

A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models

Prepared For

U.S. Department of Transportation
Federal Transit Administration

by

Frank S. Koppelman and Chandra Bhat
with technical support from
Vaneet Sethi, Sriram Subramanian, Vincent Bernardin and Jian Zhang

January 31, 2006
Modified June 30, 2006

Table of Contents

ACKNOWLEDGEMENTS	vii
CHAPTER 1 : INTRODUCTION.....	1
1.1 BACKGROUND	1
1.2 USE OF DISAGGREGATE DISCRETE CHOICE MODELS	2
1.3 APPLICATION CONTEXT IN CURRENT COURSE	3
1.4 URBAN AND INTERCITY TRAVEL MODE CHOICE MODELING.....	4
1.4.1 <i>Urban Travel Mode Choice Modeling</i>	4
1.4.2 <i>Intercity Mode Choice Models</i>	4
1.5 DESCRIPTION OF THE COURSE	5
1.6 ORGANIZATION OF COURSE STRUCTURE.....	6
CHAPTER 2 : ELEMENTS OF THE CHOICE DECISION PROCESS	9
2.1 INTRODUCTION.....	9
2.2 THE DECISION MAKER	9
2.3 THE ALTERNATIVES	10
2.4 ATTRIBUTES OF ALTERNATIVES	11
2.5 THE DECISION RULE	12
CHAPTER 3 : UTILITY-BASED CHOICE THEORY.....	14
3.1 BASIC CONSTRUCT OF UTILITY THEORY	14
3.2 DETERMINISTIC CHOICE CONCEPTS	15
3.3 PROBABILISTIC CHOICE THEORY.....	17
3.4 COMPONENTS OF THE DETERMINISTIC PORTION OF THE UTILITY FUNCTION	19
3.4.1 <i>Utility Associated with the Attributes of Alternatives</i>	20
3.4.2 <i>Utility ‘Biases’ Due to Excluded Variables</i>	21
3.4.3 <i>Utility Related to the Characteristics of the Decision Maker</i>	22
3.4.4 <i>Utility Defined by Interactions between Alternative Attributes and Decision Maker Characteristics</i>	23
3.5 SPECIFICATION OF THE ADDITIVE ERROR TERM.....	24
CHAPTER 4 : THE MULTINOMIAL LOGIT MODEL.....	26
4.1 OVERVIEW DESCRIPTION AND FUNCTIONAL FORM	26
4.1.1 <i>The Sigmoid or S shape of Multinomial Logit Probabilities</i>	31
4.1.2 <i>The Equivalent Differences Property</i>	32
4.2 INDEPENDENCE OF IRRELEVANT ALTERNATIVES PROPERTY	38
4.2.1 <i>The Red Bus/Blue Bus Paradox</i>	40
4.3 EXAMPLE: PREDICTION WITH MULTINOMIAL LOGIT MODEL	41
4.4 MEASURES OF RESPONSE TO CHANGES IN ATTRIBUTES OF ALTERNATIVES	46
4.4.1 <i>Derivatives of Choice Probabilities</i>	46
4.4.2 <i>Elasticities of Choice Probabilities</i>	48
4.5 MEASURES OF RESPONSES TO CHANGES IN DECISION MAKER CHARACTERISTICS.....	51
4.5.1 <i>Derivatives of Choice Probabilities</i>	51
4.5.2 <i>Elasticities of Choice Probabilities</i>	53
4.6 MODEL ESTIMATION: CONCEPT AND METHOD.....	54
4.6.1 <i>Graphical Representation of Model Estimation</i>	54
4.6.2 <i>Maximum Likelihood Estimation Theory</i>	56
4.6.3 <i>Example of Maximum Likelihood Estimation</i>	58

CHAPTER 5 : DATA ASSEMBLY AND ESTIMATION OF SIMPLE MULTINOMIAL LOGIT MODEL.61

5.1 INTRODUCTION.....61
 5.2 DATA REQUIREMENTS OVERVIEW61
 5.3 SOURCES AND METHODS FOR TRAVELER AND TRIP RELATED DATA COLLECTION.....63
 5.3.1 *Travel Survey Types*.....63
 5.3.2 *Sampling Design Considerations*.....65
 5.4 METHODS FOR COLLECTING MODE RELATED DATA.....68
 5.5 DATA STRUCTURE FOR ESTIMATION.....69
 5.6 APPLICATION DATA FOR WORK MODE CHOICE IN THE SAN FRANCISCO BAY AREA72
 5.7 ESTIMATION OF MNL MODEL WITH BASIC SPECIFICATION74
 5.7.1 *Informal Tests*.....77
 5.7.2 *Overall Goodness-of-Fit Measures*.....79
 5.7.3 *Statistical Tests*.....82
 5.8 VALUE OF TIME.....98
 5.8.1 *Value of Time for Linear Utility Function*.....98
 5.8.2 *Value of Time when Cost is Interacted with another Variable*.....99
 5.8.3 *Value of Time for Time or Cost Transformation*.....101

