92
5	2.0	-0.5	2.5	0.92

It can be seen from the table and figure that:

1. The probabilities of choosing alternatives 1 and 2 are both equal to 0.5 when the deterministic components of the alternatives' utilities are equal (i.e., when $V_1 - V_2 = 0$). This is because $\exp(0) = 1$ (see Table 4.1).
2. The probability of choosing alternative 1 is more sensitive to changes in the deterministic component of the utility of either alternative when $Pr(1)$ is close to 0.5 (i.e., when the deterministic components of utility are approximately equal) than when $Pr(1)$ is close to 0 or 1. (The same statement applies to $Pr(2)$.) For example, in Table 4.2, $Pr(1)$ is equal to 0.5 in case 1 but closer to 1.0 than to 0.5 in case 3. V_1 increases by 0.5 from case 1 to case 2 and from case 3 to case 4. However, $Pr(1)$ increases by 0.12 from case 1 to case 2 but by only 0.04 from case

Figure 4.2: The Binomial Logit Model



3 to case 4. This property is illustrated in Figure 4.2, where the plot of $\text{Pr}(1)$ is steeper when $\text{Pr}(1)$ is close to 0.5 than when it is close to 0 or 1.

Another property of the binomial logit model is that $\text{Pr}(1)$ is affected equally by increases in the value of V_1 and decreases in the value of V_2 . This property is illustrated in Table 4.2. The change in $\text{Pr}(1)$ is the same in going from case 3 to case 4 (an increase of 0.5 in the value of V_1) as in going from case 3 to case 5 (a decrease of 0.5 in the value of V_2). This result follows from the fact that change in $V_1 - V_2$ is the same in going from case 3 to case 4 as it is in going from case 3 to case 5.

4.3 The Multinomial Logit Model

The binomial logit model can easily be extended to accommodate choices among more than two alternatives. To see how this is done, suppose, first, that there are three alternatives in the choice set and that the deterministic components of their utilities are V_1 , V_2 , and V_3 . In the extended model, the probability that alternative 1 is chosen is

$$\text{Pr}(1) = \frac{\exp(V_1)}{\exp(V_1) + \exp(V_2) + \exp(V_3)}. \quad (4.4)$$

The additional alternative is incorporated by adding an additional exponential term to the denominator of the equation for $\text{Pr}(1)$. The probability of choosing alternative 1 remains proportional to the exponential function of the deterministic component of its utility. In general, $\text{Pr}(1)$ is smaller when there are three alternatives in the choice set than when there are only two (assuming that the values of V_1 and V_2 are the same in both cases). This is because the total probability of choosing any of the alternatives, which always equals one, must be shared among more

alternatives when there are three alternatives in the choice set than when there are only two.

A model that accommodates any specified number of alternatives can be obtained by adding the appropriate exponential terms to the denominator of equation (4.4). Specifically, suppose there are J alternatives in the choice set, where J is any number greater than or equal to 2. Let the deterministic components of the utilities of the alternatives be V_1, V_2, \dots, V_J . Then the probability that alternative 1 is chosen is

$$\Pr(1) = \frac{\exp(V_1)}{\sum_{j=1}^J \exp(V_j)}. \quad (4.5)$$

The probability of choosing alternative 2 is

$$\Pr(2) = \frac{\exp(V_2)}{\sum_{j=1}^J \exp(V_j)}. \quad (4.6)$$

In general, the probability of choosing alternative i ($i = 1, \dots, J$) is

$$\Pr(i) = \frac{\exp(V_i)}{\sum_{j=1}^J \exp(V_j)}. \quad (4.7)$$

Equation (4.7) is called the multinomial logit model. In the multinomial logit model, as in the binomial logit model, the probability of choosing an alternative increases monotonically with the deterministic component of that alternative's utility and decreases monotonically with the deterministic component of the utility of any other alternative.

The multinomial logit model has all of the desirable properties of the binomial logit model and, in addition, can be applied to any number of alternatives. Thus, it has all of the desirable properties of a probabilistic choice model listed in Section 4.1.

An important additional property of both the binomial and multinomial logit models is that the choice probabilities depend only on the differences between the deterministic components of the alternatives' utilities. This property is illustrated in Equation (4.3) and Table (4.2) for the binomial logit model, where the probability of choosing either alternative depends only on $V_1 - V_2$. The corresponding dependence for the multinomial logit model is illustrated by dividing the numerator and denominator of Equation (4.4) by $\exp(V_1)$ to obtain

$$\text{Pr}(1) = \frac{1}{1 + \exp[-(V_1 - V_2)] + \exp[-(V_1 - V_3)]} . \quad (4.8)$$

In this case, the probability that alternative 1 is chosen depends only on the values of $V_1 - V_2$ and $V_1 - V_3$. If there were J alternatives in the choice set, where J exceeds 3, then the probability that alternative 1 is chosen would depend only on $V_1 - V_2, V_1 - V_3, \dots, V_1 - V_J$. In the multinomial logit model, choice probabilities never depend on ratios of utilities such as $V_1/V_2, V_1/V_3$, etc.

4.4 Application of the Multinomial Logit Model to Mode Choice Analysis

The following example illustrates the application of the multinomial logit model to mode choice analysis.

Example 4.1: Application of the Multinomial Logit Model

Consider travel to work, and let there be three modes in the choice set: drive alone, carpool, and bus. Let the deterministic components of the utilities of these modes be:

<u>Mode</u>	<u>V</u>
Drive Alone	2.5
Carpool	2.0
Bus	1.0

Then, the values of the terms $\exp(V_j)$ and of their sum are

<u>Mode</u>	<u>exp(V)</u>
Drive Alone	12.18
Carpool	7.39
<u>Bus</u>	<u>2.72</u>
Sum	22.29

Substitution of these values into Equation (4.7) yields

$$\begin{aligned} \text{Pr}(\text{Drive Alone}) &= 12.18/22.29 = 0.55 \\ \text{Pr}(\text{Carpool}) &= 7.39/22.29 = 0.33 \\ \text{Pr}(\text{bus}) &= 2.72/22.29 = 0.12 \end{aligned}$$

As expected, the mode with the highest deterministic component of utility (drive alone) has the highest probability of being chosen. Notice, also, that the sum of the probabilities over all available modes equals 1. This always happens in the multinomial logit model because one of the alternatives must be chosen.

To verify that you understand the computation of multinomial logit choice probabilities, compute the probabilities of drive alone, carpool, and bus when the deterministic components of the utilities are 2.5 for drive alone, 1.5 for carpool, and 1.0 for bus. You can obtain the values of the

exponential function from Table 4.1. The correct probabilities are 0.63 for drive alone, 0.23 for carpool, and 0.14 for bus.

4.5 Incorporation of Attributes of Alternatives and Individuals

Example 4.1 assumed a fixed value of the deterministic component of each mode's utility. In practice, the deterministic component of a mode's utility depends on attributes of that mode (but not of other modes) and of the individual making the choice. The following example illustrates how choice probabilities can be made to depend on attributes of alternatives and individuals.

Example 4.2: Choice Probabilities That Depend on Attributes

Suppose that the deterministic component of the utility of mode j (j = drive alone, carpool, or bus) is

$$V_j = -T_j - 5C_j/Y, \quad (4.9)$$

where T_j and C_j , respectively, are the travel time (in hours) and cost (in dollars) of mode j , and Y is the annual income (in thousands of dollars) of the traveler. Suppose the travel time and cost values are:

<u>Mode</u>	<u>Time</u>	<u>Cost</u>
Drive Alone	0.50	2.00
Carpool	0.75	1.00
Bus	1.00	0.75

Then the deterministic components of the modes' utilities and their exponentials for individuals with incomes of \$15,000 per year ($Y = 15$) and \$30,000 per year ($Y = 30$) are:

<u>Mode</u>	Y = 15		Y = 30	
	<u>V</u>	<u>exp(V)</u>	<u>V</u>	<u>exp(V)</u>
Drive Alone	-1.17	0.31	-0.83	0.44
Carpool	-1.08	0.34	-0.92	0.40
<u>Bus</u>	<u>-1.25</u>	<u>0.29</u>	<u>-1.13</u>	<u>0.32</u>
Sum		0.94		1.16

The corresponding choice probabilities are:

<u>Mode</u>	Y = 15 <u>Pr(Mode)</u>	Y = 30 <u>Pr(mode)</u>
Drive Alone	0.33	0.38
Carpool	0.36	0.34
<u>Bus</u>	<u>0.31</u>	<u>0.28</u>
Sum	1.00	1.00

Note that the probabilities of choosing the relatively inexpensive modes (carpool and bus) are higher for the low-income individuals than for the high-income ones.

Now suppose that it is desired to predict the effects of increasing the bus fare by \$0.25. This fare change is represented in the multinomial logit model by increasing the value of C_{bus} by \$0.25. The resulting choice probabilities are

<u>Mode</u>	<u>Y = 15</u> <u>Pr(Mode)</u>	<u>Y = 30</u> <u>Pr(mode)</u>
Drive Alone	0.34	0.38
Carpool	0.37	0.35
<u>Bus</u>	<u>0.29</u>	<u>0.27</u>
Sum	1.00	1.00

These probabilities reflect a shift away from the bus because of its increased cost and resulting lower utility. You should compute the probabilities yourself to make sure you understand how they were obtained.

4.6 Alternative-Specific Constants

In the logit model used in Example 4.2, two modes have equal probabilities of being chosen if they have equal travel times and travel costs. (As a check of your understanding of the logit model, explain why this is so.) In practice, however, other factors, such as comfort, reliability, and safety, may cause one mode to have a greater probability of being chosen than another, even if the two modes have equal travel times and costs. The best way to account for the effects of such other factors is to include variables representing them in the deterministic component of the utility function. However, this often is not possible in practice, since many of these factors are difficult to measure and predict. An alternative method that always can be implemented easily consists of adding appropriate constant terms to the deterministic components of the utility functions of all the modes except one. These constants are called alternative-specific constants. The mode whose deterministic utility component does not include such a constant is called the base mode. The alternative-specific constant

for a given mode is the average amount that factors not included in the deterministic component of the utility function contribute to the difference between the utilities of the given mode and the base mode. In other words, it is the average contribution of the error terms to the differences between the two modes' utilities. It does not matter which mode is selected as the base mode; the values of the choice probabilities will be the same for any base mode if the values of the alternative-specific constants are assigned correctly. (Alternative-specific constants are sometimes called bias constants since they seem to represent biases by travelers toward or against other modes compared to the base mode. However, this term is misleading. The constants do not represent biases. They represent the average effects of variables not present in the model.)

The following example illustrates the use of alternative-specific constants.

Example 4.3: Alternative-Specific Constants

Suppose that the deterministic components of the utility functions of drive alone, carpool, and bus are

$$V_{DA} = 0.8 - T_{DA} - 5C_{DA}/Y \quad (4.10a)$$

$$V_{CP} = 0.2 - T_{CP} - 5C_{CP}/Y \quad (4.10b)$$

$$V_B = - T_B - 5C_B/Y. \quad (4.10c)$$

In this case, bus is the base mode, and the alternative-specific constants for drive alone and carpool are 0.8 and 0.2, respectively. The signs and magnitudes of these constants indicate that on the average, factors other than travel time and cost that affect mode choice tend to favor drive alone over both carpool and bus and carpool over bus.

To illustrate the effects of alternative-specific constants on logit choice probabilities, suppose that the travel time and cost values are the same as in Example 4.2 and that $Y = 30$. Then the values of the deterministic components of utility with and without the alternative-specific constants are:

<u>Mode</u>	Without Constants		With Constants	
	<u>V</u>	<u>exp(V)</u>	<u>V</u>	<u>exp(V)</u>
Drive Alone	-0.83	0.44	-0.03	0.97
Carpool	-0.92	0.40	-0.72	0.49
<u>Bus</u>	<u>-1.13</u>	<u>0.32</u>	<u>-1.13</u>	<u>0.32</u>
Sum		1.16		1.78

The resulting logit choice probabilities with and without the alternative-specific constants are:

<u>Mode</u>	Without	With
	Constants	Constants
	<u>Pr(Mode)</u>	<u>Pr(mode)</u>
Drive Alone	0.38	0.54
Carpool	0.34	0.28
<u>Bus</u>	<u>0.28</u>	<u>0.18</u>
Sum	1.00	1.00

The choice probabilities with the alternative-specific constants are very different from and more realistic than those obtained without the constants.

Any mode can be selected as the base mode when alternative-specific constants are introduced into a model. It does not matter which mode is the

base. The choice probabilities will be the same, regardless of base, if the differences between the values of the alternative-specific constants for any two alternatives are the same for all choices of base. For example, suppose that drive alone, rather than bus, had been selected as the base mode in Example 4.3. Suppose, in addition, that the deterministic components of the utility functions had been

$$V_{DA} = -T_{DA} - 5C_{DA}/Y \quad (4.11a)$$

$$V_{CP} = -0.6 - T_{CP} - 5C_{CP}/Y \quad (4.11b)$$

$$V_B = -0.8 - T_B - 5C_B/Y. \quad (4.11c)$$

Then, just as in equations (4.10), the difference between the alternative-specific constants for drive alone and carpool is 0.6 (that is, $0.0 - (-0.6)$), the difference between the constants for drive alone and bus is 0.8 (that is, $0.0 - (-0.8)$), and the difference between the constants for carpool and bus is 0.2 (that is, $-0.6 - (-0.8)$). You should verify that logit models based on Equations (4.10) and (4.11) yield the same choice probabilities by evaluating these probabilities for the values of the travel time, cost, and income variables used in Example 4.2.

4.7 Independence from Irrelevant Alternatives

One of the most important properties of the multinomial logit model is independence from irrelevant alternatives (IIA). The IIA property states that for any individual, the ratio of the probabilities of choosing two alternatives is independent of the availability or attributes of any other alternatives. For example, in a multinomial logit model of choice between drive-alone, carpool, and bus, the probabilities of choosing drive alone and carpool are

$$\Pr(\text{DA}) = \frac{\exp(V_{\text{DA}})}{\exp(V_{\text{DA}}) + \exp(V_{\text{CP}}) + \exp(V_{\text{B}})}, \quad (4.12a)$$

and

$$\Pr(\text{CP}) = \frac{\exp(V_{\text{CP}})}{\exp(V_{\text{DA}}) + \exp(V_{\text{CP}}) + \exp(V_{\text{B}})}. \quad (4.12b)$$

The ratio of these probabilities is

$$\frac{\Pr(\text{DA})}{\Pr(\text{CP})} = \frac{\exp(V_{\text{DA}})}{\exp(V_{\text{CP}})} = \exp(V_{\text{DA}} - V_{\text{CP}}). \quad (4.13)$$

This ratio is independent of the attributes and availability of bus. The ratio is the same regardless of whether bus is an available alternative.

In the general multinomial logit model, the probability of choosing alternative i when there are J alternatives in the choice set is given by Equation (4.7). Equation (4.7) implies that for any two alternatives i and k ,

$$\frac{\Pr(i)}{\Pr(k)} = \frac{\exp(V_i)}{\exp(V_k)} = \exp(V_i - V_k). \quad (4.14)$$

This equation shows that the ratio $\Pr(i)/\Pr(k)$ depends only on V_i and V_k . The ratio is the same regardless of which other alternatives, if any, are in the choice set and regardless of the attributes of any other alternatives.

The IIA property limits the responses to transportation changes that can be predicted by the multinomial logit model. For example, if the available modes are drive alone, carpool, and bus, a multinomial logit model predicts that the proportion of non-bus travelers choosing carpool (the ratio $\Pr(\text{CP})/[\Pr(\text{DA}) + \Pr(\text{CP})]$) is independent of the quality of bus service. Therefore, an improvement in bus service would be predicted by a multinomial logit model to draw travelers from drive alone and carpool in proportion to the original shares of these modes. The improvement in bus service would not be predicted to draw travelers mainly from carpools, say, unless carpooling were the dominant non-bus mode. This is an important

consequence of the IIA property that will be discussed further in Subsection 4.7a.

There are two other important practical consequences of the IIA property, in addition to the limitations it places on the predictions that can be made by multinomial logit models. These are:

1. It greatly simplifies the process of predicting the consequences of adding a mode to the choice set.
2. It provides great flexibility in the forms of data that can be used to calibrate models.

The first of these consequences is discussed in Subsection 4.7b. Discussion of the second consequence is an advanced topic that will not be treated in this course.

4.7a Limits on the Applicability of Multinomial Logit Due to IIA

The IIA property limits the effectiveness of the multinomial logit model in predicting choices and changes in choices in certain circumstances. An extreme example of this problem is called the red bus/blue bus paradox.

Example 4.4: The Red Bus/Blue Bus Paradox

Suppose the modes available for travel between home and work are drive alone and a bus that is painted red (red bus or RB). Assume that the attributes of drive alone and red bus are such that $V_{DA} = V_{RB}$. Then the binomial logit formula (Equations (4.1) and (4.2)) implies that $\Pr(DA) = \Pr(RB) = 0.5$. Now suppose a competing bus operator starts operating a bus painted blue (blue bus or BB) on the same route as the red bus. The blue bus uses exactly the same kind of vehicle, runs on exactly the same schedule, and serves exactly the same stops as the red bus. The only

difference between the red and blue buses is their color. If color does not affect choice of mode, then initiation of blue bus service should cause existing bus riders to divide evenly between the red and blue buses. The addition of blue bus to travelers' choice sets should have no effect on travelers who choose to drive alone because it does not affect the relative service quality of drive alone and bus. (The assumption that the red and blue buses have identical schedules implies that effective service frequency is unchanged by the initiation of blue bus service.) Therefore, the choice probabilities following the initiation of blue bus service should be $\Pr(\text{DA}) = 0.5$, $\Pr(\text{RB}) = 0.25$, and $\Pr(\text{BB}) = 0.25$.

Now consider the prediction made by the logit model. Since the red and blue buses are identical in all attributes relevant to mode choice, $V_{\text{RB}} = V_{\text{BB}}$. In addition, $V_{\text{DA}} = V_{\text{RB}}$ by assumption. Therefore, the deterministic components of the utilities of the three modes drive alone, red bus, and blue bus are equal. Let V denote this common value. Then for any of the three modes

$$\Pr(\text{mode}) = \frac{\exp(V)}{\exp(V) + \exp(V) + \exp(V)} = 1/3. \quad (4.15)$$

According to this equation, introduction of the blue bus causes the share of drive alone to decrease from 1/2 to 1/3 of the travelers. That is, 1/3 of the original drive alone travelers are predicted to switch to bus. (To obtain this result, suppose that there are 30 travelers in all. Then before the initiation blue bus service, the number predicted to drive alone is $0.5(30) = 15$. If choices after the initiation of blue bus service are given by Equation (4.15), then the number of travelers choosing to drive alone after blue bus service starts is predicted to be $30(1/3) = 10$. Thus 1/3 of the drivers alone are predicted to switch to bus.) This result is both

inconsistent with the expectations developed in the previous paragraph and unreasonable.

The red bus/blue bus paradox provides an important illustration of the possible consequences IIA, but it is extreme. A more realistic example of the effects of IIA is the following:

Example 4.5: Effects of the IIA Property

Consider an individual who has a choice between drive alone, carpool, bus, and light rail. Let the deterministic component of the logit utility function be

$$V_{DA} = 0.8 - T_{DA} - 0.25C_{DA} \quad (4.16a)$$

$$V_{CP} = 0.2 - T_{CP} - 0.25C_{CP} \quad (4.16b)$$

$$V_B = -0.2 - T_B - 0.25C_B \quad (4.16c)$$

$$V_{LR} = \quad - T_{LR} - 0.25C_{LR} \quad (4.16d)$$

where LR denotes light rail, and T and C are the travel time in hours and travel cost in dollars. Let the values of T and C be

<u>Mode</u>	<u>Time</u>	<u>Cost</u>
Drive Alone	0.50	2.00
Carpool	0.75	1.00
Bus	1.20	0.50
Light Rail	1.00	0.75

Then the values of V and the choice probabilities are:

<u>Mode</u>	<u>V</u>	<u>exp(V)</u>	<u>Pr(Mode)</u>
Drive Alone	-0.20	0.819	0.458
Carpool	-0.80	0.449	0.251
Bus	-1.53	0.217	0.121
<u>Light Rail</u>	<u>-1.19</u>	<u>0.304</u>	<u>0.170</u>
Sum		1.789	1.000

Now suppose that the cost of traveling by light rail increases by \$0.50. If bus and light rail operate in the same corridors, we would expect that most individuals diverted away from light rail would choose to travel by bus. However, according to the logit model, the new choice probabilities are:

<u>Mode</u>	<u>V</u>	<u>exp(V)</u>	<u>Pr(Mode)</u>
Drive Alone	-0.20	0.819	0.467
Carpool	-0.80	0.449	0.256
Bus	-1.53	0.217	0.123
<u>Light Rail</u>	<u>-1.31</u>	<u>0.270</u>	<u>0.154</u>
Sum		1.755	1.000

The logit model's prediction of change in the probability of choosing each mode is shown in the following table:

<u>Mode</u>	Pr(Mode)		<u>Change</u>
	<u>Before Cost Increase</u>	<u>After Cost Increase</u>	
Drive Alone	0.458	0.467	+0.009
Carpool	0.251	0.256	+0.005
Bus	0.121	0.123	+0.002
<u>Light Rail</u>	<u>0.170</u>	<u>0.154</u>	<u>-0.016</u>
Sum	1.000	1.000	0.000

Notice that the probability of choosing each mode other than light rail is predicted to increase in proportion to its original share. This is a consequence of the IIA property, which requires the ratios $\text{Pr}(\text{drive alone})/\text{Pr}(\text{carpool})$ and $\text{Pr}(\text{drive alone})/\text{Pr}(\text{bus})$ to stay constant when the cost of light rail travel increases (see equation (4.14)). In aggregate terms, the riders who stop using light rail when its cost increases are predicted to distribute themselves among the remaining modes in proportion to the initial probabilities of choosing the remaining modes. Therefore, most of the riders who leave light rail are predicted to drive alone since drive alone has the highest initial choice probability (0.458). However, such a result, though possible (e.g., if bus and light rail operate in different corridors so that bus is not a feasible alternative for light rail travelers), is not necessarily realistic. For example, it is not consistent with our expectations if bus is an alternative to light rail. This inconsistency between the predictions of the logit model and reasonable expectations limits the usefulness of the multinomial logit model in situations such as this one.

In most existing models, IIA problems involving trade-offs between competing transit modes are avoided through the simplifying assumption that transit travelers choose the transit mode that provides the fastest travel. Thus, the choice between bus and light rail is most commonly treated during the building of paths through the transit network, rather than in predicting mode shares. The mode choice model includes only a single, generic transit mode that involves bus travel for certain trips and light rail travel for others.

Combining transit modes in this way, however, can lead to serious prediction errors. Moreover, the potential problems posed by IIA are not restricted to choices among transit modes. They can also arise in choices among automobile-based modes, such as drive alone and carpool. Therefore, it is worthwhile to consider how IIA problems can be avoided without combining modes.

Frequently, it is possible to avoid unrealistic consequences of IIA by including additional variables in the deterministic component of the utility function. As an illustration of this, suppose that in Example 4.5 the light rail travelers are mainly individuals who do not have cars and, therefore, are highly unlikely to choose to drive alone. Then individuals who choose not to use light rail after the cost increases will switch mainly to carpool and bus. If carpooling is difficult for individuals who do not have cars available, then individuals who stop using light rail will switch mainly to bus. These effects can be accommodated within a multinomial logit model by including the variable automobile ownership in the deterministic component of the utility function. The following example illustrates this.