CHAPTER 6 : MODEL SPECIFICATION REFINEMENT: SAN FRANCISCO BAY AREA WORK MODE CHOICE106

6.1 INTRODUCTION.....106
 6.2 ALTERNATIVE SPECIFICATIONS.....107
 6.2.1 *Refinement of Specification for Alternative Specific Income Effects*.....108
 6.2.2 *Different Specifications of Travel Time*.....111
 6.2.3 *Including Additional Decision Maker Related Variables*.....119
 6.2.4 *Including Trip Context Variables*.....121
 6.2.5 *Interactions between Trip Maker and/or Context Characteristics and Mode Attributes*.....124
 6.2.6 *Additional Model Refinement*.....127
 6.3 MARKET SEGMENTATION.....129
 6.3.1 *Market Segmentation Tests*.....131
 6.3.2 *Market Segmentation Example*.....133
 6.4 SUMMARY137

CHAPTER 7 : SAN FRANCISCO BAY AREA SHOP/OTHER MODE CHOICE139

7.1 INTRODUCTION.....139
 7.2 SPECIFICATION FOR SHOP/OTHER MODE CHOICE MODEL.....141
 7.3 INITIAL MODEL SPECIFICATION.....141
 7.4 EXPLORING ALTERNATIVE SPECIFICATIONS.....144

CHAPTER 8 : NESTED LOGIT MODEL157

8.1 MOTIVATION.....157
 8.2 FORMULATION OF NESTED LOGIT MODEL159
 8.2.1 *Interpretation of the Logsum Parameter*.....163
 8.2.2 *Disaggregate Direct and Cross-Elasticities*.....163
 8.3 NESTING STRUCTURES165
 8.4 STATISTICAL TESTING OF NESTED LOGIT STRUCTURES172

CHAPTER 9 : SELECTING A PREFERRED NESTING STRUCTURE.....175

9.1 INTRODUCTION.....175
 9.2 NESTED MODELS FOR WORK TRIPS176

9.3 NESTED MODELS FOR SHOP/OTHER TRIPS192
 9.4 PRACTICAL ISSUES AND IMPLICATIONS199
CHAPTER 10 : MULTIPLE MAXIMA IN THE ESTIMATION OF NESTED LOGIT MODELS201
 10.1 MULTIPLE OPTIMA201
CHAPTER 11 : AGGREGATE FORECASTING, ASSESSMENT, AND APPLICATION.....209
 11.1 BACKGROUND209
 11.2 AGGREGATE FORECASTING209
 11.3 AGGREGATE ASSESSMENT OF TRAVEL MODE CHOICE MODELS212
CHAPTER 12 : RECENT ADVANCES IN DISCRETE CHOICE MODELING215
 12.1 BACKGROUND215
 12.2 THE GEV CLASS OF MODELS.....216
 12.3 THE MMNL CLASS OF MODELS218
 12.4 THE MIXED GEV CLASS OF MODELS.....221
 12.5 SUMMARY223
REFERENCES224
APPENDIX A : ALOGIT, LIMDEP AND ELM.....231
APPENDIX B : EXAMPLE MATLAB FILES ON CD240

List of Figures

Figure 3.1 Illustration of Deterministic Choice 16
 Figure 4.1 Probability Density Function for Gumbel and Normal Distributions 27
 Figure 4.2 Cumulative Distribution Function for Gumbel and Normal Distribution with the Same Mean and Variance 27
 Figure 4.3 Relationship between V_i and $\text{Exp}(V_i)$ 29
 Figure 4.4 Logit Model Probability Curve 32
 Figure 4.5 Iso-Utility Lines for Cost-Sensitive versus Time-Sensitive Travelers..... 55
 Figure 4.6 Estimation of Iso-Utility Line Slope with Observed Choice Data 56
 Figure 4.7 Likelihood and Log-likelihood as a Function of a Parameter Value..... 60
 Figure 5.1 Data Structure for Model Estimation 71
 Figure 5.2 Relationship between Different Log-likelihood Measures..... 79
 Figure 5.3 t-Distribution Showing 90% and 95% Confidence Intervals 84
 Figure 5.4 Chi-Squared Distributions for 5, 10, and 15 Degrees of Freedom..... 89
 Figure 5.5 Chi-Squared Distribution for 5 Degrees of Freedom Showing 90% and 95% Confidence Thresholds 90
 Figure 5.6 Value of Time vs. Income 100
 Figure 5.7 Value of Time for Log of Time Model 104
 Figure 5.8 Value of Time for Log of Cost Model..... 105
 Figure 6.1 Ratio of Out-of-Vehicle and In-Vehicle Time Coefficients..... 119

Figure 8.1 Two-Level Nest Structure with Two Alternatives in Lower Nest 161

Figure 8.2 Three Types of Two Level Nests 166

Figure 8.3 Three-Level Nest Structure for Four Alternatives..... 167

Figure 9.1 Single Nest Models..... 176

Figure 9.2 Non-Motorized Nest in Parallel with Motorized, Private Automobile and Shared Ride
Nests..... 179

Figure 9.3 Hierarchically Nested Models 182

Figure 9.4 Complex Nested Models 185

Figure 9.5 Motorized – Shared Ride Nest (Model 26W)..... 190

Figure 9.6 Elasticities for MNL (17W) and NL Model (26W)..... 191

Figure A.1 ALOGIT Input Command File 233

Figure A.2 Estimation Results for Basic Model Specification using ALOGIT 234

Figure A.3 LIMDEP Input Command File..... 235

Figure A.4 Estimation Results for Basic Model Specification using LIMDEP 236

Figure A.5 ELM Model Specification 237

Figure A.6 ELM Model Estimation 238

Figure A.7 ELM Estimation Results Reported in Excel..... 239

List of Tables

Table 4-1 Probability Values for Drive Alone as a Function of Drive Alone Utility..... 30

Table 4-2 Probability Values for Drive Alone as a Function of Shared Ride and Transit Utilities
..... 31

Table 4-3 Numerical Example Illustrating Equivalent Difference Property: 34

Table 4-4 Numerical Example Illustrating Equivalent Difference Property: 34

Table 4-5 Utility and Probability Calculation with TRansit as Base Alternative..... 37

Table 4-6 Utility and Probability Calculation with Drive Alone as Base Alternative..... 38

Table 4-7 Changes in Alternative Specific Constants and Income Parameters..... 38

Table 4-8 MNL Probabilities for Constants Only Model 42

Table 4-9 MNL Probabilities for Time and Cost Model 43

Table 4-10 MNL Probabilities for In and Out of Vehicle Time and Cost Model 44

Table 4-11 MNL Probabilities for In and Out of Vehicle Time, Cost and Income Model 45

Table 4-12 MNL Probabilities for In and Out of Vehicle Time, and Cost/Income Model 46

Table 5-1 Sample Statistics for Bay Area Journey-to-Work Modal Data 73

Table 5-2 Estimation Results for Zero Coefficient, Constants Only and Base Models 76

Table 5-3 Critical t-Values for Selected Confidence Levels and Large Samples..... 84

Table 5-4 Parameter Estimates, t-statistics and Significance for Base Model 86

Table 5-5 Critical Chi-Squared (χ^2) Values for Selected Confidence Levels by Number of
Restrictions..... 90

Table 5-6 Likelihood Ratio Test for Hypothesis $H_{0,a}$ and $H_{0,b}$ 92

Table 5-7 Estimation Results for Base Models and its Restricted Versions.....	93
Table 5-8 Likelihood Ratio Test for Hypothesis $H_{0,c}$ and $H_{0,d}$	94
Table 5-9 Models with Cost vs. Cost/Income and Cost/Ln(Income).....	97
Table 5-10 Value of Time vs. Income.....	100
Table 5-11 Base Model and Log Transformations.....	103
Table 5-12 Value of Time for Log of Time Model.....	104
Table 5-13 Value of Time for Log of Cost Model.....	105
Table 6-1 Alternative Specifications of Income Variable.....	110
Table 6-2 Likelihood Ratio Tests between Models in Table 6-1.....	110
Table 6-3 Estimation Results for Alternative Specifications of Travel Time.....	113
Table 6-4 Implied Value of Time in Models 1W, 5W, and 6W.....	113
Table 6-5 Estimation Results for Additional Travel Time Specification Testing.....	116
Table 6-6 Model 7W Implied Values of Time as a Function of Trip Distance.....	118
Table 6-7 Implied Values of Time in Models 6W, 8W, 9W.....	118
Table 6-8 Estimation Results for Auto Availability Specification Testing.....	121
Table 6-9 Estimation Results for Models with Trip Context Variables.....	122
Table 6-10 Implied Values of Time in Models 13W, 14W, and 15W.....	124
Table 6-11 Comparison of Models with and without Income as Interaction Term.....	125
Table 6-12 Implied Value of Time in Models 15W and 16W.....	127
Table 6-13 Estimation Results for Model 16W and its Constrained Version.....	128
Table 6-14 Estimation Results for Market Segmentation by Automobile Ownership.....	133
Table 6-15 Estimation Results for Market Segmentation by Gender.....	135
Table 7-1 Sample Statistics for Bay Area Home-Based Shop/Other Trip Modal Data.....	140
Table 7-2 Base Shopping/Other Mode Choice Model.....	142
Table 7-3 Implied Value of Time in Base S/O Model.....	144
Table 7-4 Alternative Specifications for Household Size.....	145
Table 7-5 Alternative Specifications for Vehicle Availability.....	146
Table 7-6 Alternative Specifications for Income.....	148
Table 7-7 Alternative Specifications for Travel Time.....	149
Table 7-8 Alternative Specifications for Cost.....	151
Table 7-9 Composite Specifications from Earlier Results Compared with other Possible Preferred Specifications.....	153
Table 7-10 Refinement of Final Specification Eliminating Insignificant Variables.....	155
Table 8-1 Illustration of IIA Property on Predicted Choice Probabilities.....	158
Table 8-2 Elasticity Comparison of Nested Logit vs. MNL Models.....	165
Table 8-3 Number of Possible Nesting Structures.....	172
Table 9-1 Single Nest Work Trip Models.....	177
Table 9-2 Parallel Two Nest Work Trip Models.....	180
Table 9-3 Hierarchical Two Nest Work Trip Models.....	183
Table 9-4 Complex and Constrained Nested Models for Work Trips.....	186
Table 9-5 MNL (17W) vs. NL Model 26W.....	188
Table 9-6 Single Nest Shop/Other Trip Models.....	192