Example 4.6: Avoiding the Unrealistic Consequences of IIA

Suppose, as in Example 4.5, that travelers choose between the modes drive alone, carpool, bus, and light rail. However, let the deterministic components of the utilities of these modes be

$$V_{DA} = -2.84 - T_{DA} - 0.25C_{DA} + 4.5A \quad (4.17a)$$

$$V_{CP} = -2.17 - T_{CP} - 0.25C_{CP} + 3.5A \quad (4.17b)$$

$$V_B = -0.20 - T_B - 0.25C_B \quad (4.17c)$$

$$V_{LR} = - T_{LR} - 0.25C_{LR} \quad (4.17d)$$

where A is the number of automobiles owned by the traveler's household. As in Example 4.5, let the values of travel time (T) and travel cost (C) be:

<u>Mode</u>	<u>Time (Hrs.)</u>	<u>Cost (\$)</u>
Drive Alone	0.50	2.00
Carpool	0.75	1.00
Bus	1.20	0.50
Light Rail	1.00	0.75

Then the values of V and exp(V) for travelers whose households own 0, 1, and 2 cars are:

<u>Mode</u>	<u>0 Cars</u>		<u>1 Car</u>		<u>2 Cars</u>	
	<u>V</u>	<u>exp(V)</u>	<u>V</u>	<u>exp(V)</u>	<u>V</u>	<u>exp(V)</u>
Drive Alone	-3.83	0.022	0.664	1.94	5.16	174.
Carpool	-3.17	0.042	0.334	1.40	3.83	46.2
Bus	-1.53	0.217	-1.53	0.217	-1.53	0.217
<u>Light Rail</u>	<u>-1.19</u>	<u>0.305</u>	<u>-1.19</u>	<u>0.304</u>	<u>-1.19</u>	<u>0.304</u>
Sum		0.586		3.86		221.

The multinomial logit choice probabilities according to automobile ownership level are:

<u>Mode</u>	Pr(Mode)		
	<u>0 Cars</u>	<u>1 Car</u>	<u>2 Cars</u>
Drive Alone	0.0368	0.503	0.789
Carpool	0.0720	0.362	0.209
Bus	0.370	0.0562	0.0001
<u>Light Rail</u>	<u>0.521</u>	<u>0.0790</u>	<u>0.0014</u>
Sum	1.000	1.000	1.000

Notice that bus and light rail are used mainly by 0-car owners and that drive alone and carpool are used mainly by 1- and 2-car owners. Suppose that 25% of the travelers under consideration own 0 cars, 50% own 1 car, and 25% own 2 cars. Then the aggregate share of each mode in the population as a whole can be obtained by substituting the choice probabilities according to automobile ownership level into the formula

$$\begin{aligned} \text{Share}(\text{Mode}) = & 0.25\text{Pr}(\text{Mode for } A = 0) + 0.50\text{Pr}(\text{Mode for } A = 1) \\ & + 0.25\text{Pr}(\text{Mode for } A = 2). \end{aligned} \quad (4.18)$$

The results of this substitution are shown in the following table:

<u>Mode</u>	<u>Aggregate Share</u>
Drive Alone	0.458
Carpool	0.251
Bus	0.121
Light Rail	0.170

Notice that these aggregate shares are exactly the same as the choice probabilities in Example 4.5.

Now assume that the cost of light rail transit increases by \$0.50. The following table shows the resulting values of V and $\exp(V)$ according to automobile ownership for each mode:

<u>Mode</u>	0 Cars		1 Car		2 Cars	
	<u>V</u>	<u>exp(V)</u>	<u>V</u>	<u>exp(V)</u>	<u>V</u>	<u>exp(V)</u>
Drive Alone	-3.83	0.022	0.664	1.94	5.16	174.
Carpool	-3.17	0.042	0.334	1.40	3.83	46.2
Bus	-1.53	0.217	-1.53	0.217	-1.53	0.217
<u>Light Rail</u>	<u>-1.31</u>	<u>0.269</u>	<u>-1.31</u>	<u>0.269</u>	<u>-1.31</u>	<u>0.269</u>
Sum		0.550		3.83		221.

The new choice probabilities according to automobile ownership level are:

<u>Mode</u>	Pr(Mode)		
	<u>0 Cars</u>	<u>1 Car</u>	<u>2 Cars</u>
Drive Alone	0.0392	0.508	0.789
Carpool	0.0767	0.365	0.209
Bus	0.395	0.0567	0.0001
<u>Light Rail</u>	<u>0.489</u>	<u>0.0704</u>	<u>0.0012</u>
Sum	1.000	1.000	1.000

The changes in the choice probabilities according to automobile ownership level can be obtained by subtracting Pr(Mode) before the increase in light rail cost from Pr(Mode) after the increase. The results are:

<u>Mode</u>	Change in Pr(Mode)		
	<u>0 Cars</u>	<u>1 Car</u>	<u>2 Cars</u>
Drive Alone	+0.0024	+0.005	0.000
Carpool	+0.0047	+0.003	0.000
Bus	+0.025	+0.0005	0.000
Light Rail	-0.032	-0.0086	-0.0002

Notice that, as expected, the travelers who have changed mode are mainly those who own 0 cars and that the mode change by these travelers consists mainly of switching from light rail to bus.

The aggregate shares following the increase in the cost of light rail travel can be obtained from Equation (4.17). These shares and the changes in shares caused by the cost increase are:

<u>Mode</u>	Share		<u>Change</u>
	<u>Before Cost Increase</u>	<u>After Cost Increase</u>	
Drive Alone	0.458	0.461	+0.003
Carpool	0.251	0.254	+0.003
Bus	0.121	0.127	+0.006
<u>Light Rail</u>	<u>0.170</u>	<u>0.158</u>	<u>-0.012</u>
Sum	1.000	1.000	0.000

Notice that in contrast to the situation in Example 4.5, the bus share now increases by twice as much as either the drive alone or the carpool share. This result is consistent with expectations when light rail and bus serve the same corridors and the light rail travelers consist mainly of individuals without cars. (Recall that expectations under these conditions were developed in Example 4.5.) Thus, the change in the specification of

the deterministic component of the utility function has remedied the unreasonable consequences of the IIA property that were found in Example 4.5.

Example 4.6 has shown how the unreasonable consequences of the IIA property can be alleviated by including an additional variable in the deterministic component of the utility function. Another way to alleviate these consequences is to base predictions on a model, other than the multinomial logit model, that does not have the IIA property. Discussion of such models is beyond the scope of this course. Readers interested in learning about them should consult the book by Ben-Akiva and Lerman listed in the references at the end of Module 1.

4.7b Introduction of New Modes

An important problem in transportation analysis is the prediction of ridership on new travel modes. One of the advantages of the IIA property is that it greatly simplifies the process of predicting the effects of adding a new mode to the choice set. The following example illustrates how this is done.

Example 4.7: Introduction of a New Mode

Consider a traveler who can choose between drive alone and carpool. Let the probabilities with which these modes are chosen be given by a binomial logit model in which the deterministic component of the utility function is as in Equation (4.9). Let the the traveler's income be \$20,000 per year, and let the values of the travel time and cost variables be:

<u>Mode</u>	<u>Time (Hours)</u>	<u>Cost (\$)</u>
Drive Alone	0.5	2.00
Carpool	0.6	1.00

Then, the values of V and the choice probabilities are

<u>Mode</u>	<u>V</u>	<u>exp(V)</u>	<u>Pr(mode)</u>
Drive Alone	-1.00	0.37	0.46
<u>Carpool</u>	<u>-0.85</u>	<u>0.43</u>	<u>0.54</u>
Sum		0.80	1.00

Now suppose that bus service is initiated and that it has a travel time of 0.8 hr. and a fare of \$0.60. The probability that a traveler will choose bus and the new probabilities that drive alone and carpool are chosen can be obtained from the multinomial logit model (Equation 4.7) if the value of the deterministic component of bus utility is known. This value can be obtained by substituting the values of T and C for bus into Equation (4.9) to obtain $V_B = -0.95$. The resulting computation of the mode choice probabilities for the three-mode choice set is:

<u>Mode</u>	<u>V</u>	<u>exp(V)</u>	<u>Pr(mode)</u>
Drive Alone	-1.00	0.37	0.31
Carpool	-0.85	0.43	0.36
<u>Bus</u>	<u>-0.95</u>	<u>0.39</u>	<u>0.33</u>
Sum		1.19	1.00

In Example 4.7, the deterministic component of the utility function does not contain alternative-specific constants. In practice, these constants usually are present, which makes it necessary to assign a value to the alternative-specific constant for the new mode before predictions of the effects of adding this mode can be made. This value usually must be assigned judgmentally in practice. Although guidance as to an appropriate range of values sometimes can be obtained by examining mode choice models developed for cities where the new mode already is in operation, there is almost always considerable uncertainty as to the best value to use. As a result, there is likely to be considerable uncertainty as to the effects of introducing the new mode.

4.8 Summary

This module has presented the binomial and multinomial logit models and has explained how they can be used to describe probabilistic choice behavior. Several examples have illustrated the properties of these models, including their sensitivity to changes in the deterministic components of utility, their dependence on utility differences, the importance of alternative-specific constants, the IIA property, and their ability to facilitate prediction of the effects of introducing a new mode.

EXERCISES

4.1 Use Equation (4.3) to compute the probability that alternative 1 is chosen for the following values of the deterministic components of utility:

<u>Case</u>	<u>V₁</u>	<u>V₂</u>
6	1.0	-1.5
7	0.5	-2.0
8	3.0	-1.0
9	0.0	-0.5
10	0.0	2.5

- a. Referring to Table 4.2, compare cases 4, 5, and 7. What can you conclude about the importance of utility values as opposed to utility differences?
 - b. Compare cases 4 and 10. What can you conclude about the relation between these cases?
- 4.2 Repeat the analysis shown in Example 4.1 for the case where the deterministic component of the utility of carpool is 1.5. Repeat again using 1.0 as the deterministic component of the utility of carpool.
- 4.3 Suppose that in Example 4.7 two new modes were added, bus and bicycle. Let the travel time and cost of bus be as in the example. Let the travel time of bicycle be 1.0, and let its cost be 0.
- a. Compute the choice probabilities of all four modes, drive alone, carpool, bus, and bicycle, using Equation (4.9). Assume that income is \$20,000 per year.
 - b. Predict the choice probabilities for all four modes using Equations (4.12) after adding an appropriate equation for the deterministic component of the bicycle utility function. Let the value of the alternative-specific constant for bicycle be -1.0.

MODULE 5

VARIABLES OF MODE CHOICE MODELS

5.1 Introduction

Probabilistic choice models generally, and logit models in particular, make it possible to develop useful mode choice models that do not include all of the variables that influence mode choice. This is a very important property of probabilistic choice models since, as was discussed in Module 3, the variables that influence mode choice are not all known to analysts and not all of the known variables can be measured in practice. It does not follow, however, that a model based on any subset of the influential variables will be useful. On the contrary, there are certain types of variables that must be included to obtain a useful model. This module identifies these classes of variables and explains the forms that the variables can take.

This module also identifies variables that have been found useful in previously developed mode choice models. However, it does not provide a list of standard variables that always should be used in mode choice models or a standard procedure for selecting variables. No such list or procedure exists. The appropriate variables to use in a model depend on the purposes for which the model is to be used and on the available data. They also depend on behavioral relations that normally are revealed in the process of model development. In fact, the process of selecting variables for a practical mode choice model is as much an art as a science. It relies as much on judgment and experience as on statistical techniques. The purpose of this module is to provide information that will contribute to informed

and sensible judgments in the selection of variables. Statistical techniques that can help to guide the selection of variables are discussed in Section 6.5 of Module 6.

Throughout this module, the term "utility function" will mean the deterministic component of the utility function of a logit model. The modifier "deterministic component of the" will not be used.

5.2 Classes of Variables That Must Be Included in Models

There are three kinds of variables that must be included in a model to make it useful: (1) policy variables, (2) variables that affect mode choice and that identify any demographic characteristics or population groups of interest, and (3) other variables that influence mode choice and are correlated with either the policy variables or the variables used to identify demographic characteristics and population groups.

In addition, the utility function of each mode except one should include an alternative-specific constant. The use of alternative-specific constants was discussed in Section 4.6 of Module 4. As was explained in Section 4.6, it does not matter which mode is the selected as the base mode whose utility function does not include an alternative-specific constant. (To minimize the complexity of the subsequent discussion, the utility functions in the examples presented in this module and in Module 6 do not all include alternative-specific constants. The omission of alternative-specific constants from the examples is for reasons of expository clarity and should not be interpreted as contradicting the principle that these constants should be included in the utility functions of practical models.)

5.2a Policy Variables

One of the most important uses of mode choice models is predicting the effects of policy measures. For example, a transportation planner may want to predict the change in bus ridership that will occur if bus fares change or bus travel becomes faster. Such predictions can be made only if the model includes explanatory variables, called policy variables, that represent the policy measures being considered. For example, policy variables such as bus fare and bus travel time must be included in a model to make it useful for predicting the effects of policy measures that would change fares or travel times. A model that will be used to predict the effects of measures to improve the reliability of bus service must include policy variables, such as the percent of on-time arrivals, that can represent the effects of the measures being considered.

Example 5.1: Policy Variables

Consider the following binomial logit model of work-trip mode choice. The available modes are automobile and bus, and the probabilities of choosing these modes are:

$$P_{\text{auto}} = \frac{\exp(-T_a - 5C_a/Y)}{\exp(-T_a - 5C_a/Y) + \exp(-T_b - 5C_b/Y)} \quad (5.1)$$

$$P_{\text{bus}} = 1.0 - P_{\text{auto}}, \quad (5.2)$$

where

T_a, T_b = Automobile and bus travel times in hours;

C_a, C_b = Automobile and bus travel costs in dollars;

Y = Income of the traveler's household in thousands of dollars per year.

In this model, T_a , T_b , C_a , and C_b are policy variables since their values can be influenced by transportation policy measures. Y is not a policy variable since individuals' incomes are not directly influenced by transportation policy measures.

The model of equations (5.1) and (5.2) can be used to predict the effects on mode choice of policy measures that change automobile or bus travel times and costs since travel times and costs are policy variables of the model. In contrast, comfort and safety are not policy variables of the model. The model cannot predict the effects of policy measures designed to influence these variables.

5.2b Variables That Affect Mode Choice and Identify Demographic Characteristics or Population Groups

Often, it is important to be able to predict the effects of policy measures on different groups in the population or to predict the effects of changes in demographic characteristics of the population. For example, it may be important to know whether increasing bus fares will be particularly burdensome to low-income travelers or whether a certain improvement in transit service will succeed in attracting members of multi-car households to transit. A model can answer questions such as these only if it includes variables that permit the effects of policy measures on different population groups of interest to be differentiated. In the example of bus fares just given, the variables income and automobiles owned would serve this purpose. The model discussed in Example 5.1 includes the variable Y (income) and, therefore, is capable of differentiating among the effects of changes in travel time (T) and travel cost (C) on different income groups. The model also is capable of predicting the effects on mode choice of any changes in Y

that may occur in the future. The model does not include a variable for automobile ownership and, therefore, is not capable of differentiating among households with different levels of automobile ownership or of predicting the effects of changes in automobile ownership.

Example 5.2: Effects on Different Income Groups of a Change in Bus Fare

Suppose that the model of Equation (5.2) is used to predict the effects on bus ridership of increasing the fare from \$0.50 to \$1.00 on a certain route. Assume that $C_a = \$1.50$, $T_a = 0.50$ hr., and $T_b = 1.0$ hr. for the affected travelers and that these travelers include individuals whose incomes are \$20,000 and \$40,000 per year. Then, before the fare increase, the probability that the lower-income travelers choose bus is

$$P_{bus}(\text{low, before}) = \frac{\exp[-1.0 - 5(0.5)/20]}{\exp[-0.5 - 5(1.5)/20] + \exp[-1.0 - 5(0.5)/20]} \\ = 0.44,$$

and the probability that the higher-income travelers choose bus is

$$P_{bus}(\text{hi, before}) = \frac{\exp[-1.0 - 5(0.5)/40]}{\exp[-0.5 - 5(1.5)/40] + \exp[-1.0 - 5(0.5)/40]} \\ = 0.41.$$

After the fare increase, the probabilities of bus choice by the low- and high-income travelers are

$$P_{bus}(\text{low, after}) = \frac{\exp[-1.0 - 5(1.0)/20]}{\exp[-0.5 - 5(1.5)/20] + \exp[-1.0 - 5(1.0)/20]} \\ = 0.41,$$

and

$$P_{bus}(\text{hi, after}) = \frac{\exp[-1.0 - 5(1.0)/40]}{\exp[-0.5 - 5(1.5)/40] + \exp[-1.0 - 5(1.0)/40]} \\ = 0.39.$$

Therefore, the fare increase is predicted to cause a reduction of 7 percent ($100 \times 0.03 / 0.44$) in the probability that a low-income traveler chooses bus but only a 5 percent ($100 \times 0.02 / 0.41$) reduction in the probability that a high-income traveler chooses bus. Accordingly, the fare increase is predicted to have a greater impact on the low-income travelers than on the high-income travelers. This prediction could not have been made if the model did not include the variable Y that enables income groups to be distinguished.

5.2c Other Variables That Influence Mode Choice and Are Correlated with the Policy, Demographic, or Grouping Variables

Frequently, it is possible to identify variables that are not of interest themselves but that affect mode choice and are correlated with one or more of the policy or grouping variables in a model. Such variables also must be included in the model or else the model will give incorrect predictions of the effects of the policy and grouping variables. For example, suppose that travel time (a policy variable) and income (a grouping variable) are included in a model of mode choice. Suppose, also, that the number of automobiles owned by a traveler's household -- a variable that is known to have a strong effect on mode choice -- is not of interest in a particular study but that multi-car households tend to have higher incomes and to live farther from their workplaces than do single-car and non-car-owning households. Since automobile ownership has an independent effect on mode choice, apart from its association with travel time and income, a model that does not include automobile ownership as a variable will give incorrect predictions of the effects of travel time and income on mode choice. Such a model's predictions of the effects of travel time will reflect not only the

true effects of travel time but, also, the effects of differences in automobile ownership that are associated with differences in travel time through the tendency of households with large travel times to own many cars. In other words, the travel time variable will operate, in part, as a surrogate for automobile ownership and, therefore, will not correctly describe the true effects of changes in travel time alone. Similarly, the model's predictions of the effects of income on mode choice will reflect both the true effects of income and the effects of differences in automobile ownership that are associated with differences in income through the tendency of high-income households to own many cars.

The prediction errors caused by omitting a variable that is correlated with a policy variable are further illustrated by the following example.

Example 5.3: Effect of Omitting a Variable That Is Correlated with a Policy Variable

Suppose a model of choice between the modes automobile and bus has the binomial logit form:

$$P_{\text{auto}} = \frac{\exp(V_a)}{\exp(V_a) + \exp(V_b)} \quad (5.3)$$

$$P_{\text{bus}} = \frac{\exp(V_b)}{\exp(V_a) + \exp(V_b)} \quad (5.4)$$

Let V_a and V_b , the automobile and bus utility functions, have the forms

$$V_a = -T_a + 0.5A \quad (5.5)$$

$$V_b = -T_b, \quad (5.6)$$

where T_a and T_b , respectively, are automobile and bus travel time in hours, and A is the number of automobiles owned by the traveler's household.

Substitution of Equations (5.5) and (5.6) into Equation (5.3), followed by some algebra, yields

$$P_{\text{auto}} = 1 / \{1 + \exp[-(T_b - T_a) - 0.5A]\}. \quad (5.7)$$

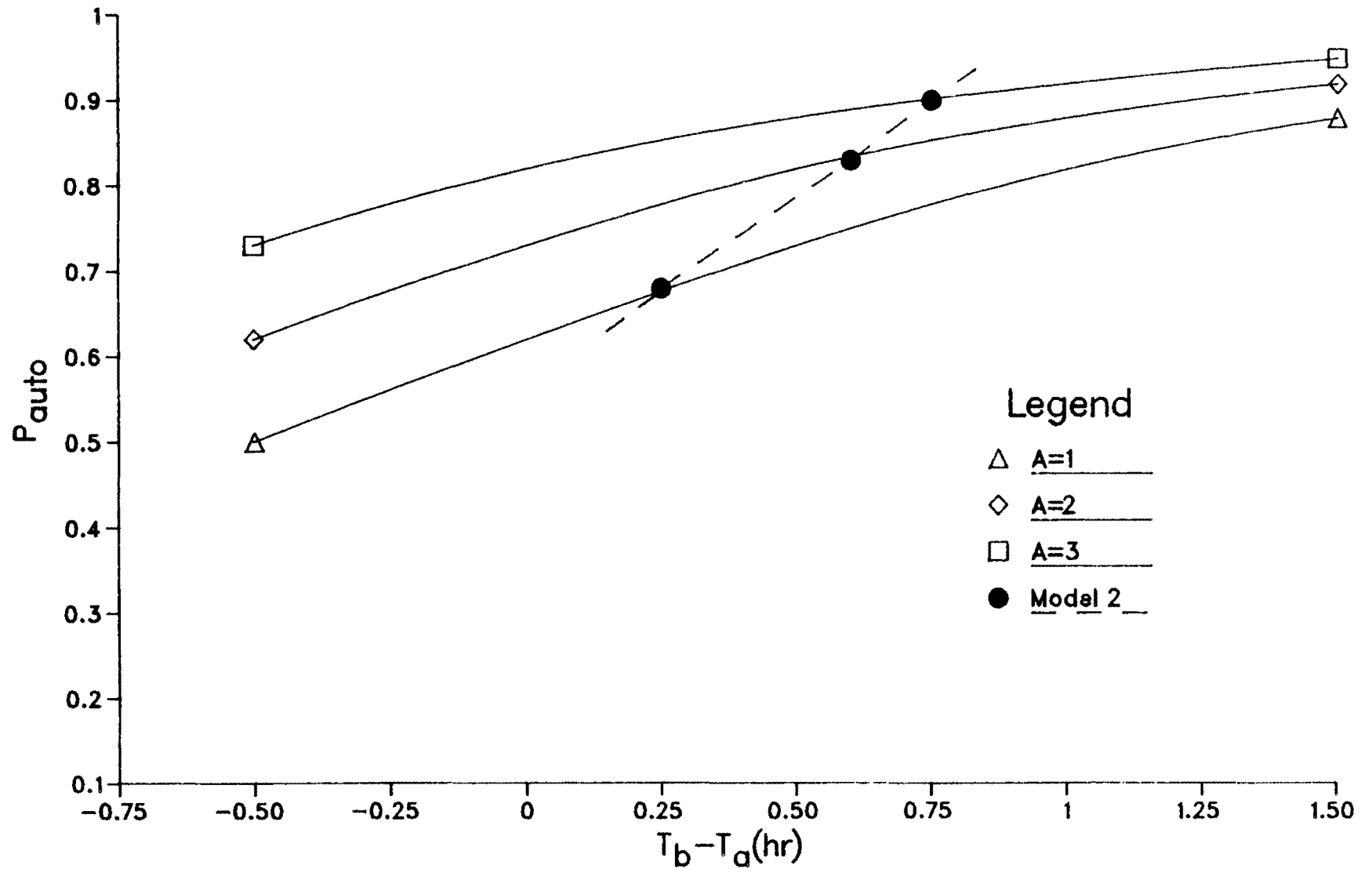
The solid lines in Figure 5.1 show graphs of P_{auto} as a function of $T_b - T_a$ for each of three different values of A . As expected, the graphs show that, given the same value of $T_a - T_b$, an individual whose household owns several automobiles has a higher probability of choosing automobile than does an individual whose household owns only one automobile.