Table 9-7 Parallel Two Nest Models for Shop/Other Trips.....	194
Table 9-8 Hierarchical Two Nest Models for Shop/Other Trips	196
Table 9-9 Complex Nested Models for Shop/Other Trips.....	198
Table 10-1 Multiple Solutions for Model 27W (See Table 9-3)	202
Table 10-2 Multiple Solutions for Model 20 S/O (See Table 9-7).....	204
Table 10-3 Multiple Solutions for Model 22 S/O (See Table 9-8).....	206
Table 10-4 Multiple Solutions for Complex S/O Models (See Table 9-9).....	207
Table B-1 Files / Directory Structure.....	240

Acknowledgements

This manual was prepared under funding of the United States Department of Transportation through the Federal Transit Administration (Agmt. 8-17-04-A1/DTFT60-99-D-4013/0012) to AECOMConsult and Northwestern University.

Valuable reviews and comments were provided by students in travel demand modeling classes at Northwestern University and the Georgia Institute of Technology. In addition, valuable comments, suggestions and questions were given by Rick Donnelly, Laurie Garrow, Joel Freedman, Chuck Purvis, Kimon Proussaloglou, Bruce Williams, Bill Woodford and others. The authors are indebted to all who commented on any version of this report but retain responsibility for any errors or omissions.

CHAPTER 1: Introduction

1.1 Background

Discrete choice models can be used to analyze and predict a decision maker's choice of one alternative from a finite set of mutually exclusive and collectively exhaustive alternatives. Such models have numerous applications since many behavioral responses are discrete or qualitative in nature; that is, they correspond to choices of one or another of a set of alternatives.

The ultimate interest in discrete choice modeling, as in most econometric modeling, lies in being able to predict the decision making behavior of a group of individuals (we will use the term "individual" and "decision maker" interchangeably, though the decision maker may be an individual, a household, a shipper, an organization, or some other decision making entity). A further interest is to determine the relative influence of different attributes of alternatives and characteristics of decision makers when they make choice decisions. For example, transportation analysts may be interested in predicting the fraction of commuters using each of several travel modes under a variety of service conditions, or marketing researchers may be interested in examining the fraction of car buyers selecting each of several makes and models with different prices and attributes. Further, they may be interested in predicting this fraction for different groups of individuals and identifying individuals who are most likely to favor one or another alternative. Similarly, they may be interested in understanding how different groups value different attributes of an alternative; for example are business air travelers more sensitive to total travel time or the frequency of flight departures for a chosen destination.

There are two basic ways of modeling such aggregate (or group) behavior. One approach directly models the aggregate share of all or a segment of decision makers choosing each alternative as a function of the characteristics of the alternatives and socio-demographic attributes of the group. This approach is commonly referred to as the aggregate approach. The second approach is to recognize that aggregate behavior is the result of numerous individual decisions and to model individual choice responses as a function of the characteristics of the

alternatives available to and socio-demographic attributes of each individual. This second approach is referred to as the disaggregate approach.

The disaggregate approach has several important advantages over the aggregate approach to modeling the decision making behavior of a group of individuals. First, the disaggregate approach explains why an individual makes a particular choice given her/his circumstances and is, therefore, better able to reflect changes in choice behavior due to changes in individual characteristics and attributes of alternatives. The aggregate approach, on the other hand, rests primarily on statistical associations among relevant variables at a level other than that of the decision maker; as a result, it is unable to provide accurate and reliable estimates of the change in choice behavior due changes in service or in the population. Second, the disaggregate approach, because of its causal nature, is likely to be more transferable to a different point in time and to a different geographic context, a critical requirement for prediction. Third, discrete choice models are being increasingly used to understand behavior so that the behavior may be changed in a proactive manner through carefully designed strategies that modify the attributes of alternatives which are important to individual decision makers. The disaggregate approach is more suited for proactive policy analysis since it is causal, less tied to the estimation data and more likely to include a range of relevant policy variables. Fourth, the disaggregate approach is more efficient than the aggregate approach in terms of model reliability per unit cost of data collection. Disaggregate data provide substantial variation in the behavior of interest and in the determinants of that behavior, enabling the efficient estimation of model parameters. On the other hand, aggregation leads to considerable loss in variability, thus requiring much more data to obtain the same level of model precision. Finally, disaggregate models, if properly specified, will obtain un-biased parameter estimates, while aggregate model estimates are known to produce biased (*i.e.* incorrect) parameter estimates.