Now suppose that the available data consist of measurements of P_{auto} , $T_b - T_a$, and A for three groups of individuals, as follows:

<u>Group</u>	<u>$T_b - T_a$</u>	<u>A</u>	<u>P_{auto}</u>
1	0.25	1	0.68
2	0.60	2	0.83
3	0.75	3	0.90

These data are plotted as large dots in Figure 5.1. Notice that increases in automobile ownership are associated with increases in $T_b - T_a$. In other words, automobile ownership is positively correlated with $T_b - T_a$. A model based on these data that included the policy variable $T_b - T_a$ but not the automobile ownership variable A would conclude that the relation between P_{auto} and $T_b - T_a$ is the one obtained by connecting the three dots. This relation is shown by the dashed line in Figure 5.1. Notice that this line is much steeper than the solid lines. In other words, the model that omits the automobile ownership variable predicts that policy changes (i.e., changes in $T_b - T_a$) have larger effects on mode choice than they really have. The model makes this prediction error because, owing to the omission

Figure 5.1: P_{auto} vs $T_b - T_a$ for Example 5.3



of the variable A, the predicted effects of changes in $T_b - T_a$ reflect not only the true effects of changes in this variable, but also the effects of the changes in A that are associated in the data with changes in $T_a - T_b$.

5.2d Alternative-Specific Constants

An alternative-specific constant is a constant that is added to the utility function of a mode and whose numerical value may be different for different modes. Example 4.3 in Module 4 illustrates the use of alternative-specific constants. For example, in the utility function

$$V_{DA} = 0.8 - T_{DA} - 5C_{DA}/Y \quad (5.8a)$$

$$V_{CP} = 0.2 - T_{CP} - 5C_{CP}/Y \quad (5.8b)$$

$$V_B = - T_B - C_B/Y, \quad (5.8c)$$

the alternative-specific constants are 0.8 for drive alone and 0.2 for carpool.

Alternative-specific constants provide a convenient way to account for the average effects of all variables affecting choice that are not explanatory variables of the model. The number of alternative-specific constants in a model must not exceed the number of modes in the model minus one. The predictions of the model are the same, regardless of which mode is selected to be the one that has no alternative-specific constant in its utility function. The following example illustrates the prediction errors that can occur when alternative-specific constants are not included in a model.

Example 5.4: Alternative-Specific Constants

Suppose that choice among the modes automobile and bus is described by a logit model. Let the logit utility function be

$$V_a = 0.5 - T_a \quad (5.9a)$$

$$V_b = -T_b, \quad (5.9b)$$

where T_a and T_b , respectively, denote automobile and bus travel time, and 0.5 is the value of the alternative-specific constant in the automobile utility function. The probability that automobile is chosen is

$$P_{\text{auto}} = \frac{\exp(0.5 - T_a)}{\exp(0.5 - T_a) + \exp(-T_b)}. \quad (5.10)$$

Equivalently,

$$P_{\text{auto}} = \frac{1}{1 + \exp[-(T_b - T_a) - 0.5]}. \quad (5.11)$$

The solid line in Figure 5.2 shows a graph of P_{auto} as a function of $T_b - T_a$.

Now suppose that the available data consist of observations of the mode choices of individuals for whom $T_b - T_a = 0.50$ hr. The probability that such individuals choose automobile can be computed from equation (5.11) and is

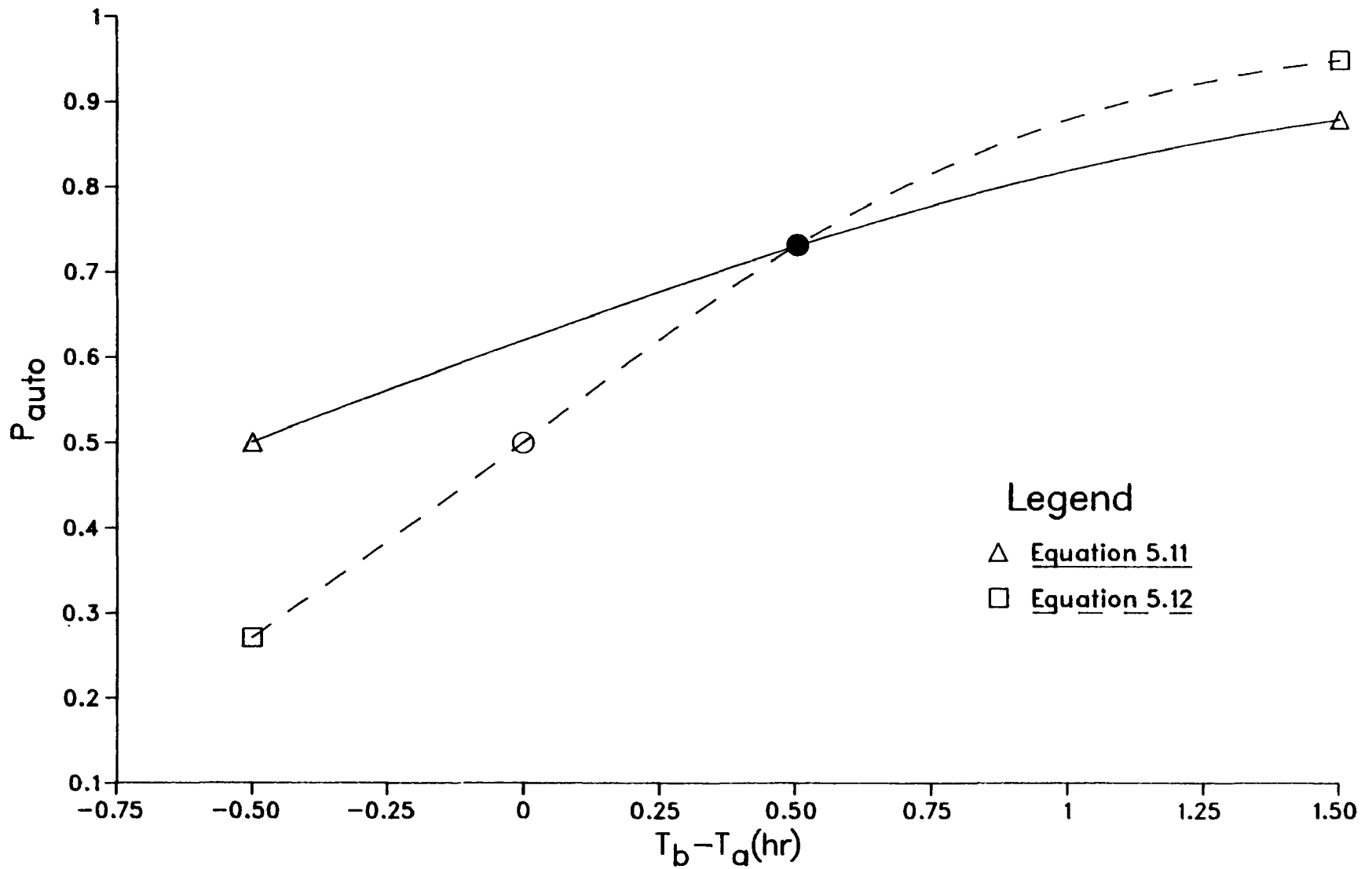
$$P_{\text{auto}} = \frac{1}{1 + \exp(-0.5 - 0.5)} = 0.73.$$

The point $P_{\text{auto}} = 0.73$, $T_b - T_a = 0.50$ is identified by the solid dot in Figure 5.2. A logit model of choice between automobile and bus that did not include an alternative-specific constant would have the form

$$P_{\text{auto}} = \frac{1}{1 + \exp[-c(T_b - T_a)]}, \quad (5.12)$$

where c is a positive constant. According to this model, $P_{\text{auto}} = 0.5$ when $T_b - T_a = 0$. The open dot in Figure 5.2 identifies the point $P_{\text{auto}} = 0.5$, $T_b - T_a = 0$. The logit model without an alternative specific constant that fits the observed choices is the one illustrated by the dashed line in Figure 5.2. This line corresponds to equation (5.12) with $c = 2.0$. Notice that the dashed line is much steeper than the solid line. In other words,

Figure 5.2: P_{auto} vs $T_b - T_a$ for Example 5.4



the model without the alternative-specific constant predicts that changes in travel time have larger effects on mode choice than they really have.

5.3 Functional Forms and Disaggregation of Variables

An important aspect of selecting variables for a model is deciding how these variables should depend on the observed attributes of modes and individuals. For example, the attribute "travel time" might be represented in a model by the variable T (travel time measured in, say, hours) or it might be represented by the variable $\ln(T)$, the natural logarithm of T . Alternatively, T might be disaggregated into the components in-vehicle travel time, walk time, wait time, transfer time, etc., and each of these components (or possibly its logarithm or some other transformation) used as a separate variable of the model. The form in which an attribute such as travel time enters a model frequently has important behavioral implications, and it can have a large effect on the model's forecasts.

Although the need to decide the relation between an attribute and the variable representing it can arise with virtually any attribute that might enter a model, the attributes that seem to cause the greatest difficulty in practical mode choice modeling are travel time, travel cost, income, and automobile ownership. This section describes the variables most frequently used to represent these attributes in practice and explains the implications of different choices of variables.

5.3a Travel Time -- Disaggregation

The most important decision that must be made with respect to travel time is whether it should be disaggregated into components (e.g., in-vehicle travel time, walk time, etc.) and, if so, what the components should be.

Disaggregation admits the possibility that equal changes in different components of travel time may have different effects on mode choice. For example, disaggregating travel time into in-vehicle and out-of-vehicle components admits the possibility that a 5 minute increase in in-vehicle travel time and a 5 minute increase in out-of-vehicle travel time have different effects on mode choice. (In fact, experience indicates that travelers consider out-of-vehicle travel time to be more burdensome than in-vehicle travel time, so a 5 minute increase in out-of-vehicle travel time does have a greater effect on mode choice than does a 5 minute increase in in-vehicle travel time.) Similarly, disaggregating out-of-vehicle travel time into the components walk time and wait time admits the possibility that equal changes in the values of these components have different effects on mode choice. In contrast, representation of travel time by the single variable "total travel time" is equivalent to assuming that equal changes in the various components of travel time have equal effects on mode choice. Similarly, combining all of the components of out-of-vehicle travel time into the single variable "total out-of-vehicle travel time" is equivalent to assuming that equal changes in the various components of out-of-vehicle travel time (e.g., walk time, wait time, etc.) have equal effects on mode choice.

Example 5.5: Components of Travel Time

Consider the following two logit models of choice between automobile and bus:

$$\text{Model 1 -- } P_{\text{auto}} = \frac{\exp(-T_a)}{\exp(-T_a) + \exp(-T_b)} \quad (5.13)$$

$$\text{Model 2 -- } P_{\text{auto}} = \frac{\exp(-0.48TI_a - 1.21TO_a)}{\exp(-0.48TI_a - 1.21TO_a) + \exp(-0.48TI_b - 1.21TO_b)} \quad (5.14)$$

$$\text{Models 1 and 2 -- } P_{\text{bus}} = 1.0 - P_{\text{auto}} \quad (5.15)$$

In these models,

T_a, T_b = Total travel time by auto and bus in hours;

TI_a, TI_b = In-vehicle travel time by auto and bus in hours;

TO_a, TO_b = Out-of-vehicle travel time by auto and bus in hours.

Thus, Model 2 disaggregates travel time into the components in-vehicle travel time and out-of-vehicle travel time, whereas Model 1 does not disaggregate travel time. Equations (5.13) and (5.14) are equivalent to

$$\text{Model 1 -- } P_{\text{auto}} = 1/(1 + \exp[-(T_b - T_a)]) \quad (5.16)$$

$$\text{Model 2 -- } P_{\text{auto}} = 1/(1 + \exp[-0.48(TI_b - TI_a) - 1.21(TO_b - TO_a)]). \quad (5.17)$$

Suppose that at present, $TI_b = 0.5$ hr., $TI_a = 0.4$ hr., $TO_b = 0.30$ hr., and $TO_a = 0.05$ hr. (the base case). Then $T_b = 0.80$ hr., and $T_a = 0.45$ hr.

In Model 1,

$$P_{\text{auto}} = 1/(1 + \exp[-(0.80 - 0.45)]) = 0.59,$$

and in Model 2,

$$P_{\text{auto}} = 1/(1 + \exp[-0.48(0.5 - 0.4) - 1.21(0.30 - 0.05)]) = 0.59.$$

The probability that automobile is chosen is the same in both models.

Now consider the effects of increasing TI_b (in-vehicle travel time) by 0.1 hr. while TO_b (out-of-vehicle travel time) remains unchanged and of increasing TO_b by 0.1 hr. while TI_b remains unchanged. The values of TI_a and TO_a remain as before. The following table shows the new values of P_{auto} obtained from the two models as well as the value of P_{auto} in the base case:

<u>Case</u>	P_{auto} According to		Change from Base Case	
	<u>Model 1</u>	<u>Model 2</u>	<u>Model 1</u>	<u>Model 2</u>
Base	0.59	0.59	0.0	0.0
Increase TI_b	0.61	0.60	0.02	0.01
Increase TO_b	0.61	0.62	0.02	0.03

It can be seen that according to Model 1, which does not disaggregate travel time into its components, increasing TI_b by 0.1 hr. and increasing TO_b by 0.1 hr. have the same effect on mode choice -- they both increase P_{auto} by 0.02. However, in Model 2, which does disaggregate travel time, increasing TI_b by 0.1 hr. increases P_{auto} by only 0.01, whereas increasing TO_b by 0.1 hr. increases P_{auto} by 0.03. In other words, in Model 2, the increase in out-of-vehicle travel time has three times the effect of the same increase in in-vehicle travel time. In Model 1, travelers are equally sensitive to changes in in-vehicle and out-of-vehicle travel time. In Model 2, travelers are more sensitive to changes in out-of-vehicle travel time than to changes in in-vehicle travel time.

The differences between Models 1 and 2 have practical policy consequences. For example, suppose that Model 2 is correct. Then use of Model 1 overstates the importance of in-vehicle travel time savings and understates the importance of out-of-vehicle travel time savings. Use of Model 1 will cause a bus operator trying to increase service quality and ridership to place too much emphasis on in-vehicle travel time and too little on out-of-vehicle travel time.

5.3b Travel Time -- Mode-Specific Representation

An issue that is related to the disaggregation issue is whether travel time (or one or more of its components) should be represented as a generic

or a mode-specific variable. Travel time (or one of its components) is generic if it is represented by the same variable in all modes. It is mode-specific if it is represented by different variables in different modes. For example, automobile travel time and bus travel time might be represented by separate variables in the utility function, with the value of the automobile travel time variable being zero for the transit mode and the value of the transit travel time variable being zero for the automobile mode.

Use of a mode-specific variable to represent an attribute admits the possibility that travelers evaluate that attribute differently for different modes. Use of a generic variable excludes this possibility. In-vehicle travel time is a travel time component that sometimes is represented by a mode-specific variable in mode choice models. The behavioral rationale for this is that transit travelers often can spend their in-vehicle time reading or sleeping, whereas automobile travelers (particularly if they are drivers) may not be able to do these things. Therefore, travelers may perceive transit in-vehicle travel time as being less burdensome than automobile in-vehicle travel time.

Example 5.6: Generic and Mode-Specific Travel Time Variables

Consider the following two models of choice between automobile and bus:

$$\text{Model 1 -- } P_{\text{auto}} = \frac{\exp(-T_a)}{\exp(-T_a) + \exp(-T_b)} \quad (5.18)$$

$$= 1/(1 + \exp[-(T_b - T_a)]) \quad (5.19)$$

$$\text{Model 2 -- } P_{\text{auto}} = \frac{\exp(-4.2TA_a - 2.8TB_a)}{\exp(-4.2TA_a - 2.8TB_a) + \exp(-4.2TA_b - 2.8TB_b)} \quad (5.20)$$

$$= 1/(1 + \exp[-4.2(TA_b - TA_a) - 2.8(TB_b - TB_a)]). \quad (5.21)$$

$$\text{Models 1 and 2 -- } P_{\text{bus}} = 1.0 - P_{\text{auto}} \quad (5.22)$$

In these models,

T_a, T_b = Value of the generic variable "total travel time" for automobile (a) and bus (b).

TA_a, TA_b = Value of the mode-specific variable "automobile travel time" for automobile (a) and bus (b). $TA_b = 0$ always since no time is spent traveling by automobile if bus is chosen. The value of TA_a is the same as the value of T_a .

TB_a, TB_b = Value of the mode-specific variable "bus travel time" for automobile (a) and bus (b). $TB_a = 0$ always since no time is spent traveling by bus if automobile is chosen. The value of TB_b is the same as the value of T_b .

Since $TA_b = TB_a = 0$, $TA_a = T_a$, and $TB_b = T_b$, Model 2 is equivalent to

$$\text{Model 2 -- } P_{\text{auto}} = 1/[1 + \exp(4.2T_a - 2.8T_b)]. \quad (5.23)$$

The difference between Models 1 and 2 is that in Model 1, travel time is a generic variable, whereas travel time is a mode-specific variable in Model 2.

Suppose that at present, $T_b = 0.80$ hr. and $T_a = 0.45$ hr. (the base case). Then in Model 1,

$$P_{\text{auto}} = 1/(1 + \exp[-(0.80 - 0.45)]) = 0.59,$$

and in Model 2,

$$P_{\text{auto}} = 1/(1 + \exp[4.2(0.45) - 2.8(0.80)]) = 0.59.$$

The probability that automobile is chosen is the same in both models.

Now consider the effects of increasing T_b by 0.1 hr. while T_a remains unchanged and of decreasing T_a by 0.1 hr. while T_b remains unchanged. The following table shows the new values of P_{auto} obtained from the two models as well as the value of P_{auto} in the base case:

<u>Case</u>	<u>P_{auto} According to</u>		<u>Change from Base Case</u>	
	<u>Model 1</u>	<u>Model 2</u>	<u>Model 1</u>	<u>Model 2</u>
Base	0.59	0.59	0.0	0.0
Increase T _b	0.61	0.65	0.02	0.06
Decrease T _a	0.61	0.68	0.02	0.09

It can be seen that according to Model 1, in which travel time is a generic variable, increasing T_b by 0.1 hr. and decreasing T_a by 0.1 hr. have the same effect on mode choice -- they both increase P_{auto} by 0.02. However, in Model 2, which treats travel time as a mode-specific variable, increasing T_b increases P_{auto} by 0.06, whereas decreasing T_a increases P_{auto} by 0.09. In other words, the change in automobile travel time has a 50 percent larger effect than an equal but opposite change in bus travel time. In Model 1, travelers are equally sensitive to changes in automobile and bus travel time. In Model 2, travelers are more sensitive to changes in automobile travel time than to changes in bus travel time.

5.3c Travel Time -- Functional Form

The final decision that must be made about travel time variables is what the functional form of the relation between the variables and the physically measured attributes should be. Two functional forms are used frequently in practice, the linear and the logarithmic. In the linear form, the travel time variable is simply measured travel time (T), and in the logarithmic form, the variable is the natural logarithm of measured travel time (ln(T)). T can represent either total travel time or any component of travel time. Adoption of the linear form is equivalent to assuming that travelers find a given increase in T equally burdensome regardless of the current value of T. For example, if T denotes total travel time, use of

the linear form means that adding 5 minutes to a 1-hr. trip is perceived by travelers as being just as burdensome as adding 5 minutes to a 10-min. trip. In the logarithmic form, travelers find a given percentage increase in T to be equally burdensome regardless of the current value of T. Thus, for example, adding 5 minutes to a 10-min. trip -- a 50 percent increase in travel time -- is perceived as being just as burdensome as adding 30 minutes to a 1-hr. trip.

The linear and logarithmic forms of travel time yield predictions of mode choice that may be very different from one another, as is illustrated by the following example.

Example 5.7: Linear and Logarithmic Forms of Travel Time

Consider the following two models of choice between two modes that will be called mode 1 and mode 2:

$$\text{Model 1 -- } P_1 = \frac{\exp(-T_1)}{\exp(-T_1) + \exp(-T_2)} \quad (5.24)$$

$$= 1/(1 + \exp[-(T_2 - T_1)]) \quad (5.25)$$

$$\text{Model 2 -- } P_1 = \frac{\exp(-\ln T_1)}{\exp(-\ln T_1) + \exp(-\ln T_2)} \quad (5.26)$$

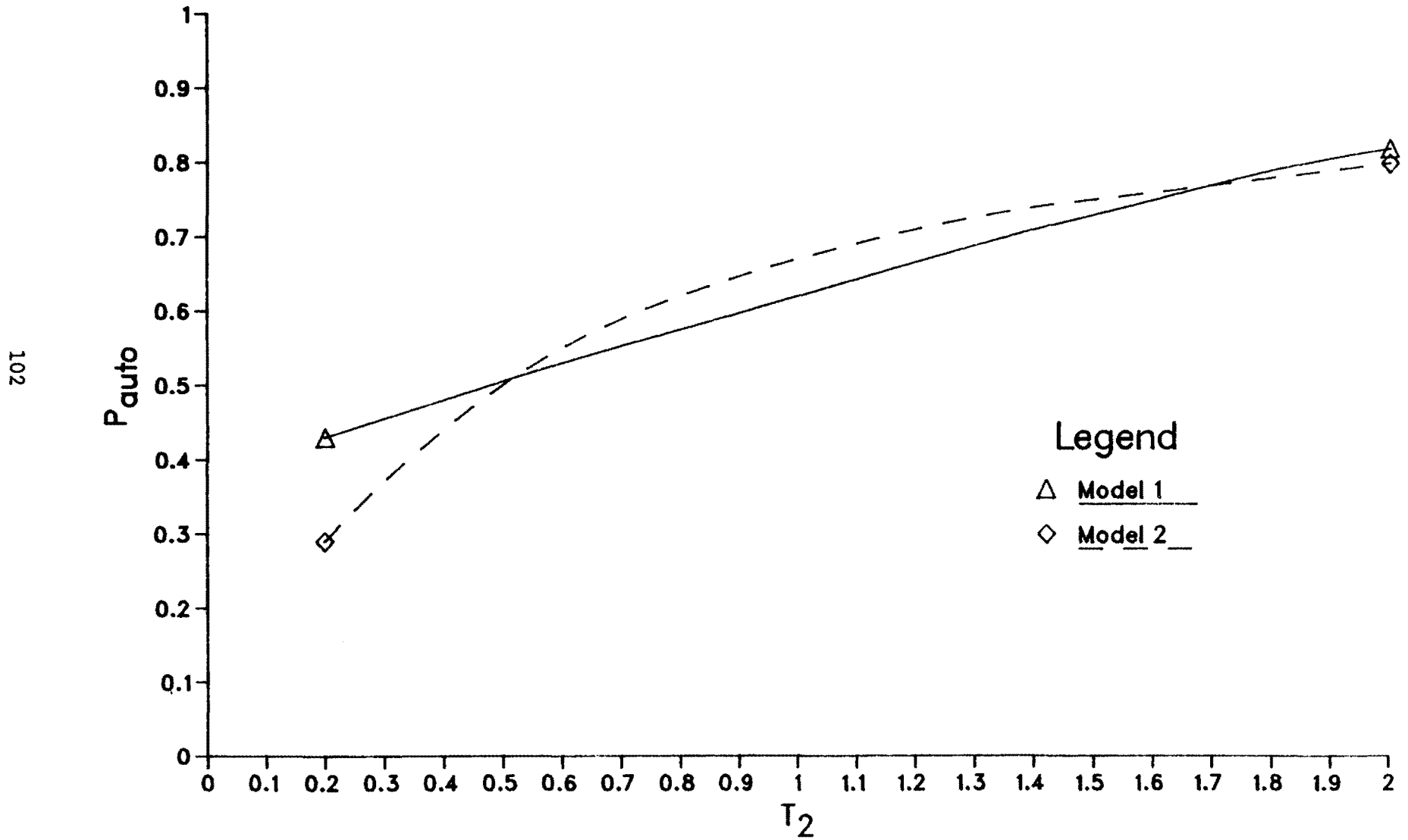
$$= 1/(1 + \exp[-\ln(T_2/T_1)]) \quad (5.27)$$

$$= 1/(1 + T_1/T_2) \quad (5.28)$$

$$\text{Models 1 and 2 -- } P_2 = 1.0 - P_1, \quad (5.29)$$

where T_1 and T_2 denote total travel time in hours by modes 1 and 2. Figure 5.3 shows a graph of the relation between P_1 and T_2 for each model when $T_1 = 0.5$ hr. It can be seen from the figure that when mode 1 is the faster mode, the two models yield similar values of P_1 . However, the values of P_1

Figure 5.3: Linear vs Logarithmic Models of Travel Time



obtained from the two models differ greatly when T_2 is less than about 0.3 hr.

5.3d Travel Cost and Income

Like travel time, travel cost can be divided into components (e.g., automobile fuel and maintenance costs, parking costs, tolls and fares, etc.) and can be represented as either a generic or mode-specific attribute. In most mode choice models, however, travel cost is treated as generic and is not divided into components.

An important consideration in the selection of travel cost variables is their interaction with income. Economic theory and everyday experience both suggest that a traveler's sensitivity to changes in travel costs may depend on his income, with high-income travelers being less sensitive than low-income ones. To represent this income dependence, the travel cost variable in mode choice models often takes the form C/Y , where Y is the total or after-tax income of the traveler's household. (After-tax income is the better variable to use, because only after-tax income can be allocated among travel and non-travel uses at the discretion of the household. After-tax income can be computed from total income -- the only type of income data normally available to transportation planners -- if the proportion of total income paid in taxes as a function of income level is known. Average values of this proportion are published in Vital Statistics of the United States.)

Income also can be used as a surrogate for unobserved personal attributes that affect choice. For example, suppose that commuters whose households have high incomes have jobs or tend to engage in other activities that cause them to particularly value the schedule flexibility provided by the automobile. This tendency can be represented in a model of choice

between automobile and transit by adding to the utility function of the automobile mode a variable equal to the income (or, possibly, the logarithm of the income) of the traveler's household. Such an income variable then acts as a surrogate for unobserved attributes, such as a preference for schedule flexibility, that tend to make the automobile mode particularly attractive to high-income travelers. Either total income or after-tax income can be used for this purpose.

When income is used in this way, it always is a mode-specific variable, and there must always be at least one mode whose utility function does not contain such a variable. Thus, for example, in a model of choice between automobile and transit, income can be a variable of either the automobile or the transit utility function but not of both utility functions. In a model of choice between drive-alone, carpool, and transit, income can enter the utility functions of any two of the three modes. The model yields the same predictions of mode choice, regardless of which alternatives are assigned the mode-specific income variables and which alternative has no such variable.

The following example illustrates the use of income as a surrogate for unobserved personal attributes that affect mode choice.

Example 5.8: Use of the Income Variable

In a logit model of choice between automobile and bus, let the utility function be:

$$V_a = -T_a - 5C_a/Y + 0.001Y \quad (5.30a)$$

$$V_b = -T_b - 5C_b/Y + 0.001Y, \quad (5.30b)$$

where T, C, and Y, respectively, denote travel time in hours, travel cost in dollars, and after-tax income in thousands of dollars per year. Notice that

the additive income term $0.001Y$ is present in the utilities of both modes.

The probability that automobile is chosen is

$$P_{\text{auto}} = \frac{\exp(-T_a - 5C_a/Y + 0.001Y)}{\exp(-T_a - 5C_a/Y + 0.001Y) + \exp(-T_b - 5C_b/Y + 0.001Y)} \quad (5.31)$$

Dividing the numerator and denominator of equation (5.30) by $\exp(-T_a - 5C_a/Y - 0.001Y)$ yields the equivalent model

$$P_{\text{auto}} = \frac{1}{1 + \exp[-(T_a - T_b) - 5(C_a - C_b)/Y]} \quad (5.32)$$

Notice that the additive income term is no longer present in the model: it has cancelled out in the division and, therefore, has no effect on the probability that auto is chosen or the probability that transit is chosen. Income affects the choice probabilities only through the term $5(C_a - C_b)/Y$ in which income interacts with travel cost.

Now suppose that the utilities are

$$V_a = -T_a - 5C_a/Y + 0.001Y \quad (5.33a)$$

$$V_b = -T_b - 5C_b/Y \quad (5.33b)$$

In this case the additive income term is present in the utility function of only one of the two modes. The probability that automobile is chosen is

$$P_{\text{auto}} = \frac{\exp(-T_a - 5C_a/Y + 0.001Y)}{\exp(-T_a - 5C_a/Y + 0.001Y) + \exp(-T_b - 5C_b/Y)} \quad (5.34)$$

Dividing the numerator and denominator of (5.34) by $\exp(-T_a - 5C_a/Y + 0.001Y)$ yields the equivalent model

$$P_{\text{auto}} = \frac{1}{1 + \exp[-(T_b - T_a) - 5(C_b - C_a)/Y - 0.001Y]} \quad (5.35)$$

Notice that the additive income term is present in equation (5.35). It has not cancelled in the division. Its form is such that P_{auto} increases when Y increases, as is to be expected from the form of equations (5.33) for the utility function. Thus, the additive income term affects the choice

probabilities when it is excluded from the utility function of one mode but not when it is included in the utility functions of all modes.

5.3e Automobile Ownership

The automobile ownership variable in mode choice models usually takes one of the following three forms:

1. A -- Total number of automobiles owned by the traveler's household
2. A/LD -- Number of automobiles per licensed driver in the traveler's household
3. A/W -- Number of automobiles per worker in the traveler's household.

The second two forms represent the possibility that as the number of licensed drivers or workers in a household increases, the likelihood that any particular individual can have the use of an automobile decreases.

Regardless of which form of the automobile ownership variable is used, it is usually mode-specific and enters the utility function additively. As with additive income variables, there must be, at most, one fewer additive automobile ownership variables than there are modes in the model.

The following example illustrates the use of automobile ownership in a mode choice model.

Example 5.9: Automobile Ownership Variables

Suppose that choice among the modes drive alone, carpool, and bus is described by a trinomial logit model. Let the utility function be

$$V_{DA} = -T_{DA} - 5C_{DA}/Y + 0.1A/W \quad (5.36a)$$

$$V_{CP} = -T_{CP} - 5C_{CP}/Y + 0.05A/W \quad (5.36b)$$

$$V_B = -T_B - 5C_B/Y, \quad (5.36c)$$

where T and C, respectively, denote travel time in hours and travel cost in dollars, Y denotes the after-tax income of the traveler's household in thousands of dollars per year, and A/W denotes the number of automobiles per worker in the traveler's household. Notice that A/W is mode specific (that is, its coefficient is different for different modes) and that it enters the utility functions of only two of the three modes. The choice probabilities are:

$$P_{DA} = \frac{\exp(V_{DA})}{\exp(V_{DA}) + \exp(V_{CP}) + \exp(V_B)} \quad (5.37a)$$

$$P_{CP} = \frac{\exp(V_{CP})}{\exp(V_{DA}) + \exp(V_{CP}) + \exp(V_B)} \quad (5.37b)$$

$$P_B = \frac{\exp(V_B)}{\exp(V_{DA}) + \exp(V_{CP}) + \exp(V_B)} \quad (5.37c)$$

5.4 Other Variables

Many other variables, in addition to those already discussed, can be included in disaggregate mode choice models. Table 5.1 lists some of the variables that have been used in such models in the past.

As an illustration of the use of some of these variables (no model includes all of them), Table 5.2 shows the utility function of a mode choice model that was developed using data from the San Francisco area. (This model includes more variables than do many mode choice models, which is why it was picked for this illustration.) There are three modes in the model: drive alone, carpool, and transit. Access to transit can be either on foot or by automobile. The table shows the variables of the model and the values

of their coefficients in the utility function. Thus, the utility function for mode i is

$$V_i = -4.697X_{i,DA} - 3.658X_{i,CP} - 21.43C/Y - 0.0122IVTT \\ - \dots + 0.0327NW + 0.0000137YD. \quad (5.38)$$

The roles of the variables of the model that have not previously been discussed in this course are as follows. HD1 and HD2 represent the inconvenience of having to wait for transit. These variables account for the effect of headway on wait time and, through its effect on wait time, on mode choice. The choice of headway variables is based on the assumption that an increase in transit headway when the headway exceeds 8 min. is less onerous than an equal increase when the headway is less than 8 min. (e.g., because when the headway exceeds 8 min., additional waiting time can be spent at home or the office, rather than at the transit stop). The relative magnitudes of the coefficients of HD1 and HD2 support this assumption. CBD1 and CBD2 represent the effects of variables other than in-vehicle travel time and walk time that may make automobile travel to the CBD particularly difficult. Examples of such variables are the need to drive in heavily congested traffic and the difficulty of finding a parking space. CA accounts for the possibility that the random components of the utilities of transit with auto access and transit with walk access not have the same average values. PW captures the possibility that the "primary worker" in a household may have a greater claim to an automobile than do other household members. NW captures the possibility that members of a multi-worker household can form a carpool among themselves, thereby reducing the difficulty of carpool formation and increasing the probability that the carpool mode is chosen.

5.5 The Selection of Choice Sets

A problem related to that of selecting explanatory variables for a model is that of selecting the choice set. In principle, a traveler's choice set consists of every mode whose probability of being chosen exceeds zero. In practice, this can include a large number of infrequently chosen modes for which data acquisition may be difficult (e.g., walk, bicycle, boat). Except in studies where these modes are of particular interest, little is lost by excluding them from the choice set (i.e., making the approximation that they are never chosen). Thus, in practice, the choice set contains every mode whose probability of being chosen is large enough to be practically significant.

Even when obviously infrequently used modes (e.g., boat) are excluded, selecting choice sets can present difficult decisions. For example, should drive alone be included in the choice set of a traveler whose household does not own an automobile? The answer is no if there is no significant likelihood that such a traveler has access to an automobile. However, it may be yes if substantial numbers of non-automobile-owning travelers borrow or lease cars or drive cars provided by their employers. (The difficulty of deciding whether drive alone should be included in the choice set is greatly reduced if the data include information on the number of cars available to a household, including cars not owned. Drive alone usually can be safely excluded from the choice set of a traveler whose household has no car available.)

There are no rigorous analytic methods for assigning choice sets to travelers. The assignment must be based mainly on the experience and judgment of the analyst. This judgment should be exercised carefully since, as is illustrated by the following example, the choice set can have a

substantial effect on a model's choice probabilities. The method used to assign choice sets in applying a model must be the same as the method used in developing it.

Example 5.10: The Effect of the Choice Set on Choice Probabilities

Suppose that for travelers who have access to all three modes, choice among the modes drive alone, carpool, and bus is described by a multinomial logit model in which the utility function is

$$V_{DA} = -T_{DA} - 5C_{DA}/Y + 0.1A/W \quad (5.39a)$$

$$V_{CP} = -T_{CP} - 5C_{CP}/Y + 0.05A/W \quad (5.39b)$$

$$V_B = -T_B - 5C_B/Y, \quad (5.39c)$$

where T_i is the travel time (in hours) by mode i , C_i is the cost (in dollars) of travel by mode i , Y is the income of the traveler's household in thousands of dollars per year, and A/W is the number of automobiles that the traveler's household owns per worker in the household. Then, the choice probabilities are

$$P_{DA} = \frac{\exp(V_{DA})}{\exp(V_{DA}) + \exp(V_{CP}) + \exp(V_B)} \quad (5.40)$$

$$P_{CP} = \frac{\exp(V_{CP})}{\exp(V_{DA}) + \exp(V_{CP}) + \exp(V_B)} \quad (5.41)$$

$$P_B = \frac{\exp(V_B)}{\exp(V_{DA}) + \exp(V_{CP}) + \exp(V_B)} \quad (5.42)$$

It is a consequence of the IIA property of the logit model that if Equations (5.40) - (5.42) describe mode choice by a traveler with access to all three modes, then mode choice by a traveler who lacks access to drive alone (i.e., a traveler whose choice set consists of carpool and bus but not drive alone) is described by the following binomial logit model:

$$P_{CP} = \frac{\exp(V_{CP})}{\exp(V_{CP}) + \exp(V_B)} \quad (5.43)$$

$$P_B = \frac{\exp(V_B)}{\exp(V_{CP}) + \exp(V_B)} \quad (5.44)$$

In this model $P_{DA} = 0$.

Now suppose that for a particular traveler, $T_{DA} = 0.50$ hr., $T_{CP} = 0.60$ hr, $T_B = 1.0$ hr., $C_{DA} = \$1.00$, $C_{CP} = C_B = \$0.50$, $Y = 20$, and $A = 0$. The traveler's household owns no car. If drive alone is included in this traveler's choice set, the choice probabilities are

$$P_{DA} = \frac{\exp[-0.5 - 5(1/20)]}{\exp[-0.5 - 5(1/20)] + \exp[-0.6 - 5(0.50/20)] + \exp[-1.0 - 5(0.50/20)]} = 0.37$$

$$P_{CP} = \frac{\exp[-0.6 - 5(0.50/20)]}{\exp[-0.5 - 5(1/20)] + \exp[-0.6 - 5(0.50/20)] + \exp[-1.0 - 5(0.50/20)]} = 0.38$$

$$P_B = \frac{\exp[-1.0 - 5(0.50/20)]}{\exp[-0.5 - 5(1/20)] + \exp[-0.6 - 5(0.50/20)] + \exp[-1.0 - 5(0.50/20)]} = 0.25.$$

If drive alone is not included in the choice set, the choice probabilities are

$$P_{DA} = 0$$

$$P_{CP} = \frac{\exp[-0.6 - 5(0.50/20)]}{\exp[-0.6 - 5(0.50/20)] + \exp[-1.0 - 5(0.50/20)]} = 0.60$$

$$P_B = \frac{\exp[-1.0 - 5(0.50/20)]}{\exp[-0.6 - 5(0.50/20)] + \exp[-1.0 - 5(0.50/20)]} = 0.40.$$

In this case, the decision whether to include drive alone in the choice set makes a difference of 0.37 in the probability that drive alone is chosen, 0.22 in the probability that carpool is chosen, and 0.15 in the probability that bus is chosen.

5.6 Summary

This module has been concerned with the problem of selecting variables and choice sets for multinomial logit mode choice models. It can be seen that the analyst has considerable flexibility in making these selections. To a large extent, the analyst must rely on judgment and past experience in deciding which variables to include in a model. However, there also are systematic, empirical procedures for testing models. These procedures help to guide the selection of variables by enabling the analyst to determine whether a particular selection is seriously inconsistent with the available data. Procedures for testing models in this way are discussed in Section 6.5 of Module 6.

TABLE 5.1

VARIABLES THAT HAVE BEEN USED IN MODE CHOICE MODELS

Total travel time	Income of traveler's household
Logarithm of total travel time	Size of traveler's household
In-vehicle travel time	Number of automobiles (automobiles
Logarithm of in-vehicle travel time	per worker, automobiles per
Out-of-vehicle travel time	licensed driver) owned by the
Out-of-vehicle travel time divided by	traveler's household
travel distance	Number of licensed drivers in the
Walk time	traveler's household
Walk distance	Employment density at traveler's
Wait time	workplace
Transit headway	Dummy variable indicating whether
Wait time for transfers	the traveler's workplace is
Number of transfers	in the CBD
Total travel cost	Dummy variable indicating whether
Total travel cost divided by income	traveler is the "primary worker"
of traveler's household	in household or head of household
Auto mileage-related cost divided by	Number of workers in traveler's
income of traveler's household	household
Auto parking cost divided by income	Sex of traveler
of traveler's household	Age of traveler
Auto tolls divided by income of	
traveler's household	
Bus fare divided by income of	
traveler's household	

TABLE 5.2

VARIABLES AND COEFFICIENTS OF WORK-TRIP A MODE CHOICE MODEL
FOR THE SAN FRANCISCO AREA^a

<u>Symbol of Variable</u>	<u>Definition of Variable</u>	<u>Coefficient</u>
$X_{i,DA}$	Dummy variable equal to 1 if i = DA and 0 otherwise	-4.697
$X_{i,CP}$	Dummy variable equal to 1 if i = CP and 0 otherwise	-3.658
C/Y	Travel cost (round trip cents) divided by household income (annual dollars)	-21.43
IVTT	In-vehicle travel time (round trip minutes)	-0.0122
WT	Walk time (round trip minutes)	-0.0335
HD1	Transit headway up to 8 minutes (round trip minutes)	-0.0155
HD2	Transit headway greater than 8 minutes	-0.0107
TW	Transfer wait time (round trip minutes)	-0.0302
CBD1	Dummy variable equal to 1 for drive alone if the traveler's workplace is in the CBD, and equal to 0 for other modes and workplaces	-1.067

<u>Symbol of Variable</u>	<u>Definition of Variable</u>	<u>Coefficient</u>
CBD2	Dummy variable equal to 1 for carpool if the traveler's workplace is in the CBD, and equal to 0 for other modes and workplaces	-0.347
AWDA	Automobiles per worker in the traveler's household for drive alone, and 0 otherwise	1.958
AWCP	Automobiles per worker in the traveler's household for carpool, and 0 otherwise	1.763
AWTA	Automobiles per worker in the traveler's household for transit with auto access, and zero otherwise	1.389
CA	Dummy variable equal to 1 for transit with auto access, and 0 otherwise	-1.237
PW	Dummy variable equal to 1 for drive alone if the traveler is the primary worker in his household, and 0 otherwise	0.677

<u>Symbol of Variable</u>	<u>Definition of Variable</u>	<u>Coefficient</u>
NW	Dummy variable equal to the number of workers in the traveler's household for carpool, and 0 for other modes	0.327
YD	Disposable income (dollars) of the traveler's household for drive alone and carpool, and zero for transit	0.0000137

^a Source: Cambridge Systematics, Inc. (1978), Analytic Procedures for Estimating Changes in Travel Demand and Fuel Consumption, Report DOE/PE/8628-1, Vol. II, U.S. Department of Energy, Washington, D.C.

EXERCISES

5.1 Suppose you want to predict the effects on transit ridership of improving the reliability of transit service. Give three examples of policy variables that could be used to represent transit reliability in a mode choice model.

5.2 Suppose the utility function of a mode choice model is

$$V = -T^2,$$

where T is travel time in hours. According to this utility function, is adding 5 minutes to a 15-minute trip more or less burdensome to travelers than adding 5 minutes to a 1-hour trip? What would your answer be if the utility function were

$$V = -T^{1/2}?$$

(Note: $T^{1/2}$ is the square root of T .)

5.3 Let T , C , and Y , respectively, denote travel time in hours, travel cost in dollars, and after-tax income in thousands of dollars per year. For each of the following utility functions, determine whether high-income travelers are more sensitive, equally sensitive, or less sensitive to changes in travel cost than are low-income travelers:

a. $V = -T - 0.01CY$

b. $V = -0.05TY - 0.025C$

c. $V = -T + 5\ln(Y - C)$

d. $V = -T - 0.01C + 0.2Y$

e. $V = -T - 0.01C - 4C/Y$

5.4 Suppose, as in Example 5.9, that the utility function of a logit model of choice between drive alone, carpool, and bus is

$$V_{DA} = -T_{DA} - 5C_{DA}/Y + 0.1A/W$$

$$V_{CP} = -T_{CP} - 5C_{CP}/Y + 0.05A/W$$

$$V_B = -T_B - 5C_B/Y.$$

Show that if the coefficients b_1 and b_2 have appropriate values, you obtain the same choice probabilities from the logit model whose utility function is

$$V_{DA} = -T_{DA} - 5C_{DA}/Y$$

$$V_{CP} = -T_{CP} - 5C_{CP}/Y + b_1A/W$$

$$V_B = -T_B - 5C_B/Y + b_2A/W.$$

Also, show that by choosing b_3 and b_4 appropriately, you obtain the same choice probabilities from the logit model whose utility function is

$$V_{DA} = -T_{DA} - 5C_{DA}/Y + b_3A/W$$

$$V_{CP} = -T_{CP} - 5C_{CP}/Y$$

$$V_B = -T_B - 5C_B/Y + b_4A/W.$$

What are the appropriate values of the b's?

MODULE 6

ESTIMATING THE UTILITY FUNCTIONS OF CHOICE MODELS

6.1 Introduction

In practice, the deterministic component of a logit model's utility function (called simply the utility function in this module) is never known a priori. In fact, an analyst who is in the initial stages of developing a logit model typically has very little information about the utility function. At best, he is likely to have a list of variables that he thinks should be present in the utility function. But he may not be certain that all of the variables are needed, and he is highly unlikely to know how the variables should be transformed (if at all) or the numerical values of any parameters that enter the utility function. For example, the analyst may believe strongly that a logit model of choice between automobile and bus should include the variables in-vehicle travel time (IVTT), out-of-vehicle travel time (OVTT), and travel cost (C). But he is less likely to have strong a priori beliefs about such matters as:

- a. Whether OVTT should be subdivided into walking time and waiting time.
- b. Whether $\log(\text{IVTT})$ gives a better representation of the effects of in-vehicle travel time than does IVTT.
- c. The values of the numerical coefficients that enter the utility function. For example, suppose it has been determined that the appropriate form of the utility function is

$$V = a_1 \text{IVTT} + a_2 \text{OVTT} + a_3 C.$$

What numerical values should be assigned to the coefficients a_1 , a_2 , and a_3 ?

These questions must be answered empirically by fitting one or more models to appropriate data and then testing and comparing the models to see which one best describes the data. This module describes methods for fitting and testing logit models.

The need to fit a model to data arises in the process of developing virtually any travel demand model, not just logit models. This fitting process is often called calibration. However, the words "estimation" and "testing" give a more precise description of the process that will be discussed in this module since this process consists of estimating the values of the numerical coefficients of models and testing different models to determine which one best explains the available data.