1.2 Use of Disaggregate Discrete Choice Models

The behavioral nature of disaggregate models, and the associated advantages of such models over aggregate models, has led to the widespread use of disaggregate discrete choice methods in travel demand modeling. A few of these application contexts below with references to recent

work in these areas are: travel mode choice (reviewed in detail later), destination choice (Bhat *et al.*, 1998; Train, 1998), route choice (Yai *et al.*, 1998; Cascetta *et al.*, 1997, Erhardt *et al.*, 2004, Gliebe and Koppelman, 2002), air travel choices (Proussaloglou and Koppelman, 1999) activity analysis (Wen and Koppelman, 1999) and auto ownership, brand and model choice (Hensher *et al.*, 1992; Bhat and Pulugurta, 1998). Choice models have also been applied in several other fields such as purchase incidence and brand choice in marketing (Kalyanam and Putler, 1997; Bucklin *et al.*, 1995), housing type and location choice in geography (Waddell, 1993; Evers, 1990; Sermons and Koppelman, 1998), choice of intercity air carrier (Proussaloglou and Koppelman, 1998) and investment choices of finance firms (Corres *et al.*, 1993).

1.3 Application Context in Current Course

In this self-instructing course, we focus on the travel mode choice decision. Within the travel demand modeling field, mode choice is arguably the single most important determinant of the number of vehicles on roadways. The use of high-occupancy vehicle modes (such as ridesharing arrangements and transit) leads to more efficient use of the roadway infrastructure, less traffic congestion, and lower mobile-source emissions as compared to the use of single-occupancy vehicles. Further, the mode choice decision is the most easily influenced travel decision for many trips. There is a vast literature on travel mode choice modeling which has provided a good understanding of factors which influence mode choice and the general range of trade-offs individuals are willing to make among level-of-service variables (such as travel time and travel cost).

The emphasis on travel mode choice in this course is a result of its important policy implications, the extensive literature to guide its development, and the limited number of alternatives involved in this decision (typically, 3 – 7 alternatives). While the methods discussed here are equally applicable to cases with many alternatives, a limited number of mode choice alternatives enable us to focus the course on important concepts and issues in discrete choice modeling without being distracted by the mechanics and presentation complexity associated with larger choice sets.

1.4 Urban and Intercity Travel Mode Choice Modeling

The mode choice decision has been examined both in the context of urban travel as well as intercity travel.

1.4.1 Urban Travel Mode Choice Modeling

Many metropolitan areas are plagued by a continuing increase in traffic congestion resulting in motorist frustration, longer travel times, lost productivity, increased accidents and automobile insurance rates, more fuel consumption, increased freight transportation costs, and deterioration in air quality. Aware of these serious consequences of traffic congestion, metropolitan areas are examining and implementing transportation congestion management (TCM) policies. Urban travel mode choice models are used to evaluate the effectiveness of TCM policies in shifting single-occupancy vehicle users to high-occupancy vehicle modes.

The focus of urban travel mode choice modeling has been on the home-based work trip. All major metropolitan planning organizations estimate home-based work travel mode choice models as part of their transportation planning process. Most of these models include only motorized modes, though increasingly non-motorized modes (walk and bike) are being included (Lawton, 1989; Purvis, 1997).

The modeling of home-based non-work trips and non-home-based trips has received less attention in the urban travel mode choice literature. However, the increasing number of these trips and their contribution to traffic congestion has recently led to more extensive development of models for these trip purposes in some metropolitan regions (for example, see Iglesias, 1997; Marshall and Ballard, 1998).

In this course, we discuss model-building and specification issues for home-based work and home-based shop/other trips within an urban context, though the same concepts can be immediately extended to other trip purposes and locales.

1.4.2 Intercity Mode Choice Models

Increasing congestion on intercity highways and at intercity air terminals has raised serious concerns about the adverse impacts of such congestion on regional economic development,

national productivity and competitiveness, and environmental quality. To alleviate current and projected congestion, attention has been directed toward identifying and evaluating alternative proposals to improve intercity transportation services. These proposals include expanding or constructing new express roadways and airports, upgrading conventional rail services and providing new high-speed ground transportation services using advanced technologies. Among other things, the *a priori* evaluation of such large scale projects requires the estimation of reliable intercity mode choice models to predict ridership share on the proposed new or improved intercity service and identify the modes from which existing intercity travelers will be diverted to the new (or improved) service.

Intercity travel mode choice models are usually segmented by purpose (business versus pleasure), day of travel (weekday versus weekend), party size (traveling individually versus group travel), *etc.* The travel modes in such models typically include car, rail, air, and bus modes (Koppelman and Wen, 1998; Bhat, 1998; and KPMG Peat Marwick *et al.*, 1993).

This manual examines issues of urban model choice; however, the vast majority of approaches and specifications can and have been used in intercity mode choice modeling.