6.2 Acquisition of Data

The first step of any estimation and testing process is acquisition of appropriate data. The data needed to develop a logit mode choice model are:

- a. Observations of actual mode choices by a random sample of individuals who made the types of trips to which the model will apply. For example, if a model of work-trip mode choice is being developed, observations of mode choices by a sample of travelers to work are needed. This sample is called the estimation sample.
- b. The corresponding values of all attributes of both the chosen and the non-chosen modes that may be used as variables of the model. For example, suppose that total travel time is being considered for use as a variable of the model. Then for each individual in the estimation sample, the data must include the observed value of total travel time for each mode available to that individual. A logit model cannot be developed if the attribute data pertain only

to the chosen mode. (In practice, modes that are highly unlikely to be chosen can be considered unavailable and ignored in the data-acquisition step.)

- c. The values of any attributes of individuals that may be used as variables of the model. For example, if automobile ownership is being considered for inclusion in the model, then the data must include the number of automobiles owned by the household of each individual in the estimation sample.

Observations of the choices and attributes of individuals usually are obtained from home-interview or telephone-interview surveys of randomly sampled individuals or households. Data on the attributes of modes (e.g., travel times and costs) can be obtained from analyses of highway and transit networks and transit schedules. Since it is necessary to have data on individual travelers, aggregate data sets, such as the U.S. Census, cannot be used as sources of travel data unless special procedures are implemented to remove the biases caused by aggregation of the data. These procedures are quite complex, and discussion of them is beyond the scope of this course.

The size of the sample needed to develop a logit mode choice model usually is in the range 1000-3000 individuals, not counting non-respondents and unusable responses. Although the upper end of this range is preferable to the lower, a mode choice model usually can be developed satisfactorily from a sample of 1000 observations if cost or other considerations prohibit acquisition of a larger data set.

The data used for developing a logit model must consist of observations of choices by individuals, and the attribute data must pertain to these individuals. Use of aggregate data, such as mode shares according to

traffic zone and average values of attributes according to traffic zone, can result in the development of a highly erroneous model unless special procedures are used to remove the effects of aggregation bias.

Table 6.1 illustrates the data needed to develop a simple model of choice between the modes drive alone, carpool, and bus. The attributes of modes and travelers that will be used to form the variables of this model are total travel time and the number of automobiles owned by the traveler's household. The entry NA for the travel time of a mode signifies that this mode either is not available to the traveler or is so unlikely to be chosen that it can be treated as unavailable without making a serious error. Of course, in practice, many more attributes would be included in the data set than are shown in Table 6.1, and the data table would be correspondingly larger. However, its form and structure would be as shown in Table 6.1.

TABLE 6.1 -- DATA FOR DEVELOPMENT OF A SIMPLE MODE CHOICE MODEL

Person	Autos Owned	Chosen Mode	Travel Time (Min.) By		
			Drive Alone	Carpool	Bus
1	1	Drive Alone	20	30	45
2	0	Bus	NA	25	35
3	2	Drive Alone	15	22	60
4	1	Carpool	30	35	55
5	1	Carpool	10	12	25
6	1	Bus	20	30	15
7	0	Carpool	NA	20	15
8	3	Drive Alone	30	40	75
9	2	Carpool	10	12	8

TABLE 6.1 (cont)

Person	Autos Owned	Chosen Mode	Travel Time (Min.) By		
			Drive Alone	Carpool	Bus
10	1	Bus	50	60	40
.
.
.

6.3 Specifying the Model

After the data have been acquired, the next step in developing a logit model consists of specifying, tentatively, one or more forms of the utility function. This specification step usually includes identifying the variables of the utility function, including any transformations of the attribute data, and the functional form of the relation between the variables and the deterministic component of utility. It usually does not include specifying the numerical values of the constant coefficients that enter the utility function. For example, in developing a logit mode choice model from the data shown in Table 6.1, the following two alternative forms of the utility function might be specified:

$$\text{Form 1: } V_{DA} = a_1 T_{DA} + a_2 A + a_3 \quad (6.1a)$$

$$V_{CP} = a_1 T_{CP} + a_4 A + a_5 \quad (6.1b)$$

$$V_B = a_1 T_B. \quad (6.1c)$$

$$\text{Form 2: } V_{DA} = b_1 \log(T_{DA}) + b_2 A + b_3 \quad (6.2a)$$

$$V_{CP} = b_1 \log(T_{CP}) + b_4 A + b_5 \quad (6.2b)$$

$$V_B = b_1 \log(T_B). \quad (6.2c)$$

In these equations, T denotes travel time in minutes, A denotes automobiles owned by the traveler's household, and a_1 - a_5 and b_1 - b_5 are constant coefficients. The specification of the forms (6.1) and (6.2) at this stage does not imply that the analyst necessarily believes either to be correct. Rather, (6.1) and (6.2) are forms that the analyst believes are worthy of estimation and testing. During the process of estimation and testing, information will be obtained that helps to determine whether these forms should be modified (e.g., by deleting one or more variable from either or both) and that provides a way to determine which form best explains the observed choices of the individuals in the estimation sample.

6.4 Estimation -- The Maximum Likelihood Method

The third step of model development consists of estimating the numerical values of the models' coefficients (e.g., the values of a_1 to a_5 and b_1 to b_5 in equations (6.1) and (6.2)) by fitting the models to the available data. The fitting technique that is usually used in practice is called the maximum likelihood method. This consists of choosing the values of the coefficients so as to maximize the likelihood (or probability) according to the model being developed of observing the choices made by the individuals in the estimation sample. It can be shown that the maximum likelihood method yields estimates of the coefficients and predictions of choice probabilities that have the greatest possible accuracy.

The values of the coefficients obtained by the maximum likelihood method are called estimates because they are based on a data set that is a random sample of all travelers who might make the choices being modeled. Re-estimation of the coefficients using a different random sample from the same population of travelers would give different values of the coefficients

owing to the effects of random sampling error. The "true" coefficient values could be found only if estimation were carried out using all individuals in the population of interest. This is never possible in practice. Therefore, the values of the coefficients can be known in practice only up to the effects of random sampling error. This is why the numerical values of the coefficients are called estimates.

The following example illustrates the maximum likelihood method.

Example 6.1: The Maximum Likelihood Method

Suppose a logit model of choice between the modes automobile and bus is being developed. The only variable of the model is total travel time, T. The deterministic component of the model's utility function is specified as

$$V = aT, \tag{6.3}$$

where a is a constant coefficient. Suppose that the estimation sample consists of observations of the mode choices of three individuals. Of course, a sample size of three is far too small to be useful in practice, but it is convenient for illustrative purposes because it enables all the necessary computations to be performed with a desk calculator. Maximum likelihood estimation with a sample of realistic size requires the use of a digital computer. Let the choices and travel times for the individuals in the estimation sample be:

<u>Person</u>	<u>Chosen Mode</u>	<u>Travel Time (Min.)</u>	
		<u>Automobile</u>	<u>Bus</u>
1	Auto	50	30
2	Auto	10	20
3	Bus	30	40

According to the model, the probabilities of the observed mode choices are

$$\text{Individual 1: } P(\text{Auto}) = \frac{\exp(50a)}{\exp(50a) + \exp(30a)}$$

$$\text{Individual 2: } P(\text{Auto}) = \frac{\exp(10a)}{\exp(10a) + \exp(20a)}$$

$$\text{Individual 3: } P(\text{bus}) = \frac{\exp(40a)}{\exp(30a) + \exp(40a)}.$$

Equivalently, the probabilities are

$$\text{Individual 1: } P(\text{Auto}) = \frac{1}{1 + \exp(-20a)}$$

$$\text{Individual 2: } P(\text{Auto}) = \frac{1}{1 + \exp(10a)}$$

$$\text{Individual 3: } P(\text{Bus}) = \frac{1}{1 + \exp(-10a)}.$$

The probability of the entire estimation sample is

$$L = P(\text{person 1 chooses auto}) \times P(\text{person 2 chooses auto}) \\ \times P(\text{person 3 chooses bus}).$$

Therefore,

$$L = \frac{1}{1 + \exp(-20a)} \times \frac{1}{1 + \exp(10a)} \times \frac{1}{1 + \exp(-10a)} \quad (6.4)$$

L is called the sample likelihood. In practice it is customary to work with

the natural logarithm of L, which is called the log likelihood and is

denoted by $\log L$. In this example

$$\log L = -\{\log[1 + \exp(-20a)] + \log[1 + \exp(10a)] \\ + \log[1 + \exp(-10a)]\}. \quad (6.5)$$

The likelihood of a logit model is always a number whose value is between zero and one (because the likelihood is a probability). Therefore, the log likelihood is always a negative number.

The maximum likelihood method chooses the value of a so as to maximize L or, equivalently, $\log L$. Although this usually requires the use of a digital computer, it can be done graphically in this case. Figure 6.1 shows the graph of $\log L$ in Equation (6.5) as a function of a . It can be seen that the maximum occurs at $a = 0.08$. This value is called the maximum likelihood estimate of a .

The only difference between the use of the maximum likelihood method in the foregoing example and the use of the method in actual practice is that in practice, the model being estimated has many unknown coefficients (e.g., the work trip mode choice model for San Francisco discussed in Section 5.4 of Module 5 has 17 coefficients) and the estimation data set contains many more than three observations. As a result, the values of the coefficients that maximize the likelihood and log likelihood functions cannot be found graphically; they must be found using a digital computer. Software for carrying out maximum likelihood estimation of logit models is available for both mainframe computers (e.g., the subroutine ULOGIT in the UTMS system) and microcomputers (e.g., the MDA software package).

6.5 Interpreting the Estimation Results -- Testing the Model

The outputs of logit estimation software include, in addition to the estimated values of the model's coefficients, a variety of information that is useful for interpreting the estimated coefficients, deciding which variables should be included in the model, and comparing one model with another to determine which best describes the data.

6.5a The Precision of the Estimates -- Standard Errors of the Estimates

The outputs of most logit estimation software include, along with the estimated values of the coefficients, a set of numbers called the standard errors of the estimates. The standard error of the estimate of the value of a particular coefficient is an indicator of the amount by which the estimated value of the coefficient is likely to differ from the true value as a result of random sampling error. Thus, the standard error of the

estimate is an indicator of the precision with which the coefficient has been estimated. If the model is correctly specified, then there is a probability of 0.95 that the true coefficient value is within 1.96x(standard error of the estimate) of the estimated value. In other words, if b_{est} is the estimated value of a coefficient, b_{true} is its unknown true value, and s is its standard error of estimate, the following inequality holds with probability 0.95:

$$b_{est} - 1.96s \leq b_{true} \leq b_{est} + 1.96s. \quad (6.6)$$

Changing the number 1.96 to 1.645 or 2.575 causes the inequality to hold with probability 0.90 or 0.99. The interval $b_{est} - 1.96s$ to $b_{est} + 1.96s$ is called a 95 percent confidence interval for b_{true} . The analogous intervals formed by replacing 1.96 by 1.645 and 2.575 are called 90 percent and 99 percent confidence intervals.

Example 6.2 -- Confidence Intervals

Suppose the estimated value of a particular coefficient is 2.337 and its standard error of estimate is 0.875. Then a 95 percent confidence interval for the true value of the coefficient is 0.622 to 4.052. In other words, the interval 0.622 to 4.052 contains the true value of the coefficient with probability 0.95.

6.5b Deciding Whether to Retain a Variable -- The t Statistic

In addition to the standard errors of the coefficient estimates, most logit software reports numbers called t statistics of the coefficients. The t statistic of a coefficient is simply the estimated value of the coefficient divided by the standard error of the estimate. In other words, $t = b_{est}/s$.

The t statistic of a coefficient is useful for deciding whether the variable associated with that coefficient contributes significantly to the ability of the model to describe or explain the data. This makes the t statistic useful for deciding whether a variable should be retained in the model or dropped. Variables with significant explanatory power should be retained, whereas variables with little explanatory power should be dropped. In general, variables whose coefficients have large positive or negative t statistics have greater explanatory power than variables whose coefficients have t statistics that are close to zero. Thus, variables whose coefficients have large positive or negative t statistics should be retained, whereas variables whose coefficients have t statistics close to zero may be dropped from a model. There is no single "correct" dividing line between t values that indicate a variable should be retained and t values that indicate a variable can be dropped. However, experience suggests that it is a good policy to retain variables whose coefficients have t statistics that are greater than 1.0 or less than -1.0. Failure to retain variables whose coefficients have t statistics outside of the range -1.0 to 1.0 may cause the estimates of the remaining coefficients to be seriously biased and the resulting model to be highly erroneous.

Variables whose coefficients have t statistics between -1.0 and 1.0 can be considered candidates for being dropped from the model. Before such variables are dropped, however, it is necessary to take account of certain qualifications that will be discussed later in this subsection and in subsection 6.5c. In addition, the analyst may have strong reasons for believing a variable has an important influence on choice, even if its t statistic is close to 0. In such a case, the analyst may decide to retain the variable, despite its low t statistic. However, the coefficient of the

variable and, therefore, its influence on choice will be known very imprecisely. It will not be possible to conclude with confidence that predictions obtained from a model that includes the variable are more accurate than predictions obtained from a model that excludes it.

Example 6.3 -- t Statistics

Suppose that in a model of choice between automobile (A) and bus (B) for travel to work, the utility function is specified as

$$V_A = b_1 + b_2IVTT_A + b_3OVTT_A + b_4C_A + b_5A + b_6D \quad (6.7a)$$

$$V_B = b_2IVTT_B + b_3OVTT_B + b_4C_B, \quad (6.7b)$$

where IVTT denotes in-vehicle travel time, OVTT denotes out-of-vehicle travel time, C denotes travel cost, A denotes the number of automobiles owned by the traveler's household, and D equals 1 if the traveler's workplace is in the central business district and 0 otherwise. Let the estimation results be:

<u>Coefficient</u>	<u>Variable</u>	<u>Estimated Value</u>	<u>Standard Error of Estimate</u>	<u>t Statistic</u>
b_1	Intercept	1.45	0.390	3.72
b_2	IVTT	-0.00897	-0.00632	-1.42
b_3	OVTT	-0.0308	-0.0106	-2.91
b_4	C	-0.115	-0.0262	-4.39
b_5	A	0.770	0.244	3.16
b_6	D	-0.561	0.783	-0.716

The t statistic of b_6 is between -1.0 and 1.0, which suggests that the variable D has little explanatory power and that this variable can be

dropped from the model. None of the other coefficients have t statistics between -1.0 and 1.0, so none of the other variables can be dropped.

The fact that a coefficient has a low t statistic does not automatically mean that the corresponding variable should be dropped from the model. Errors in specifying the model's utility function can cause one or more coefficients to have low t statistics even if the attributes their variables represent are important to mode choice. For example, if the correct way to represent a certain attribute is with the variable X^2 but in the estimated model the attribute is represented incorrectly by X, then the coefficient of X may have a low t statistic, even if the attribute represented by X is important to mode choice. In this case, if the model were re-estimated with the variable X^2 in place of X, the coefficient of X^2 might be found to have a very high t statistic. Thus, it is often useful to experiment with different transformations of an attribute before concluding on the basis of t statistics that the attribute need not be represented in the utility function.

Another situation in which a low t statistic may not indicate that a variable should be dropped is when two or more coefficients have low t statistics. It is possible for the t statistics of several coefficients to be low even though the variables associated with these coefficients collectively have significant explanatory power. In other words, it is possible for individual variables to have low explanatory power while a group of such variables has high explanatory power. In this case, it would be undesirable to drop any of the variables, despite the low t statistics of their coefficients. A method for identifying the occurrence of this situation is presented in the next subsection.

There is one situation in which a t statistic outside of the range -1.0 to 1.0 indicates that a model is erroneously specified. If a coefficient has a t statistic outside this range and its sign is inconsistent with well-established theory, then the model involved is almost certainly incorrect. For example, the coefficient of travel cost in a mode choice model should be negative. Thus, for example, if estimating a particular model yields a travel cost coefficient of +0.50 with a t statistic of 2.7, the model is almost certainly incorrect and should be reformulated.

6.5c Deciding Whether to Retain a Group of Variables -- The Likelihood Ratio Test

Most logit estimation software reports the value that the sample log likelihood has when the values of the coefficients equal the maximum likelihood estimates. This maximum value of the log likelihood provides the basis of a procedure for deciding whether a group of variables can be dropped from a model. The procedure is called a likelihood ratio test. Intuitively, it works as follows. If the group of variables in question has little explanatory power, then dropping them from the model should have little effect on the maximum value of the log likelihood. Dropping one or more variables always will cause the maximum value of the log likelihood to decrease, but it will not decrease by much if the variables that have been dropped have little explanatory power. In other words, if the group of variables in question has little explanatory power, the difference between the log likelihoods of models estimated with and without these variables will be close to zero.

The likelihood ratio test is carried out quantitatively as follows:

- a. Estimate the model with all variables included. Let $\log L_1$ denote the resulting maximum value of the log likelihood.
- b. Drop the variables in question and re-estimate the model. Let $\log L_2$ denote the resulting maximum value of the log likelihood.
- c. Compute the quantity $LR = 2(\log L_1 - \log L_2)$. LR is called the likelihood ratio test statistic. It usually must be computed by hand using the value of $\log L_1$ and $\log L_2$ reported by the logit estimation software. LR is always a positive number.
- d. If LR exceeds an appropriately determined critical value, CV, then the variables being tested should be retained in the model, even if all of their coefficients have t statistics in the range -1.0 to 1.0. If LR is less than C, then it may be desirable to drop the variables from the model.

The critical value, CV, for the likelihood ratio test statistic depends on the number of variables being tested. Table 6.2 shows the appropriate values of CV for testing 2 to 5 variables. A likelihood ratio test of one variable is equivalent to the t-test procedure described in subsection 6.5b. Thus, there is no need to carry out a likelihood ratio test of a single variable.

The group of variables to which the likelihood ratio test is applied should be selected before looking at the variables' t statistics. It is an incorrect use of the test to define the group as the set of variables whose t statistics are between -1.0 and 1.0.

TABLE 6.2 -- CRITICAL VALUES OF THE LIKELIHOOD RATIO TEST STATISTIC

<u>Number of Variables Being Tested</u>	<u>Critical Value</u>
2	2.408
3	3.665
4	4.878
5	6.064

The use of the likelihood ratio test is illustrated by the following example.

Example 6.4: The Likelihood Ratio Test

Suppose that estimation of the logit model of Example 6.3 (i.e., the model whose utility function is specified in Equations 6.7) had yielded the following results:

<u>Coefficient</u>	<u>Variable</u>	<u>Estimated Value</u>	<u>Standard Error of Estimate</u>	<u>t Statistic</u>
b ₁	Intercept	1.45	0.390	3.72
b ₂	IVTT	-0.00897	-0.0163	-0.549
b ₃	OVTI	-0.0308	-0.0106	-2.91
b ₄	C	-0.115	-0.0262	-4.39
b ₅	A	0.770	0.244	3.16
b ₆	D	-0.561	0.783	-0.716

log L = -374.4

Suppose, also, that it is uncertain whether the variables IVTT and D contribute significantly to the explanatory power of the model. To determine whether these variables can be dropped, re-estimate the model using the utility function

$$V_A = b_1 + b_3 \text{OVTT}_A + b_4 C_A + b_5 A \quad (6.8a)$$

$$V_B = b_3 \text{OVTT}_B + b_4 C_B \quad (6.8b)$$

Suppose the results are as follows:

Coefficient	Variable	Estimated Value	Standard Error of Estimate	t Statistic
b_1	Intercept	2.67	0.438	6.10
b_3	OVTT	-0.0291	-0.0143	-2.04
b_4	C	-0.175	-0.0482	-3.63
b_5	A	0.567	0.163	3.48

$$\log L = -377.2$$

Then the likelihood ratio test statistic is $LR = 2[(-374.4) - (-377.2)] = 5.60$. There are two variables being tested. According to Table 6.2, the critical value of the likelihood ratio statistic for testing two variables is 2.408. Since LR exceeds this value, the variables IVTT and DX_a together have significant explanatory power, even though neither has a t statistic outside of the range -1.0 to 1.0. Although the influence of each variable on choice can be estimated only very imprecisely, neither variable should be dropped from the model. Doing so may seriously bias the remaining coefficients and lead to large prediction errors. In other words, it is not possible with this model to make accurate predictions of the effects on mode choice of changes in in-vehicle travel time or workplace location. However,

it is necessary to retain these variables in the model to prevent predictions of the effects of changes in the other variables from being biased.

6.5d Other Uses of t Statistics and Likelihood Ratio Tests

Other important uses of t statistics and likelihood ratio tests are to determine whether attributes such as travel time should be decomposed into components and to determine whether such attributes are generic or mode specific. The following examples illustrate how these determinations can be made.

Example 6.5: Determining Whether Out-of-Vehicle Travel Time Should Be Subdivided into the Components Walk Time and Wait Time

Consider, again, the logit mode choice model of Example 6.3 whose utility function is given by Equations (6.7). If walk time and wait time are evaluated differently by travelers, then the variable OVTT should be replaced by its components WK (walk time) and WT (wait time). The term b_3OVTT in the utility function should be replaced by $b_3WK + aWT$, where b_3 and a are constant coefficients. But

$$b_3WK + aWT = b_3(WK + WT) + (a - b_3)WT. \quad (6.9)$$

Since OVTT is the sum of WK and WT,

$$b_3WK + aWT = b_3OVTT + (a - b_3)WT. \quad (6.10)$$

Let $b_7 = (a - b_3)$. Then

$$b_3WK + aWT = b_3OVTT + b_7WT. \quad (6.11)$$

Therefore, the utility function for the model in which out-of-vehicle travel time is decomposed into the components walk time and wait time can be written in the form

$$V_A = b_1 + b_2 IVTT_A + b_3 OVTT_A + b_4 C_A + b_5 A + b_6 D + b_7 WT_A. \quad (6.12a)$$

$$V_B = b_2 IVTT_B + b_3 OVTT_B + b_4 C_B + b_7 WT_B. \quad (6.12b)$$

To determine whether the decomposition is worthwhile, estimate the logit model using the utility function (6.12). If the t statistic of the coefficient b_7 is outside of the range -1.0 to 1.0, then the separate variable WT should not be dropped from the model. The decomposition of out-of-vehicle travel time into components adds significant explanatory power to the model. If the t statistic of b_7 is between -1.0 and 1.0, then WT does not have significant explanatory power and probably can be dropped. In other words, if the t statistic is between -1.0 and 1.0, decomposition of out-of-vehicle travel time into its components is unnecessary.

Example 6.6: Generic or Mode-Specific Travel Time and Cost Variables

Refer, again, to the mode choice model of Example 6.3. Suppose it is desired to determine whether in-vehicle travel time and cost should be represented as mode-specific variables. If in-vehicle travel time is mode-specific, then Equations (6.7) should be replaced by

$$V_A = b_1 + b_2 IVTT_A + b_3 OVTT_A + b_4 C_A + b_5 A + b_6 D \quad (6.13a)$$

$$V_B = b_7 IVTT_B + b_3 OVTT_B + b_4 C_B, \quad (6.13b)$$

where b_7 is a constant coefficient. This representation causes the in-vehicle travel time term of the utility function, $b_2 IVTT$ for auto and $b_7 IVTT$ for bus, to be mode specific. If, in addition, travel cost is mode specific, the utility function becomes

$$V_A = b_1 + b_2 IVTT_A + b_3 OVTT_A + b_4 C_A + b_5 A + b_6 D \quad (6.14a)$$

$$V_B = b_7 IVTT_B + b_3 OVTT_B + b_8 C_B, \quad (6.14b)$$

where b_8 is a constant coefficient. Equations (6.14) are equivalent to

$$V_A = b_1 + b_2 IVTT_A + b_3 OVTT_A + b_4 C_A + b_5 A + b_6 D \quad (6.15a)$$

$$V_B = b_2 IVTT_B + b_3 OVTT_B + b_4 C_B + (b_7 - b_2) IVTT_B + (b_8 - b_4) C_B. \quad (6.15b)$$

Define new coefficients b_9 and b_{10} by

$$b_9 = b_7 - b_2 \quad (6.16a)$$

$$b_{10} = b_8 - b_4. \quad (6.16b)$$

Then the utility function with mode specific in-vehicle travel time and travel cost variables can be written in the form:

$$V_A = b_1 + b_2 IVTT_A + b_3 OVTT_A + b_4 C_A + b_5 A + b_6 D \quad (6.17a)$$

$$V_B = b_2 IVTT_B + b_3 OVTT_B + b_4 C_B + b_9 IVTT_B + b_{10} C_B. \quad (6.17b)$$

To determine whether the mode-specific representation of in-vehicle travel time and travel cost is worthwhile, estimate the logit model whose utility function is given by Equations (6.17). Then use a likelihood ratio test to determine whether the terms $b_9 IVTT_B$ and $b_{10} C_B$ add significant explanatory power to the model. If they do not, then the mode-specific representation is unnecessary. If they do, then at least one of the attributes in-vehicle travel time and cost should be represented as a mode specific variable in the model. If only one of the coefficients b_9 and b_{10} has a t-statistic between -1.0 and 1.0, then the corresponding attribute can be represented by a generic variable. If neither or both coefficients have t-statistics in this range, then both attributes should be represented by mode-specific variables. (If the likelihood ratio test indicates that the mode-specific representations of travel time and cost add significant explanatory power to the model, then at least one of the variables travel time and travel cost must be treated as mode specific, regardless of the values of the t statistics of b_9 and b_{10} . If, in addition, b_9 and b_{10} both have t statistics between -1.0 and 1.0, it is not possible to determine

whether one of the variables -- either travel time or travel cost -- can be treated as generic.)

6.5e Comparisons of Non-Nested Models -- The Modified Likelihood Ratio

Test

All of the tests of models that have been discussed so far have been formulated as tests of whether a particular variable or group of variables should be dropped from a model. Not all tests can be formulated this way. For example, suppose that two logit models of mode choice are under consideration. It is desired to determine which model best explains the available data. Let the deterministic components of the utility functions of these models be

$$\text{Model 1: } V = a_1T + a_2C \quad (6.18)$$

$$\text{Model 2: } V = b_1 \log T + b_2C, \quad (6.19)$$

where T and C , respectively, denote travel time and travel cost, and the a 's and b 's are constant coefficients. The t and likelihood ratio test procedures discussed in the preceding subsections cannot be used to determine which model is best because neither model can be obtained by adding variables to or dropping variables from the other. Models that cannot be obtained from one another by addition or deletion of variables are said to be non-nested.

Intuitively, one might expect that if one of two non-nested models explains the available data better than the other model does, then the better model should yield a larger maximum value of the sample log likelihood. Thus, one might expect that a test similar to a likelihood ratio test can be developed for testing non-nested models against one another. This expectation is correct.

The modified likelihood ratio test procedure is as follows. Let the non-nested models be called models 1 and 2. Let $\log L_1$ and $\log L_2$, respectively, denote the maximum values of the sample log likelihood for models 1 and 2, and let K_1 and K_2 denote the numbers of estimated coefficients in the two models. (For example, in the models represented by equations (6.18) and (6.19), $K_1 = K_2 = 2$.) Assume that $\log L_1 \geq \log L_2$, which suggests that model 1 is preferred to model 2. (If $\log L_1 < \log L_2$, then renumber the models to make $\log L_1 \geq \log L_2$.) Define the modified likelihood ratio test statistic, MLR, by

$$\text{MLR} = (\log L_1 - K_1/2) - (\log L_2 - K_2/2). \quad (6.20)$$

If $\text{MLR} > 1.35$, then model 1 explains the available data substantially better than does model 2. Moreover, model 2 almost certainly is misspecified and should be dropped from further consideration.

The following example illustrates the comparison of two non-nested models using the modified likelihood ratio test.

Example 6.7 -- Comparison of Non-Nested Models

Consider the logit mode choice models whose utility functions are as follows:

$$\text{Model 1: } V = a_1 \log \text{IVTT} + a_2 \log \text{OVTT} + a_3 C \quad (6.21)$$

$$\text{Model 2: } V = b_1 T + b_2 C, \quad (6.22)$$

where T , IVTT , OVTT , and C are total travel time, in-vehicle travel time, out-of-vehicle travel time, and travel cost, respectively, and the a 's and b 's are constant coefficients. Suppose that maximum likelihood estimation of the two models yields the results $\log L_1 = -437.7$ and $\log L_2 = -440.2$. There are 3 estimated coefficients in model 1 and 2 in model 2. Therefore,

$K_1 = 3$ and $K_2 = 2$. The value of the modified likelihood ratio test statistic is

$$MLR = (-437.7 - 3/2) - (-440.2 - 2/2) = 2.00.$$

Since MLR exceeds 1.35, model 1 explains the data better than model 2 does, and model 2 is almost certainly incorrect.

6.6 Some Estimation Problems and How to Avoid Them

There are certain kinds of specification errors that make maximum likelihood estimation of logit models impossible. If one or more of these specification errors occurs, estimation software will terminate abnormally and, possibly, produce an error or warning message. Any estimates that are obtained under these conditions will be meaningless. The most frequently arising specification errors that make estimation impossible will now be described.

a. Use of too many alternative-specific constants: In most practical situations, the utility functions of logit mode choice models include alternative-specific constants. As was discussed in Section 4.6 of Module 4, the number of such constants that are included in a model must not exceed the number of modes in the model minus one. If the number of alternative-specific constants equals the number of modes, there will not be a unique set of coefficient values that maximizes the sample log likelihood. This usually will cause estimation software to terminate abnormally or to produce a message indicating that an estimation problem has occurred.

b. Incorrect specification of socioeconomic variables: Socioeconomic variables, such as income and automobile ownership, have the same values for all alternatives. As was discussed in Section 5.3 of Module 5, these variables can enter a logit model only if they are mode-specific or are

multiplied or divided by an attribute variable whose values differ across alternatives. Socioeconomic variables have no effect on choice probabilities if they enter a logit model's utility function as generic variables that do not interact with variables whose values vary across alternatives. As a result, when generic socioeconomic variables that do not interact with other variables are present in a logit model, there is not a unique set of coefficient values that maximizes the sample log likelihood, and abnormal termination of estimation software occurs.

As was also discussed in Section 5.3, the number of mode-specific socioeconomic variables must not exceed the total number of modes in the model minus one. Violation of this rule will cause coefficient estimation to fail and logit estimation software to terminate abnormally or produce an error message.

c. Perfect Collinearity of Variables: Perfect collinearity is a condition in which one or more variables of the utility function are exact linear combinations of other variables. For example, suppose that T , $IVTT$, and $OVTT$ denote total travel time, in-vehicle travel time, and out-of-vehicle travel time, respectively. Let the utility function of a logit mode choice model be specified as

$$V = b_1 T + b_2 IVTT + b_3 OVTT + \text{other terms.} \quad (6.23)$$

Then perfect collinearity exists because T is an exact linear combination of $IVTT$ and $OVTT$. Specifically, $T = IVTT + OVTT$. The problem that perfect collinearity poses for estimation can be seen by rewriting (6.23) in the form

$$V = b_1 (IVTT + OVTT) + b_2 IVTT + b_3 OVTT + \text{other terms} \quad (6.24)$$

$$= (b_1 + b_2) OVTT + (b_1 + b_3) IVTT + \text{other terms.} \quad (6.25)$$

Equation (6.26) shows that predictions of choice depend only on the values of $b_1 + b_2$ and $b_1 + b_3$. But there are infinitely many combinations of b_1 , b_2 , and b_3 that yield the same values of $b_1 + b_2$ and $b_1 + b_3$. As a result, it is not possible to find a unique set of b values that maximizes the model's log likelihood, and attempts to estimate the coefficients will result in abnormal termination of the software.

Perfect collinearity always causes there to be infinitely many combinations of coefficient values that maximize the log likelihood and, for this reason, logit estimation software terminates abnormally or produces an error message when it occurs.

d. Models with one or more unbounded coefficients: It is possible to create models in which some of the variables, when multiplied by an infinite coefficient, perfectly explain the choices of a subset of the travelers in the estimation data set without affecting the estimated choice probabilities for the other travelers. For example, suppose a certain variable always equals one for observed transit users and always equals zero for other travelers. (Such a variable might represent a special preference for transit that only transit users are thought to have.) Then, if the utility function coefficient of this variable is infinity, the predicted probability of choosing transit will be 1 for all observed transit users, but the variable will have no effect on the choice probabilities of travelers not observed to choose transit.

When this condition occurs, estimation software will not be able to find a set of coefficient values that maximizes the log likelihood (i.e., because a computer cannot represent an infinite constant). The estimation software will terminate abnormally, possibly after reporting a series of numerical overflows and usually with an indication that the coefficient

values that maximize the log likelihood have not been found. Restarting the estimation procedure will consume computer time but will not produce successful maximum likelihood estimates of the coefficients.

6.7 Conclusions

This module has explained the method normally used to estimate the coefficients of logit mode choice models. It has also described statistical procedures that can guide the selection of variables for and testing of logit models. The statistical procedures are invaluable aids in model development, but it is important to understand that the use of statistical methods alone cannot guarantee the development of a satisfactory model. As was discussed in Section 5.1 of Module 5, model development is as much an art as a science, and judgment and experience are important elements of the art.

The need for judgment and experience, even when "objective" statistical methods are available, arises mainly from the fact that statistical tests cannot determine when a model is correct (or, at least, sufficiently free of serious errors to be satisfactory for its intended uses). They can only determine that a model is wrong. If a model is found to be wrong, statistical methods rarely provide useful insight into why it is wrong or how to correct it. The analyst must use judgment and experience to identify likely sources of error in the model and to formulate modifications of the model that might remove the errors. The modified models then can be subjected to statistical tests to determine whether they are seriously erroneous. Thus, practical model development always involves alternating between statistical analysis and judgmental activities.

EXERCISES

6.1 Suppose you want to develop a binomial logit model of choice between automobile and bus for the work trip. The utility function of the model is

$$V = aT,$$

where T is total travel time in minutes, and a is a constant coefficient. Suppose that the estimation sample consists of the following three observations:

<u>Person</u>	<u>Chosen Mode</u>	<u>Travel Time (Min.)</u>	
		<u>Automobile</u>	<u>Bus</u>
1	Auto	20	25
2	Auto	25	40
3	Bus	30	40

(Of course a real model would have a much more complicated utility function and would be estimated using a much larger data set. However, this simple model and small data set are exactly right for this exercise.)

- a. What is the log likelihood of this sample according to the specified model? (Hint: Refer to Example 6.1.)
- b. Evaluate the log likelihood for $a = -0.025$, -0.045 , and -0.065 . Which value of a yields the largest value of the log likelihood?
- c. Can you find a value of a that yields a larger value of the log likelihood? (Hint: Evaluate the log likelihood using two new

values of a , one that is slightly larger and one that is slightly smaller than the value that yielded the largest log likelihood in part b.)

6.2 In a logit model of choice between automobile and bus for travel to work, the utility function has the form

$$V = a_1 IVTT + a_2 OVTT + a_3 LC + a_4 PC,$$

where IVTT denotes in-vehicle travel time, OVTT denotes walk time, LC denotes linehaul cost (e.g., cost of automobile fuel and maintenance, bus fare), PC denotes parking cost, and the a 's are constant coefficients. The values of the coefficients were estimated by maximum likelihood, and the following results were obtained:

<u>Coefficient</u>	<u>Variable</u>	<u>Estimated Value</u>	<u>Standard Error of Estimate</u>
a_1	IVTT	-1.72	0.79
a_2	OVTT	-2.36	0.52
a_3	LC	-0.79	0.55
a_4	PC	-0.84	0.41

Log likelihood: -179.37

- a. Do the estimation results suggest that there are variables of the model that do not significantly affect mode choice and, therefore, are candidates for being dropped? If so, which one or ones?
- b. Since the estimated values of a_3 and a_4 are close, you would like to determine whether LC and PC can be combined into the single variable, total travel cost. What model should be estimated in order to make this determination? Write the utility function of

the model. What estimation results will you use to decide whether to retain the decomposition travel cost into separate components?

c. A second analyst has suggested adding two additional variables to the model: CBD, which is a variable equal to 1 for the automobile mode if the traveler works in the CBD; and TR, the number of transfers required if transit is used. If the second analyst's suggestion is accepted, the utility function will be

$$V_a = a_1 IVTT_a + a_2 OVTT_a + a_3 LC_a + a_4 PC + a_5 CBD$$

$$V_b = a_1 IVTT_b + a_2 OVTT_b + a_3 LC_b + a_6 TR,$$

where the subscripts a and b denote automobile and bus, respectively, and it is assumed that there is no parking cost if bus is chosen. Estimation of this model yielded the results:

<u>Coefficient</u>	<u>Variable</u>	<u>Estimated Value</u>	<u>Standard Error of Estimate</u>
a ₁	IVTT	-1.65	0.83
a ₂	OVTT	-2.53	0.61
a ₃	LC	-0.73	0.52
a ₄	PC	-0.72	0.59
a ₅	CBD	-0.23	0.47
a ₆	TR	-0.09	0.39

Log likelihood: -178.22

Based on these results, do you believe the variables suggested by the second analyst should be retained in the model?

- d. A third analyst thinks that travelers may find out-of-vehicle travel time less onerous for long trips than for short ones. This analyst has proposed specifying the utility function as

$$V = a_1 IVTT + a_2 OVTT/DIST + a_3 LC + a_4 PC,$$

where DIST denotes travel distance. Estimation of this model yielded the results:

<u>Coefficient</u>	<u>Variable</u>	<u>Estimated Value</u>	<u>Standard Error of Estimate</u>
a_1	IVTT	-1.43	0.69
a_2	OVTT/DIST	-0.27	0.10
a_3	LC	-0.64	0.43
a_4	PC	-0.89	0.38

Log likelihood: -177.53

Based on the estimation results, does the model with OVTT/DIST explain the data significantly better than does the model with OVTT?

MODULE 7

PREDICTION WITH DISAGGREGATE MODELS

7.1 Introduction

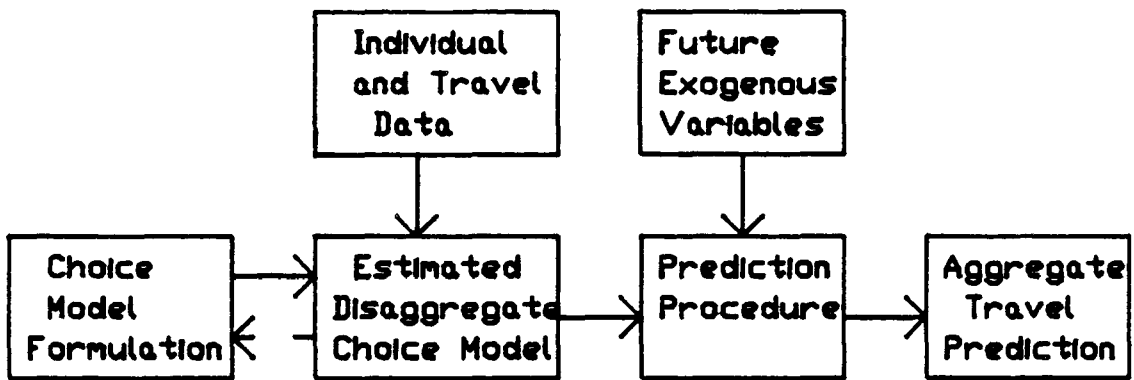
One of the main objectives of transportation analysis is to support the evaluation of transportation plans and policies and, thereby, to aid the transportation decision-making process. The evaluation is based, in part, on the predicted effects of alternative capital investment and operating decisions on travel flows, levels of service, and external or non-user impacts. The decision process requires information about aggregate travel volumes because these are important measures of system performance and they affect travel service and external impacts. Thus, aggregate travel volumes are important outputs of travel demand prediction models.

The preceding modules have described the formulation and estimation of models of individual travel behavior. The decision to use models of individual behavior, or disaggregate models, is based on theoretical and empirical evidence that such models reduce data collection costs and are necessary to properly capture the effects of changes in population characteristics and transportation service attributes on travel behavior. To use these models for making predictions of aggregate travel volumes, a method is needed to aggregate the model's predictions of the behavior of individuals. This module describes methods for obtaining aggregate forecasts from disaggregate models.

Figure 7.1 summarizes the principal steps involved in developing a disaggregate model and using it to make predictions of aggregate travel. The main flow of activities is shown on the lower line. The first block on

Figure 7.1

Development and Application of
Dissaggregate Mode Choice Models



the left represents model formulation, which includes identification of the set of alternatives and selection of variables. The next block represents estimation of the model using a disaggregate data set. The results of the estimation process may lead to revisions of the model formulation as indicated by the broken feedback arrow. The next block represents the process of prediction using the estimated disaggregate model and predicted values of the model's variables. The predicted values of the variables describe anticipated future conditions, including the effects of policy measures. The final block represents the output of the modeling and prediction process: predicted aggregate travel volumes conditional on the predicted values of the model's variables.

The preceding modules have discussed the formulation and estimation of disaggregate mode choice models. This module is concerned with methods for using an estimated disaggregate model to make aggregate predictions. Section 7.2 reviews the reasons for estimating disaggregate rather than aggregate models. Section 7.3 describes the relations between aggregate and disaggregate travel behavior. Section 7.4 describes and evaluates three methods that can be used to make aggregate predictions with disaggregate models, and Section 7.5 presents an example of the application of one of these procedures.

7.2 Reasons for Estimating Disaggregate Models

Since a special methods are needed to obtain aggregate predictions from disaggregate models, it is reasonable to ask why disaggregate models should be used for making such predictions. It may seem that a better procedure would be to estimate models directly from aggregate data and use the resulting aggregate models to make aggregate predictions. There are two

important reasons for not basing aggregate predictions on models estimated from aggregate data. First, estimation of models from aggregate data does not use the data efficiently. It wastes some of the information contained in the data. For example, aggregate data sets often are obtained by computing average values of demographic characteristics and travel behavior of individuals living in the same geographical area (usually a traffic zone or district). The use of such average values discards information about differences among individuals within the same district or zone. The loss of such information, which is expensive to collect, is a waste of resources. It is particularly important to avoid this waste in an era when data collection costs are high and the resources available for transportation studies are limited. The use of disaggregate models, which use data more efficiently than do aggregate models, makes it possible to collect less data than otherwise would be needed and, thereby, to conserve planning resources.

The second important reason for estimating disaggregate models rather than aggregate ones is that estimation of models from aggregate data often yields parameter estimates that do not correctly reflect the relations that influence travel behavior. Such incorrect estimation will lead to an improper understanding of travel behavior and may result in the design and selection of transportation alternatives that are not effective in meeting public objectives.

The incorrect estimation of parameters when aggregate data are used results from complicated statistical relations in the data. These relations are difficult to explain in non-technical terms. As an alternative, two examples will be used to illustrate the types of errors that can occur. The first example is based on estimation of a linear regression model of household trip generation as a function of automobile ownership. The second

example is based on estimation of a logit model of choice between automobile and bus. The linear regression example illustrates the effects of data aggregation in a familiar context. The mode choice example illustrates these effects in the travel prediction context that is the subject of this course.

Each example assumes the existence of an underlying relation that describes the behavior of the individual or household. This relation is used to generate simulated data about individuals or households. The data are then aggregated in various ways, and the aggregated data are used to estimate models of trip generation and mode choice. Finally, the models estimated from the aggregated data are compared with the original models to determine the extent to which the aggregate models recover the true parameter values.

Example 7.1: A Linear Regression Model of Trip Generation

Assume that the number of daily trips made by the members of a household is related to the number of automobiles owned by the household as follows:

$$\text{Trips} = 2 + 2A + U, \quad (7.1)$$

where A is the number of automobiles owned, and U is a random term that represents the effects on trip generation of variables other than automobile ownership. Assume that the probability distribution of U is:

<u>U</u>	<u>Probability</u>
+2	0.1
+1	0.2
0	0.4
-1	0.2
-2	0.1

This distribution makes it possible to compute the probability that the number of trips made by a household is $2 + 2A - 2$, $2 + 2A - 1$, and so forth up to $2 + 2A + 2$. For example, the probability that the number of trips is $2 + 2A - 2$ is 0.1.

Suppose that the following data on trip generation by a sample of households residing in 3 traffic districts have been obtained:

District 1		District 2		District 3	
<u>Autos</u>	<u>Trips</u>	<u>Autos</u>	<u>Trips</u>	<u>Autos</u>	<u>Trips</u>
1	2	1	4	1	5
1	3	1	4	1	5
1	3	1	4	1	6
1	4	2	6	2	7
2	4	2	6	2	7
2	5	2	6	2	8
2	5	2	6	3	8
3	6	3	8	3	9
3	7	3	8	3	9
3	7	3	8	3	10

If linear regression is used to estimate the relation between trips and automobiles owned from the individual data, the resulting estimated model is

$$\text{Trips} = 2 + 2A.$$

Thus, the dependence of trip generation on automobile ownership that is given by the original model is recovered exactly. This dependence is illustrated by the solid line in Figure 7.2.