1.5 Description of the Course

This self-instructing course (SIC) is designed for readers who have some familiarity with transportation planning methods and background in travel model estimation. It updates and extends the previous SIC Manual (Horowitz *et al.*, 1986) in a number of important ways. First, it is more rigorous in the mathematical details reflecting increased awareness and application of discrete choice models over the past decade. The course is intended to enhance the understanding of model structure and estimation procedures more so than it is intended to introduce discrete choice modeling (readers with no background in discrete choice modeling may want to work first with the earlier SIC). Second, this SIC emphasizes "hands-on" estimation experience using data sets obtained from planning and decision-oriented surveys. Consequently, there is more emphasis on data structure and more extensive examination of model specification issues. Various software packages available for discrete choice modeling

estimation are described briefly with the intent of providing a broad overview of their capabilities. The descriptions and examples of the command structure and output for selected models are included in Appendix A to illustrate key differences among them. Further, example command and output files for models using a module developed for Matlab (an engineering software package), as well as the module's code, are included on the accompanying CD and documented in Appendix B. Third, this SIC extends the range of travel modes to include non-motorized modes and discusses issues involved in including such modes in the analysis. Fourth, this SIC includes detailed coverage of the nested logit model which is being used more commonly in many metropolitan planning organizations today.

1.6 Organization of Course Structure

This course manual is divided into twelve chapters or modules. CHAPTER 1, this chapter, provides an introduction to the course. CHAPTER 2 describes the elements of the choice process including the decision maker, the alternatives, the attributes of the alternative, and the decision rule(s) adopted by the decision maker in making his/her choice. CHAPTER 3 introduces the basic concepts of utility theory followed by a discussion of probabilistic and deterministic choice concepts and the technical components of the utility function.

CHAPTER 4 describes the Multinomial Logit (MNL) Model in detail. The discussion includes the functional form of the model, its mathematical properties, and the practical implications of these properties in model development and application. The chapter concludes with an overview of methods used for estimating the model parameters.

In CHAPTER 5, we first discuss the data requirements for developing disaggregate mode choice models, the potential sources for these data, and the format in which these data need to be organized for estimation. Next, the data sets used in this manual, *i.e.*, the San Francisco Bay Area 1990 work trip mode choice (for urban area journey to work travel) and the San Francisco Bay Area Shop/Other 1990 mode choice data (for non-work travel), are described. This is followed by the development of a basic work mode choice model specification. The estimation results of this model specification are reviewed with a comprehensive discussion of informal and

formal tests to evaluate the appropriateness of model parameters and the overall goodness-of-fit statistics of the model.

CHAPTER 6 describes and demonstrates the process by which the utility function specification for the work mode choice model can be refined using intuition, statistical analysis, testing, and judgment. Many specifications of the utility function are explored for both data sets to demonstrate some of the most common specification forms and testing methods. Starting from a base model, incremental changes are made to the modal utility functions with the objective of finding a model specification that performs better statistically, and is consistent with theory and our *a priori* expectations about mode choice behavior. The appropriateness of each specification change is evaluated using judgment and statistical tests. This process leads to a preferred specification for the work mode choice MNL model.

CHAPTER 7 parallels CHAPTER 6 for the shop/other mode choice model.

CHAPTER 8 introduces the Nested Logit (NL) Model. The Chapter begins with the motivation for the NL model to address one of the major limitations of the MNL. The functional form and the mathematical properties of the NL are discussed in detail. This is followed by a presentation of estimation results for a number of NL model structures for the work and shop/other data sets. Based on these estimation results, statistical tests are used to compare the various NL model structures with the corresponding MNL.

CHAPTER 9 describes the issues involved in formulating, estimating¹ and selecting a preferred NL model. The results of statistical tests are used in conjunction with our *a priori* understanding of the competitive structure among different alternatives to select a final preferred nesting structure. The practical implications of choosing this preferred nesting structure in comparison to the MNL model are discussed.

CHAPTER 11 describes how models estimated from disaggregate data can be used to predict a aggregate mode choice for a group of individuals from relevant information regarding the altered value (due to socio-demographic changes or policy actions) of exogenous variables.

¹ Estimation of NL models includes the problem of searching across multiple optima and convergence difficulties that can arise.

The chapter also discusses issues related to the aggregate assessment of the performance of mode choice models and the application of the models to evaluate policy actions.

CHAPTER 12 provides an overview of the motivation for and structure of advanced discrete choice models. The discussion is intended to familiarize readers with a variety of models that allow increased flexibility in the representation of the choice behavior than those allowed by the multinomial logit and nested logit models. It does not provide the detailed mathematical formulations or the estimation techniques for these advanced models. Appropriate references are provided for readers interested in this information.

CHAPTER 2: Elements of the Choice Decision Process

2.1 Introduction

We observe individuals (or decision makers) making choices in a wide variety of decision contexts. However, we generally do not have information about the process individuals use to arrive at their observed choice. A proposed framework for the choice process is that an individual first determines the available alternatives; next, evaluates the attributes of each alternative relevant to the choice under consideration; and then, uses a decision rule to select an alternative from among the available alternatives (Ben-Akiva and Lerman, 1985, Chapter 3). Some individuals might select a particular alternative without going through the structured process presented above. For example, an individual might decide to buy a car of the same make and model as a friend because the friend is happy with the car or is a car expert. Or an individual might purchase the same brand of ice cream out of habit. However, even in these cases, one can view the behavior within the framework of a structured decision process by assuming that the individual generates only one alternative for consideration (which is also the one chosen).