Now suppose that the data are summarized by district average values as follows:

	District 1		District 2		District 3	
	<u>Autos</u>	<u>Trips</u>	<u>Autos</u>	<u>Trips</u>	<u>Autos</u>	<u>Trips</u>
Total	19	46	20	60	21	74
Average	1.9	4.6	2.0	6.0	2.1	7.4

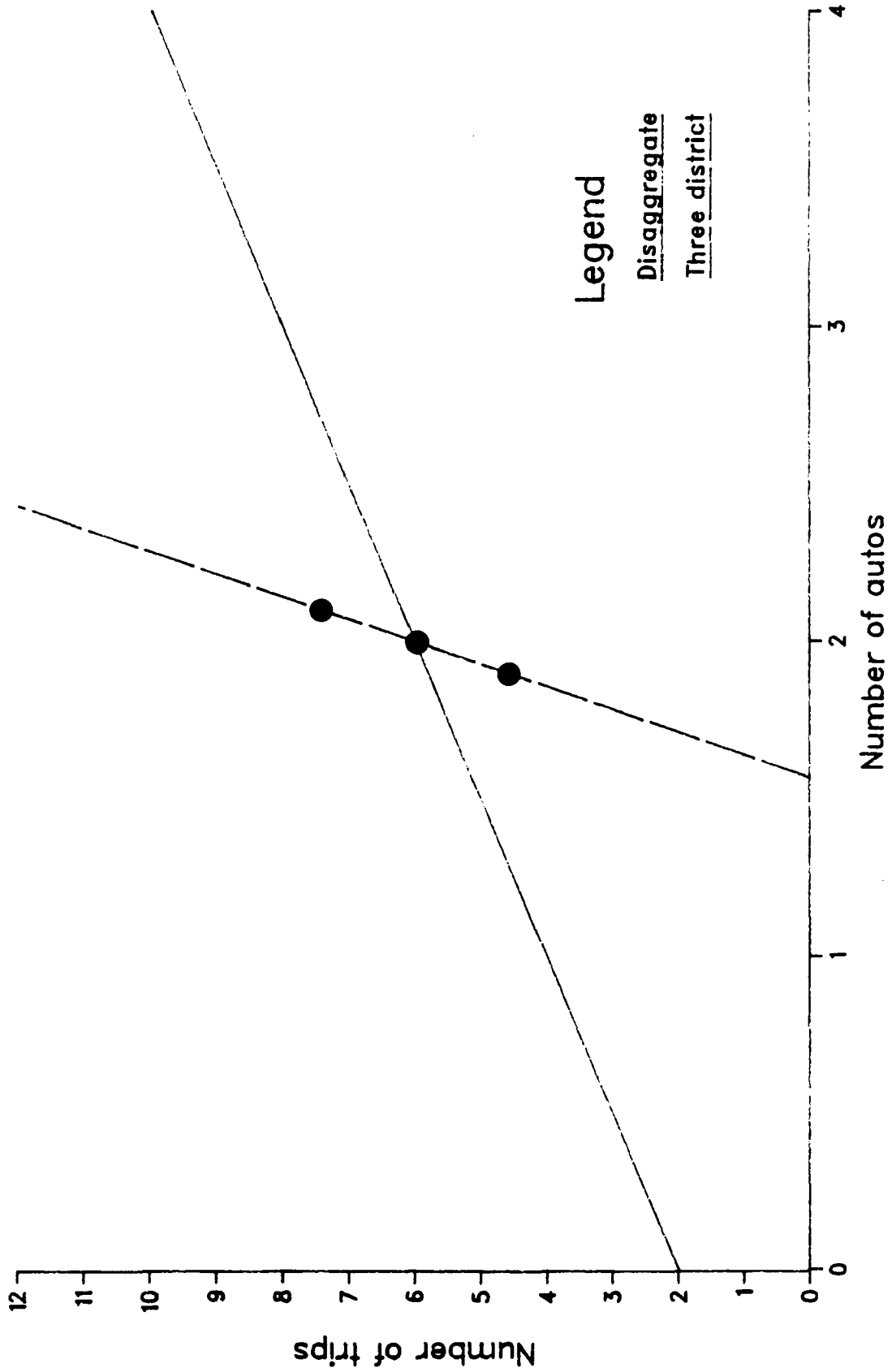
If linear regression is used to estimate average trips per household from these aggregated data, the resulting estimated model is

$$\text{Trips} = -22 + 14A. \tag{7.2}$$

This relation is illustrated by the dashed line in Figure 7.2. Although the aggregate model correctly replicates the district averages, as illustrated by the points in Figure 7.2, it incorrectly predicts that increases in automobile ownership increase trips at the rate of 14 trips per automobile on the average rather than two trips per automobile. An error of this magnitude would cause seriously erroneous predictions of future travel under conditions of changed automobile ownership.

The result just obtained is specific to the districts that were used for aggregating the data. The use of different districts produces different estimation results. For example, if the boundaries of the districts are

Figure 7.2: Disaggregate and Three District
Linear Regression



changed in a way that causes the first 5 households in district 2 to be reassigned to district 1 and the last 5 to be assigned to district 3, the total and average values of the data become:

	District 1		District 3	
	<u>Autos</u>	<u>Trips</u>	<u>Autos</u>	<u>Trips</u>
Total	26	70	34	110
Average	1.73	4.67	2.27	7.33

Application of linear regression to these aggregate data yields the estimated relation

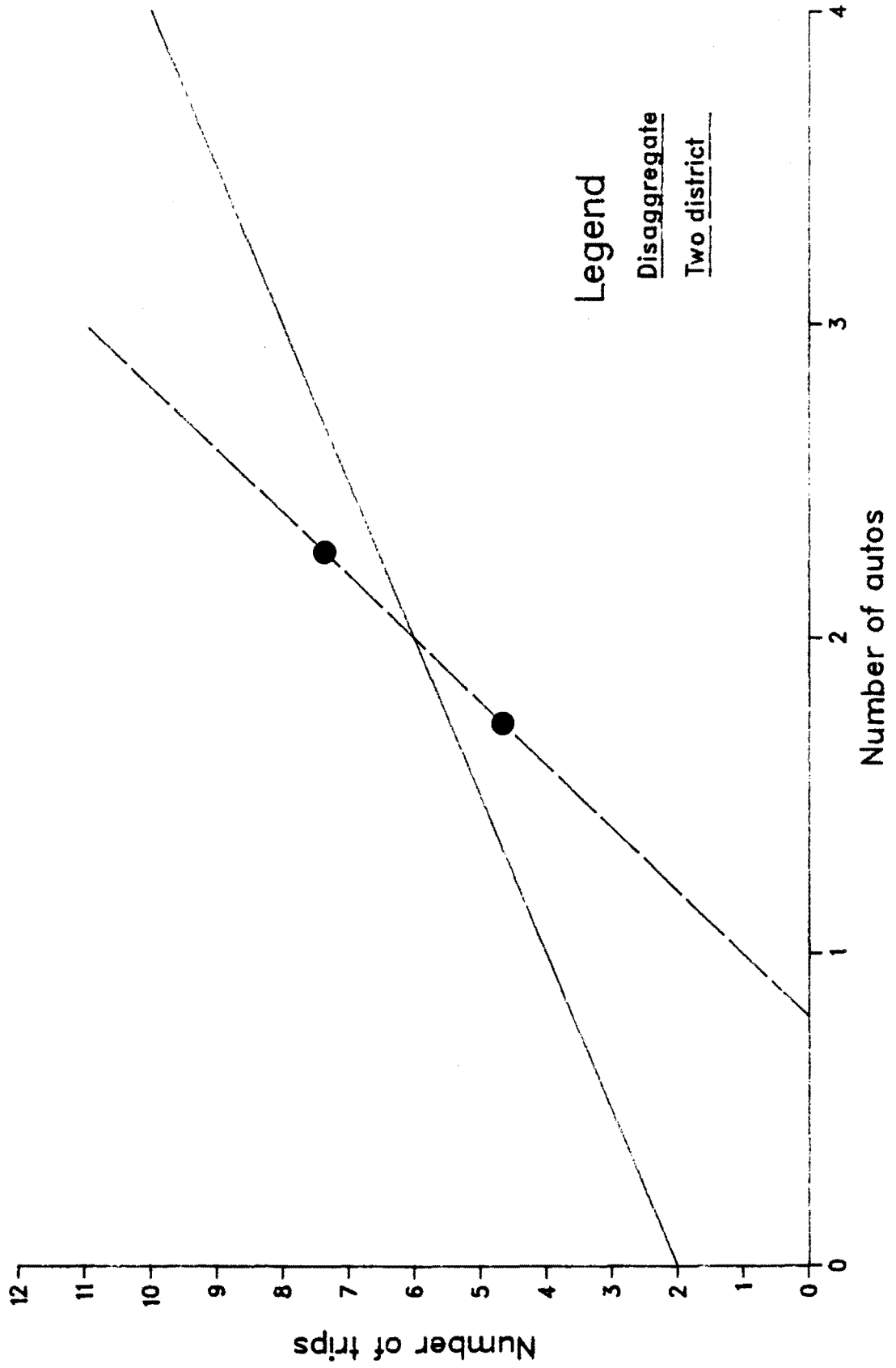
$$\text{Trips} = -4 + 5A. \quad (7.3)$$

This relation is illustrated by the dashed line in Figure 7.3. Although Equation (7.3) is closer than is Equation (7.2) to the original model, it still seriously overestimates the effect of automobile ownership on trip generation.

In practice, it is not possible to know how close a particular set of parameter values estimated from aggregate data is to the true parameter values. Therefore, it is not possible to select districts or other groupings of the data that can be certain of producing parameter estimates that are close to the true values.

A similar pattern of bias in parameter estimates occurs when aggregate mode share models are estimated instead of disaggregate mode choice models. Bias in aggregate mode share models is illustrated by the following example.

Figure 7.3: Disaggregate and Two District
Linear Regression



Example 7.2: Aggregate Mode Share Models

Let choice between automobile and bus for travel to work be described by the binomial logit model whose utility function is

$$V_a = 1.5 - 0.1T_a \quad (7.4)$$

$$V_b = \quad - 0.1T_b, \quad (7.5)$$

where the subscripts a and b signify automobile and bus, respectively, and T is travel time in minutes. Substitution of these utility functions into the formula for binomial logit choice probabilities (Equation (4.3) of Module 4) yields the following probability that automobile is chosen:

$$P_a = \frac{1}{1 + \exp[-1.5 - 0.1(T_b - T_a)]} \quad (7.6)$$

In a large sample, the proportion of individuals with a given value of $T_b - T_a$ who choose automobile is approximately equal to the probability that a single individual with the same value of $T_b - T_a$ chooses automobile. The following simulated data set is based on these relations with individuals assigned to two traffic districts:

$T_b - T_a$ (min.)	District 1		$T_b - T_a$ (min.)	District 2	
	Auto	Bus		Auto	Bus
20	49	1	15	47	3
15	48	2	10	45	5
10	48	3	5	42	8
5	46	4	0	39	11
0	43	7	-5	34	16
-5	39	11	-10	31	19

Estimation of the utility functions for automobile and bus using the methods described in Module 6 and the foregoing individual data yields

$$V_a = 1.496 - 0.101T_a \quad (7.7)$$

$$V_b = - 0.101T_b. \quad (7.8)$$

The estimated utility function is almost identical to the original one given in Equations (7.4) and (7.5).

Now suppose that the travel times and mode choices are averaged according to district to yield:

	<u>District 1</u>	<u>District 2</u>
Average Value of $T_b - T_a$ (min.)	7.5	2.5
Auto Volume	273	238
Bus Volume	27	62
Auto Share	0.910	0.793
Bus Share	0.090	0.207

The estimated utility function corresponding to this aggregate data set is

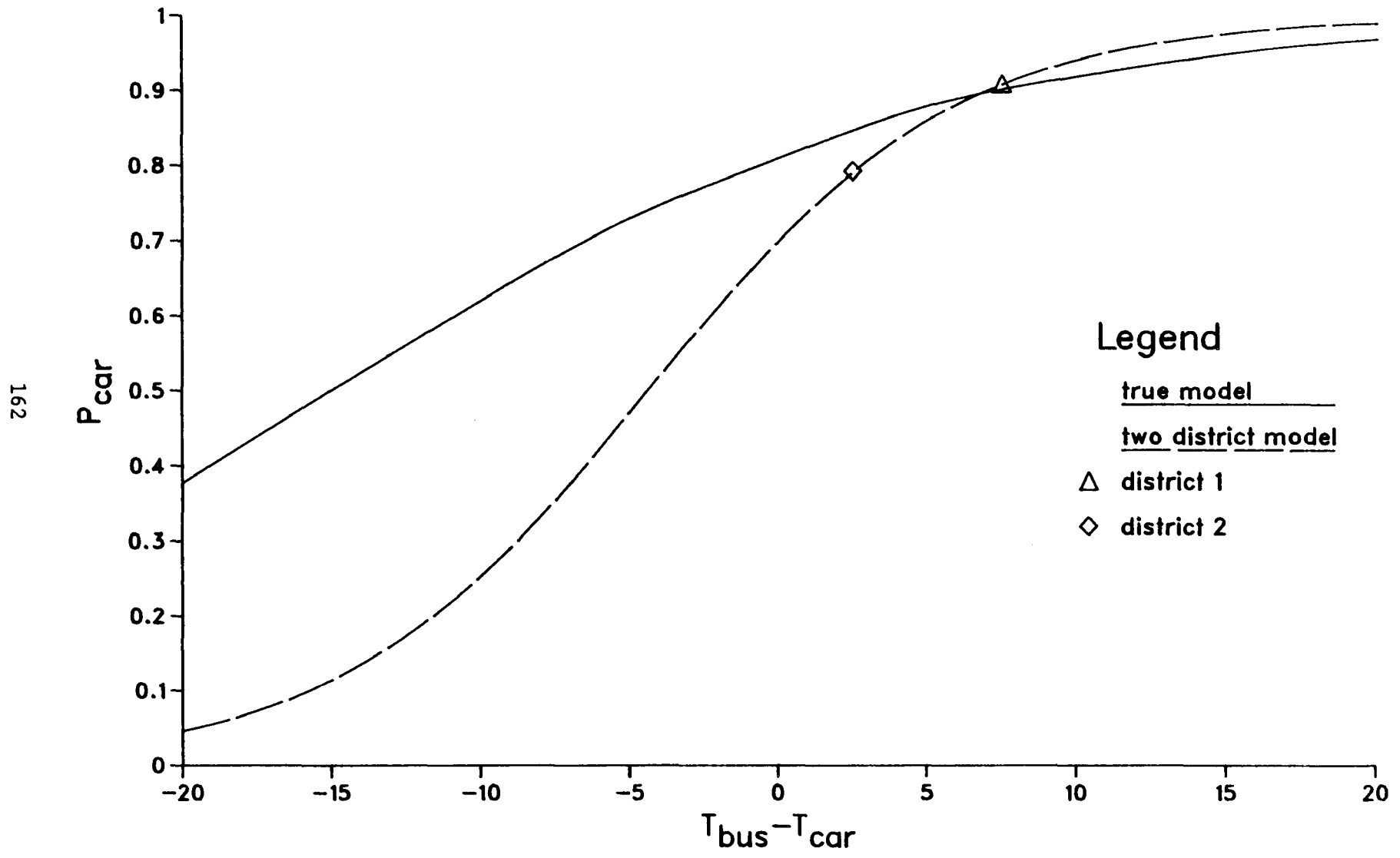
$$V_a = 0.861 - 0.194T_a \quad (7.9)$$

$$V_b = - 0.194T_b. \quad (7.10)$$

The model based on this utility function is illustrated in Figure 7.4. As in the linear regression example, use of aggregate data to estimate a model has yielded seriously biased parameter estimates.

As in the linear regression case, grouping the data into different geographical areas produces differently biased parameter estimates. For example, the districts that have been defined can be divided into two

Figure 7.4: Two District Mode Choice Model



traffic zones each such that the average time differences and mode shares are as shown in the following table and Figure 7.5:

	Zone <u>1A</u>	Zone <u>1B</u>	Zone <u>2A</u>	Zone <u>2B</u>
Average Value of $T_b - T_a$ (min)	12.5	5.0	5.0	0.0
Auto Volume	97	176	128	110
Bus Volume	3	24	22	40
Auto Share	0.970	0.880	0.853	0.733
Bus Share	0.030	0.120	0.147	0.267

The estimated utility function based on the zonal data is

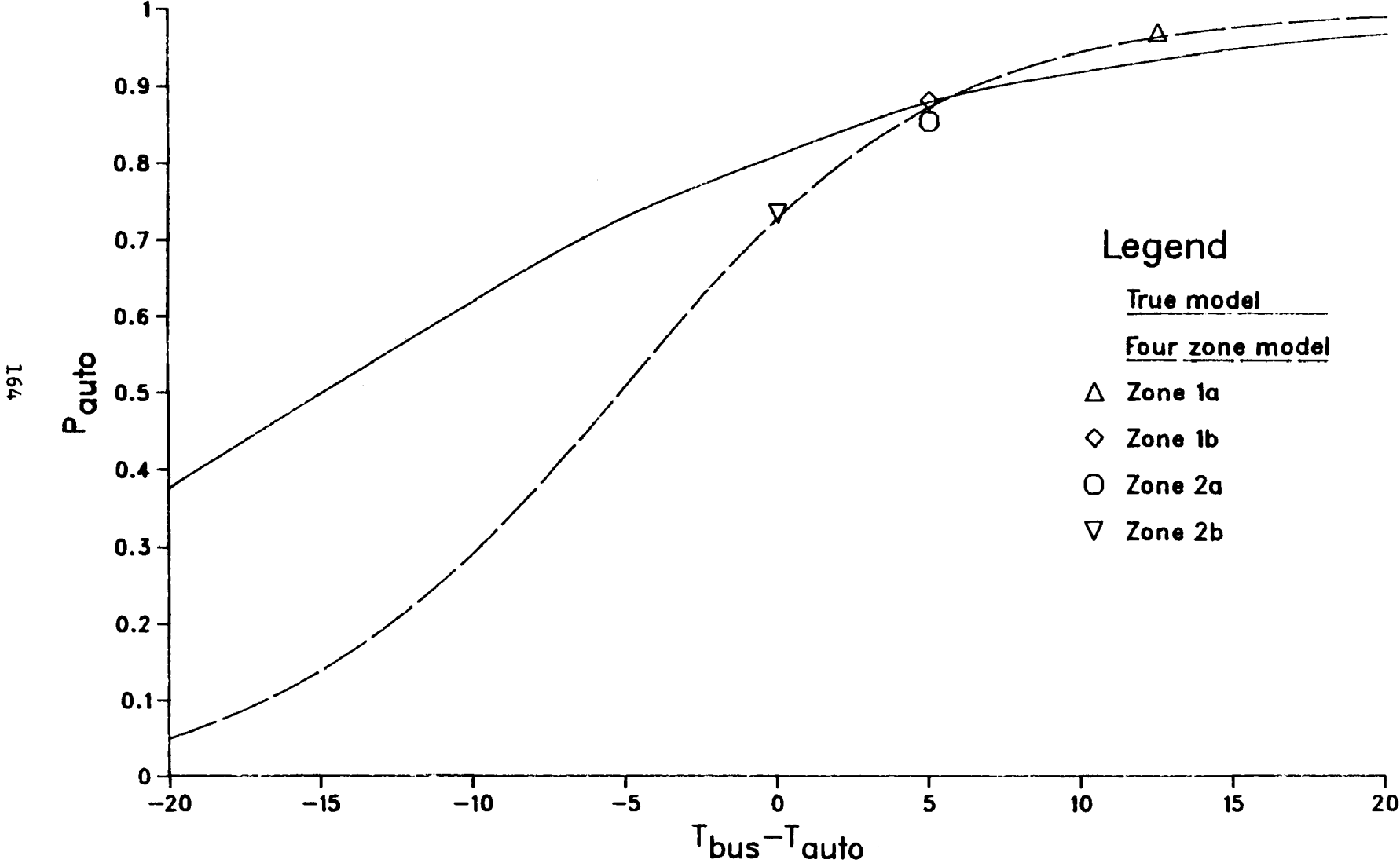
$$V_a = 0.987 - 0.187T_a \quad (7.11)$$

$$V_b = - 0.187T_b \quad (7.12)$$

This estimate is different from the one obtained using the district data and from the original model of Equations (7.4) and (7.5). Still different estimates would be obtained using different groupings of the data.

This example shows that, as with linear regression models, use of aggregate data to estimate mode choice models is likely to produce incorrect parameter estimates. The errors in parameter estimation depend on the grouping of observations. The resulting models will give incorrect predictions of the effects of changes in automobile or bus travel times. This failure to properly represent the effects of changes in travel time occurs despite the fact that the estimated model accurately replicates the aggregate mode shares in the data (as shown in Figures 7.4 and 7.5). The estimation errors that are caused by use of aggregate data typically increase in severity as explanatory variables are added to the model.

Figure 7.5: Four Zone Mode Choice Model



Despite these problems in estimating models from aggregate data, planning practice requires information on the aggregate share of individuals choosing each mode. Thus, it is necessary to develop methods for correctly estimating models of travelers' responses to service changes and to correctly predict aggregate mode shares and volumes. The preceding modules have described the formulation and estimation of disaggregate models that correctly describe travelers' responses to service changes. The following sections of this module describe how to use disaggregate models of individual choice to predict aggregate shares and volumes.

7.3 Relation between Individual Choices and Mode Shares

If the choices of individuals are known with certainty, the number of individuals choosing a given mode can be obtained by simple counting. The aggregate share of the mode is obtained by dividing the total number of individuals choosing the mode by the total number of individuals in the sample. This is what was done to obtain the mode shares in Example 7.2. For example, in the two-district case, the automobile share for district 1 was obtained by counting the individuals choosing car (273 in district 1) and dividing by the total number of individuals in the district 1 sample (300), thereby obtaining the share of 0.910.

When only probabilities of choices are known, rather than actual choices, the number of individuals predicted to choose a mode is the sum of the probabilities of choosing that mode for all of the individuals in the population of travelers. Thus, for example, if the available modes are automobile and bus, the numbers of individuals predicted to choose each are

$$N_{\text{auto}} = \sum_n P_{\text{auto}}(n), \quad (7.13)$$

and

$$N_{bus} = \sum_n P_{bus}(n), \quad (7.14)$$

where $P_{auto}(n)$ and $P_{bus}(n)$, respectively, are the probabilities that individual n chooses automobile and bus. The corresponding mode shares are obtained by dividing the predicted numbers choosing each mode by the total number of individuals in the population. That is

$$S_{auto} = (1/N) \sum_n P_{auto}(n) \quad (7.15)$$

and

$$S_{bus} = (1/N) \sum_n P_{bus}(n), \quad (7.16)$$

where S_{auto} and S_{bus} , respectively, denote the automobile and bus shares, and N is the total number of travelers by both modes combined. Thus, the predicted share of a mode is the average value of its choice probability in the population of travelers.

The following example illustrates this prediction process using a binomial logit model of choice between automobile and bus.

Example 7.3: Aggregate Prediction with a Disaggregate Model

Let the logit model's utility function be

$$V_a = 0.5 - 0.1T_a + 0.5A \quad (7.17)$$

$$V_b = - 0.1T_b, \quad (7.18)$$

where the subscripts a and b denote automobile and bus, respectively, T denotes travel time in minutes, and A is the number of automobiles owned by the traveler's household. Substitution of this utility function into Equation (4.3) yields the following formula for the probability that automobile is chosen:

$$P_a = \frac{1}{1 - \exp[-0.5 - 0.1(T_b - T_a) - 0.5A]} \quad (7.19)$$

The following table gives the values of P_a corresponding to various values of $T_b - T_a$ and A :

A = 1			A = 2		
$T_b - T_a$ (min.)	No. of Cases	P_a	$T_b - T_a$ (min.)	No. of Cases	P_a
10	20	0.881	30	20	0.989
5	20	0.818	25	20	0.982
0	20	0.731	20	20	0.971
-5	20	0.622	15	20	0.953
-10	20	0.500	10	20	0.924
-15	20	0.378	5	20	0.881

The number of individuals predicted to choose automobile is obtained by multiplying the number of cases for each time difference and automobile ownership level by the corresponding probability that automobile is chosen and, then, summing over all time differences and automobile ownership levels. The predicted automobile share is obtained by dividing the number of individuals predicted to choose automobile by the number of individuals under consideration. The results of these computations are:

Number predicted to choose auto: 192.6

Predicted auto share: 0.802.

The method used in the foregoing example to obtain aggregate predictions from a disaggregate model is straightforward. However, to use this method in practice, it would be necessary to compute mode choice

probabilities for each individual in the population of interest. In the example, this computation was easy because the numbers of travel time differences and automobile ownership levels were small. In practice, there would be many more levels, and an impractically large data set would be needed to compute choice probabilities for every individual. Therefore, it is necessary in practice to have procedures that do not require enumerating the entire population to obtain aggregate predictions from disaggregate models. Several practical procedures are described and evaluated in the next section.

7.4 Methods for Aggregate Prediction with Disaggregate Models

This section describes the three most frequently used methods for making aggregate predictions with disaggregate models. The first is the naive method, which is popular because of its simplicity and its similarity to methods used in prediction with aggregate models. The second is the market segmentation method, which is also easy to apply and can give substantially improved predictions using a small quantity of additional data. The third is the sample enumeration method, which provides very good estimates of aggregate behavior when adequate data are available.

7.4a The Naive Method

This method consists of substituting average values of all of the explanatory variables into the utility equations of the logit mode choice model. The resulting average utility values then are substituted into the logit formula to obtain estimates of mode shares. This method, however, does not yield the same aggregate share estimates that would be obtained by computing the choice probabilities of individuals and averaging these

probabilities according to Equation (7.15). The difference between the two estimates is a consequence of the fact that the average of a nonlinear function (in this case the average of logit choice probabilities) is not equal to the function evaluated at the average values of its variables (in this case the logit choice probabilities of an individual with the average values of the explanatory variables). The errors associated with the naive method are illustrated in the following example.

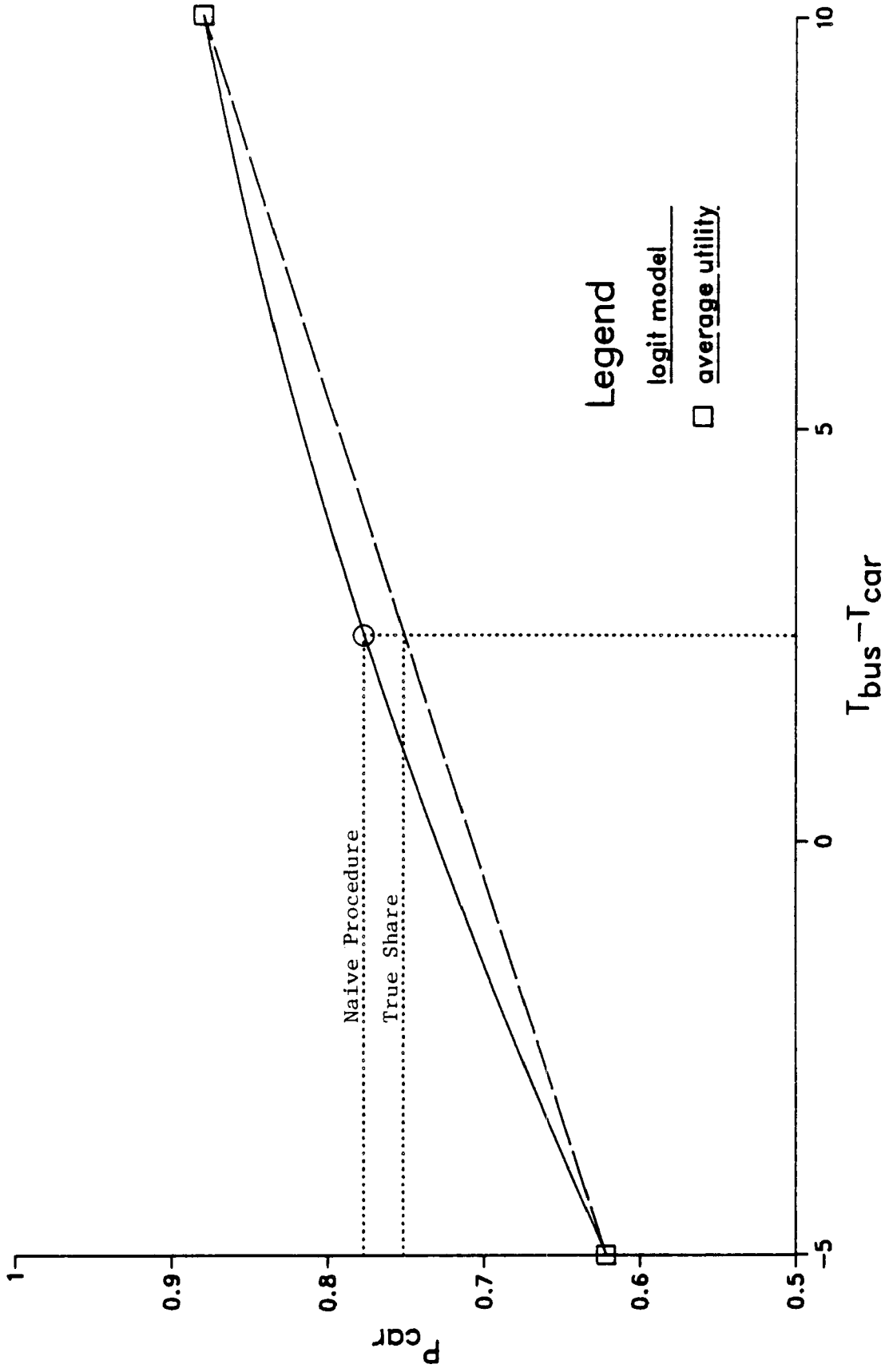
Example 7.4: Errors of the Naive Aggregation method

Consider the choice between automobile and bus for travel to work. Assume that the logit model of Equation (7.19) applies and that all of the individuals of concern own one car ($A = 1$ in Equation (7.19)). Suppose that the difference between bus and automobile travel time, $T_b - T_a$, is 10 min. for one half of the individuals and -5 minutes for the other half. Then the probabilities P_a that automobile is chosen for the two groups of individuals are as follows:

	<u>Group 1</u>	<u>Group 2</u>
$T_a - T_b$ (min.)	10	-5
$V_b - V_a$	-2.0	-0.5
P_a	0.881	0.622

The true aggregate share of automobile in this case is the weighted average of the choice probabilities of the individuals in the two groups, where the weights are proportional to group size. Since, in this example, the sizes of the groups are equal, the weights are 0.5 for each group and the aggregate share is $0.5(0.881) + 0.5(0.622) = 0.752$, as shown in Figure 7.6.

Figure 7.6: Prediction Aggregation Error



However, when the naive method is used, the aggregate share is based on the average difference between the utilities of bus and automobile. In this example, the average difference is $0.5(-2.0) + 0.5(-0.5) = -1.25$. The value of P_a corresponding to this utility difference is $1/[1 + \exp(-1.25)] = 0.777$, as is also shown in Figure 7.6. The naive method produces an error of 0.026 in the predicted shares of automobile and bus. The percentage errors are $100(0.777 - 0.752)/0.752 = 3.3\%$ for the automobile share and $100(0.223 - 0.248)/0.248 = -10.1\%$ for the bus share.

Now consider the data used in Example 7.3. As was shown in that example, the correct aggregate automobile share is 0.802. The naive aggregation method sets the aggregate share equal to the logit automobile choice probability at the average values of the variables. In Example 7.3, the average value of $T_b - T_a$ is 7.5 min. and the average value of A is 1.5. Accordingly, the aggregate automobile share predicted by the naive method is $1/(1 + \exp[-0.5 - 0.1(7.5) - 0.5(1.5)]) = 0.881$. The correct share and the share predicted by the naive method differ by 0.079. The corresponding percentage errors are 10% for the automobile share and 40% for the bus share. Thus, in this case the naive method makes very large prediction errors.

In practice, the sizes of the prediction errors made by the naive method depend on the distributions of utility values in the population for which the predictions are being made. The foregoing example shows that the errors can be large. Large errors also have been found in actual practice. The errors made by the naive method generally reduce the predicted shares of low probability modes and increase the predicted shares of high probability modes. Such prediction errors may seriously affect the evaluation of

transportation policy options. Therefore, it is important to seek other methods of aggregate prediction that reduce these prediction errors.

7.4b The Market Segmentation Method

This method divides the population for which a forecast is required into segments within which individuals are similar but not necessarily identical in terms of the values of important explanatory variables. Separate mode share predictions are made for each segment using the naive method, and the mode share for the entire population is obtained by weighted averaging of the shares for the segments using weights proportional to the segment sizes. In most cases, this method substantially improves the accuracy of the aggregate predictions relative to applying the naive method to the entire population. As an illustration of this, suppose that in Example 7.3 the market segments consist of the different levels of automobile ownership. The market segmentation method predicts the aggregate mode shares for each group by substituting into the logit utility functions the true automobile ownership levels and the average travel time differences for each of these levels. The results of this computation are summarized in the following table:

	<u>A = 1</u>	<u>A = 2</u>
Average value of $T_b - T_a$ (min.)	-2.5	17.5
Number of cases	120	120
Auto choice share	0.679	0.963

The market segmentation method's prediction of the automobile share for the entire population is $[120(0.679) + 120(0.963)]/240 = 0.821$. This prediction

is considerably closer to the true share of 0.802 than is the prediction of the naive method.

The accuracy of the market segmentation method can be increased by increasing the number of market segments. For example, in the case just described, the prediction accuracy could be improved by dividing the district into geographical subareas so as to reduce the variation of travel times within segments. The variables used to define market segments also influence the accuracy of the method. It is best to select variables whose variation is likely to have large effects on mode choice in the circumstances being modeled.

7.4c The Sample Enumeration Method

The ultimate extension of the market segmentation method is to continue subdividing the population until each segment consists of a single individual. However, data on the explanatory variables of models usually cannot be obtained for every individual in a population of practical interest. A practical alternative is to base predictions on a random sample of individuals from the population. This method, called the sample enumeration method, predicts the mode choice probabilities for each individual in the sample and averages these as in Equation (7.15) to obtain an estimate of the mode share for the population. The sample can be drawn from the same data used for model estimation. Again using the data from Example 7.3, a sample of 20 individuals might be:

$T_b - T_a$ (min.)	A = 1		$T_b - T_a$ (min.)	A = 2	
	No. of Cases	P_a		No. of Cases	P_a
10	1	0.881	30	1	0.989
5	3	0.818	25	2	0.982
0	1	0.731	20	2	0.971
-5	2	0.622	15	1	0.953
-10	2	0.500	10	3	0.924
-15	1	0.378	5	1	0.881

To estimate the automobile mode share by the sample enumeration method, multiply each probability by the corresponding number of cases, add the products, and divide the sum by the total number of cases. The resulting share is $[1(0.881) + 3(0.818) + \dots + 3(0.924) + 1(0.881)]/20 = 0.809$. This estimate is very close to the correct share of 0.802.

In general, the error made by the sample enumeration method depends on the size and representativeness of the sample. However, for random samples consisting of 20-30 individuals per traffic zone or district, the method usually gives predictions that are very close to the true shares.

The following table compares the shares predicted by the naive, market segmentation, and sample enumeration method for the data in Example 7.3:

	<u>Auto Share</u>	<u>Bus Share</u>	Percent Error in	
			<u>Auto Share</u>	<u>Bus Share</u>
True value	0.802	0.198	0.0%	0.0%
Naive method	0.881	0.119	9.9	-39.9
Market segmentation	0.821	0.179	2.4	-9.6
Sample enumeration	0.809	0.191	0.9	-3.5

In this case, as in most applications, the error associated with the naive method is large, particularly in percentage terms for the bus mode share. This error is smaller for the market segmentation method and is negligible for the sample enumeration method. As suggested by these results and confirmed by other, more complete studies, the sample enumeration method should be used whenever the required data are available. When the data are not available, the market segmentation method should be used. The naive method should not be used except in cases where data limitations preclude the use of the other methods. In such cases, the potential errors associated with the naive method should be recognized, and the uncertain accuracy of the resulting predictions should be considered in decision making.

In cases where the 20-30 observations per traffic zone or district needed to implement the sample enumeration method are not available, a modified procedure called pseudosample enumeration often can be used. This procedure consists of using estimates of the distributions of the relevant variables to generate artificially a sample of individuals in each zone or district. The sample enumeration method then is applied to this pseudosample.

7.5 Summary

This module has described methods for making forecasts of aggregate choice shares and traffic volumes using disaggregate models. The module has explained the need for using disaggregate models even when the objective is to predict aggregate travel. The module has also explained the theoretical basis for making aggregate predictions with disaggregate models. Three

practical methods for making aggregate predictions have been described and their accuracy discussed.

This module concludes the core of the course. The course has been designed to provide you with an introduction to and general understanding of the estimation and use of disaggregate mode choice models. The microcomputer-based exercises in the supplement to the course will help you to solidify your understanding of these models. The references listed at the end of Module 1 provide additional resources if you wish to further enhance your understanding of these models.

SOLUTIONS TO EXERCISES

MODULE 2

2.1

<u>Mode</u>	Utility	
	<u>Y = 40</u>	<u>Y = 10</u>
Drive Alone	2.25	1.50
Carpool	2.12	1.75
Bus	1.91	1.62

The choices are the same as in Example 2.1

2.2

<u>Mode</u>	Utility	
	<u>Y = 40</u>	<u>Y = 10</u>
Drive Alone	-1.50	-2.85
Carpool	-1.99	-2.56
<u>Bus</u>	<u>-2.37</u>	<u>-2.73</u>

Choice Drive Alone Carpool

2.3 $W = 10 - 20U$, and $X = -U^2$ The values of U, V, W, and X corresponding to the travel times and costs in Example 2.1 and two different values of Y are:

<u>Mode</u>	U	
	<u>Y = 40</u>	<u>Y = 10</u>
Drive Alone	-0.75	-1.50
Carpool	-0.88	-1.25
Bus	-1.09	-1.38

<u>Mode</u>	V	
	<u>Y = 40</u>	<u>Y = 10</u>
Drive Alone	-30.0	-15.0
Carpool	-35.0	-12.5
Bus	-43.75	-13.75

	W	
<u>Mode</u>	<u>Y = 40</u>	<u>Y = 10</u>
Drive Alone	-5.0	-20.0
Carpool	-7.5	-15.0
Bus	-11.88	-17.5

	V	
<u>Mode</u>	<u>Y = 40</u>	<u>Y = 10</u>
Drive Alone	-0.56	-2.25
Carpool	-0.77	-1.56
Bus	-1.20	-1.89

All of the utility functions predict that the individual with Y = 40 will choose drive alone and that the individual with Y = 10 will choose carpool.

2.4 The utility values and mode choices according to income group are:

<u>Income</u>	Utility			<u>Choice</u>
	<u>Drive Alone</u>	<u>Carpool</u>	<u>Bus</u>	
14	-1.39	-1.20	-1.18	Bus
18	-1.19	-1.10	-1.14	Carpool
22	-1.07	-1.03	-1.14	Carpool
26	-0.98	-0.99	-1.11	Drive Alone
30	-0.91	-0.96	-1.10	Drive Alone
34	-0.87	-0.93	-1.07	Drive Alone

The aggregate mode shares are 55% for drive alone, 40% for carpool, and 5% for bus.

The average utility values are -1.03 for drive alone, -0.96 for carpool, and -1.11 for bus. According to the average utilities, all travelers are predicted to choose carpool, which is incorrect.

After the reduction in bus travel time, the utilities and mode choices according to income are:

<u>Income</u>	<u>Utility</u>			<u>Choice</u>
	<u>Drive Alone</u>	<u>Carpool</u>	<u>Bus</u>	
14	-1.39	-1.20	-1.13	Bus
18	-1.19	-1.10	-1.09	Bus
22	-1.07	-1.03	-1.06	Carpool
26	-0.98	-0.99	-1.05	Drive Alone
30	-0.91	-0.96	-1.03	Drive Alone
34	-0.87	-0.93	-1.02	Drive Alone

The new aggregate mode shares are 55% for drive alone, 25% for carpool, and 20% for bus. Thus, the reduction in bus travel time has caused the mode share of carpool to decrease and that of bus to increase.

The average utilities after the reduction in bus travel time are -1.03 for drive alone, -0.96 for carpool, and -1.06 for bus. Using average utilities, one predicts, erroneously, that all travelers choose carpool and, therefore, that the reduction in bus travel time has caused no change in the mode shares.

MODULE 3

3.1

$$\begin{aligned}
 \text{One car: } U_{DA} &= -0.5 - 5(2/15) + 0.4(1 - 1) = -1.17 \\
 U_{CP} &= -0.75 - 5(1/15) + 0.2(1 - 1) = -1.08 \\
 U_B &= -1.0 - 5(0.75/15) = -1.25 \\
 \text{Two cars: } U_{DA} &= -0.5 - 5(2/15) + 0.4(2 - 1) = -0.77 \\
 U_{CP} &= -0.75 - 5(1/15) + 0.2(2 - 1) = -0.88 \\
 U_B &= -1.0 - 5(0.75/15) = -1.25 \\
 \text{Three cars: } U_{DA} &= -0.5 - 5(2/15) + 0.4(3 - 1) = -0.37 \\
 U_{CP} &= -0.75 - 5(1/15) + 0.2(3 - 1) = -0.68 \\
 U_B &= -1.0 - 5(0.75/15) = -1.25
 \end{aligned}$$

3.2

Percentage of Individuals	Utilities Based on the Travel Time Distributions in Example 3.2				Based on Ex. 3.1
	20%	50%	20%	10%	100%
Drive Alone	-0.67	-0.77	-0.87	-0.97	-0.77
Carpool	-0.78	-0.88	-0.98	-1.08	-0.88
Bus	-1.25	-1.25	-1.25	-1.25	-1.25
Chosen Mode	Drive alone in all cases				

3.3 0 cars:

Percentage of Individuals	Utilities Based on the Travel Time Distributions in Example 3.2			
	20%	50%	20%	10%
Drive Alone	-1.47	-1.57	-1.67	-1.77
Carpool	-1.18	-1.28	-1.38	-1.48
Bus	-1.25	-1.25	-1.25	-1.25
Chosen Mode	CP	Bus	Bus	Bus
% of All Travelers	4	10	4	2

One car:

Percentage of Individuals	Utilities Based on the Travel Time Distributions in Example 3.2			
	20%	50%	20%	10%
Drive Alone	-1.07	-1.17	-1.27	-1.37
Carpool	-0.98	-1.08	-1.18	-1.28
Bus	-1.25	-1.25	-1.25	-1.25
Chosen Mode	CP	CP	CP	Bus
% of All Travelers	10	25	10	5

Two cars:

Percentage of Individuals	Utilities Based on the Travel Time Distributions in Example 3.2			
	20%	50%	20%	10%
Drive Alone	-0.67	-0.77	-0.87	-0.97
Carpool	-0.78	-0.88	-0.98	-1.08
Bus	-1.25	-1.25	-1.25	-1.25
Chosen Mode	DA	DA	DA	DA
% of All Travelers	6	15	6	3

Aggregate mode shares: 30% (drive alone), 49% (carpool), 21% (bus).

3.4	Income (\$000)	Number of Individuals	Percent Choosing		
			Drive Alone	Carpool	Bus
	17.5	10	10	50	40
	22.5	20	80	20	0
	27.5	20	85	15	0
	32.5	30	90	10	0
	37.5	20	95	5	0
	42.5	3	100	0	0

Income (\$000)	Number of Individuals	Number Choosing		
		Drive Alone	Carpool	Bus
17.5	10	1	5	4
22.5	20	16	4	0
27.5	20	17	3	0
32.5	30	27	3	0
37.5	20	19	1	0
42.5	3	3	0	0
Total	103	83	16	4

MODULE 4

4.1	Case	V_1	V_2	$V_1 - V_2$	Pr(1)
	6	1.0	-1.5	2.5	0.92
	7	0.5	-2.0	2.5	0.92
	8	3.0	-1.0	4.0	0.98
	9	0.0	-0.5	0.5	0.62
	10	0.0	2.5	-2.5	0.08

4.2 Carpool utility = 1.5:

<u>Mode</u>	<u>V</u>	<u>Exp(V)</u>	<u>Pr</u>
DA	2.5	12.18	0.63
CP	1.5	4.48	0.23
Bus	1.0	<u>2.72</u>	<u>0.14</u>
		19.38	1.00

Carpool utility = 1.0:

<u>Mode</u>	<u>V</u>	<u>Exp(V)</u>	<u>Pr</u>
DA	2.5	12.18	0.63
CP	1.0	2.72	0.155
Bus	1.0	<u>2.72</u>	<u>0.155</u>
		19.38	1.000

4.3 a.

<u>Mode</u>	<u>T</u>	<u>C</u>	<u>V</u>	<u>Exp(V)</u>	<u>Pr</u>
DA	0.5	2.0	-1.0	0.37	0.237
CP	0.6	1.0	-0.85	0.43	0.276
Bus	0.8	0.6	-0.95	0.39	0.250
Bike	1.0	0.0	-1.0	<u>0.37</u>	<u>0.237</u>
				1.56	1.00

4.3 b. Alternative-specific constant for bicycle is -1.0:

<u>Mode</u>	<u>T</u>	<u>C</u>	<u>V</u>	<u>Exp(V)</u>	<u>Pr</u>
DA	0.5	2.0	-1.0	0.37	0.278
CP	0.6	1.0	-0.85	0.43	0.323
Bus	0.8	0.6	-0.95	0.39	0.293
Bike	1.0	0.0	-2.0	<u>0.14</u>	<u>0.105</u>
				1.33	1.00

MODULE 5

- 5.1 (1) Percentage of runs that arrive at a given point within 3 minutes of schedule.
- (2) Percentage of runs that arrive at a given point no more than 3 minutes late.
- (3) Root-mean-square deviation from scheduled arrival time.
- 5.2 When $v = -T^2$, adding 5 minutes is more burdensome for the 1-hour trip. When $V = -T^{1/2}$, adding 5 minutes is more burdensome for the 15-minute trip.
- 5.3 a. High-income travelers are more sensitive.
b. Equally sensitive
c. Low income travelers are more sensitive
d. Equally sensitive
e. Low income travelers are more sensitive.
- 5.4 For the first group of utility functions, use $b_1 = -0.05$, $b_2 = -0.10$. For the second group, use $b_3 = 0.05$, $b_4 = -0.05$.

MODULE 6

- 6.1 a. $\log L = -\log[1 + \exp(5a)] - \log[1 + \exp(15a)] - \log[1 + \exp(-10a)]$
- b.

<u>a</u>	<u>log L</u>
-0.025	-1.98
-0.045	-1.94
-0.065	-1.93
- c. To 2 significant figures, there is no value of a that gives a log L larger than -1.93.

6.2 a.	<u>Variable</u>	<u>t statistic</u>
	IVTT	-2.18
	OVTT	-4.54
	LC	-1.44
	PC	-2.05

The t statistics indicate that none of the variables should be dropped.

- b. Estimate the model whose utility function is

$$V = a_1 IVTT + a_2 OVTT + a_3(LC + PC) + a_4 PC,$$

and use the t statistic of a_4 to determine whether the last term can be dropped.

- c. The likelihood ratio test statistic is

$$LR = 2[(-178.22) - (-179.37)] = 2.30.$$

This is below the critical value for two variables, so the two additional variables can be dropped.

- d. The value of the modified likelihood ratio test statistic is

$$MLR = (-177.53 - 4/2) - (-179.37 - 4/2) = 1.84.$$

Since this exceeds the critical value for MLR, the model with OVTT/DIST is significantly better.

DOT-T-93-18

MTA DOROTHY GRAY LIBRARY & ARCHIVE
A self-instructing course in disaggreg
HE336.C5 H67 1986



100000083384

HE 336 .C5 H67 1986

16095

Horowitz, Joel.

273

A self-instructing course in
disaggregate mode choice

TECHNOLOGY SHARING

A Program of the U.S. Department of Transportation

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof.

The United States Government does not endorse manufacturers or products. Trade names appear in the document only because they are essential to the content of the report.

This report is being distributed through the U.S. Department of Transportation's Technology Sharing Program.

DOT-T-93-18