In the subsequent sections, we discuss four elements associated with the choice process; the decision maker, the alternatives, the attributes of alternatives and the decision rule.

2.2 The Decision Maker

The decision maker in each choice situation is the individual, group or institution which has the responsibility to make the decision at hand. The decision maker will depend on the specific choice situation. For example, the decision maker will be the individual in college choice, career choice, travel mode choice, *etc.*; the household in residential location choice, vacation destination choice, number of cars owned, *etc.*; the firm in office or warehouse location, carrier choice, employee hiring, *etc.* or the State (in the selection of roadway alignments). A common characteristic in the study of choice is that different decision makers face different choice situations and can have different tastes (that is, they value attributes differently). For example, in

travel mode choice modeling, two individuals with different income levels and different residential locations are likely to have different sets of modes to choose from and may place different importance weights on travel time, travel cost and other attributes. These differences among decision makers should be explicitly considered in choice modeling; consequently, it is important to develop choice models at the level of the decision maker and to include variables which represent differences among the decision makers.

2.3 The Alternatives

Individuals make a choice from a set of alternatives available to them. The set of available alternatives may be constrained by the environment. For example, high speed rail between two cities is an alternative only if the two cities are connected by high speed rail. The choice set determined by the environment is referred to as the universal choice set. However, even if an alternative is present in the universal choice set, it may not be feasible for a particular individual. Feasibility of an alternative for an individual in the context of travel mode choice may be determined by legal regulations (a person cannot drive alone until the age of 16), economic constraints (limousine service is not feasible for some people) or characteristics of the individual (no car available or a handicap that prevents one from driving). The subset of the universal choice set that is feasible for an individual is defined as the feasible choice set for that individual. Finally, not all alternatives in the feasible choice set may be considered by an individual in her/his choice process. For example, transit might be a feasible travel mode for an individual's work trip, but the individual might not be aware of the availability or schedule of the transit service. The subset of the feasible choice set that an individual actually considers is referred to as the consideration choice set. This is the choice set which should be considered when modeling choice decisions.

The choice set may also be determined by the decision context of the individual or the focus of the policy makers supporting the study. For example, a study of university choice may focus on choice of school type (private vs. public, small vs. large, urban vs. suburban or rural

location, etc.), if the perspective is national, or a choice of specific schools, if the perspective is regional.

2.4 Attributes of Alternatives

The alternatives in a choice process are characterized by a set of attribute values. Following Lancaster (1971), one can postulate that the attractiveness of an alternative is determined by the value of its attributes. The measure of uncertainty about an attribute can also be included as part of the attribute vector in addition to the attribute itself. For example, if travel time by transit is not fixed, the expected value of transit travel time and a measure of uncertainty of the transit travel time can both be included as attributes of transit.

The attributes of alternatives may be generic (that is, they apply to all alternatives equally) or alternative-specific (they apply to one or a subset of alternatives). In the travel mode choice context, in-vehicle-time is usually considered to be specific to all motorized modes because it is relevant to motorized alternatives. However, if travel time by bus is considered to be very onerous due to over-crowding, bus in-vehicle-time may be defined as a distinct variable with a distinct parameter; differences between this parameter and the in-vehicle-time parameter for other motorized modes will measure the degree to which bus time is considered onerous to the traveler relative to other in-vehicle time. Other times, such as wait time at a transit stop or transfer time at a transit transfer point are relevant only to the transit modes, not for the non-transit modes. It is also common to consider the travel times for non-motorized modes (bike and walk) as specific to only these alternatives.

An important reason for developing discrete choice models is to evaluate the effect of policy actions. To provide this capability, it is important to identify and include attributes whose values may be changed through *pro-active* policy decisions. In a travel mode choice context, these variables include measures of service (travel time, frequency, reliability of service, *etc.*) and travel cost.

2.5 The Decision Rule

An individual invokes a decision rule (*i.e.*, a mechanism to process information and evaluate alternatives) to select an alternative from a choice set with two or more alternatives. This decision rule may include random choice, variety seeking, or other processes which we refer to as being irrational. As indicated earlier, some individuals might use other decision rules such as "follow the leader" or habit in choosing alternatives which may also be considered to be irrational. However, even in this case, rational discrete choice models may be effective if the decision maker who adopts habitual behavior previously evaluated different alternatives and selected the best one for him/her and there have been no intervening changes in her/his alternatives and preferences. However, in the case of follow-the-leader behavior, the decision maker is considered to be rational if the "leader" is believed to share a similar value system. An individual is said to use a rational decision process if the process satisfies two fundamental constructs: consistency and transitivity. Consistency implies the same choice selection in repeated choices under identical circumstances. Transitivity implies an unique ordering of alternatives on a preference scale. Therefore, if alternative A is preferred to alternative B and alternative B is preferred to alternative C, then alternative A is preferred to alternative C.

A number of possible rules fall under the purview of rational decision processes (Ben-Akiva and Lerman, 1985; Chapter 3). In this course, the focus will be on one such decision rule referred to as utility maximization. The utility maximization rule is based on two fundamental concepts. The first is that the attribute vector characterizing each alternative can be reduced to a scalar utility value for that alternative. This concept implies a compensatory decision process; that is, it presumes that individuals make "trade-offs" among the attributes characterizing alternatives in determining their choice. Thus, an individual may choose a costlier travel mode if the travel time reduction offered by that mode compensates for the increased cost. The second concept is that the individual selects the alternative with the highest utility value.

The focus on utility maximization in this course is based on its strong theoretical background, extensive use in the development of human decision making concepts, and amenability to statistical testing of the effects of attributes on choice. The utility maximization

rule is also robust; that is, it provides a good description of the choice behavior even in cases where individuals use somewhat different decision rules.

In the next chapter, we discuss the concepts and underlying principles of utility based choice theory in more detail.

CHAPTER 3: Utility-Based Choice Theory

3.1 Basic Construct of Utility Theory

Utility is an indicator of value to an individual. Generally, we think about utility as being derived from the attributes of alternatives or sets of alternatives; *e.g.*, the total set of groceries purchased in a week. The utility maximization rule states that an individual will select the alternative from his/her set of available alternatives that maximizes his or her utility. Further, the rule implies that there is a function containing attributes of alternatives and characteristics of individuals that describes an individual's utility valuation for each alternative. The utility function, U , has the property that an alternative is chosen if its utility is greater than the utility of all other alternatives in the individual's choice set. Alternatively, this can be stated as alternative, ' i ', is chosen among a set of alternatives, if and only if the utility of alternative, ' i ', is greater than or equal to the utility of all alternatives², ' j ', in the choice set, C . This can be expressed mathematically as:

$$\text{If } U(X_i, S_t) \geq U(X_j, S_t) \quad \forall j \quad \Rightarrow \quad i \succ j \quad \forall j \in C \quad 3.1$$

where $U(\quad)$ is the mathematical utility function,

X_i, X_j are vectors of attributes describing alternatives i and j , respectively
(*e.g.*, travel time, travel cost, and other relevant attributes of the available modes),

S_t is a vector of characteristics describing individual t , that influence his/her preferences among alternatives (*e.g.*, household income and number of automobiles owned for travel mode choice),

² "All j includes alternative i . The case of equality of utility is included to acknowledge that the utility of i will be equal to the utility of i included in all j .

$i \succ j$ means the alternative to the left is preferred to the alternative to the right, and

$\forall j$ means all the cases, j , in the choice set.

That is, if the utility of alternative i is greater than or equal to the utility of all alternatives, j ; alternative i will be preferred and chosen from the set of alternatives, C .

The underlying concept of utility allows us to rank a series of alternatives and identify the single alternative that has highest utility. The primary implication of this ranking or ordering of alternatives is that there is no absolute reference or zero point, for utility values. Thus, the only valuation that is important is the difference in utility between pairs of alternatives; particularly whether that difference is positive or negative. Any function that produces the same preference orderings can serve as a utility function and will give the same predictions of choice, regardless of the numerical values of the utilities assigned to individual alternatives. It also follows that utility functions, which result in the same order among alternatives, are equivalent.

3.2 Deterministic Choice Concepts

The utility maximization rule, which states that an individual chooses the alternative with the highest utility, implies no uncertainty in the individual's decision process; that is, the individual is certain to choose the highest ranked alternative under the observed choice conditions. Utility models that yield certain predictions of choice are called deterministic utility models. The application of deterministic utility to the case of a decision between two alternatives is illustrated in Figure 3.1 that portrays a utility space in which the utilities of alternatives 1 and 2 are plotted along the horizontal and vertical axes, respectively, for each individual. The 45° line represents those points for which the utilities of the two alternatives are equal. Individuals B, C and D (above the equal-utility line) have higher utility for alternative 2 than for alternative 1 and are certain to choose alternative 2. Similarly, individuals A, E, and F (below the line) have higher utility for alternative 1 and are certain to choose that alternative. If deterministic utility models described behavior correctly, we would expect that an individual would make the same choice over time and that similar individuals (individuals having the same individual and household

characteristics), would make the same choices when faced with the same set of alternatives. In practice, however, we observe variations in an individual's choice and different choices among apparently similar individuals when faced with similar or even identical alternatives. For example, in studies of work trip mode choice, it is commonly observed that individuals, who are represented as having identical personal characteristics and who face the same sets of travel alternatives, choose different modes of travel to work. Further, some of these individuals vary their choices from day to day for no observable reason resulting in observed choices which appear to contradict the utility evaluations; that is, person A may choose Alternative 2 even though $U_1 > U_2$ or person C may choose Alternative 2 even though $U_2 > U_1$. These observations raise questions about the appropriateness of deterministic utility models for modeling travel or other human behavior. The challenge is to develop a model structure that provides a reasonable representation of these unexplained variations in travel behavior.

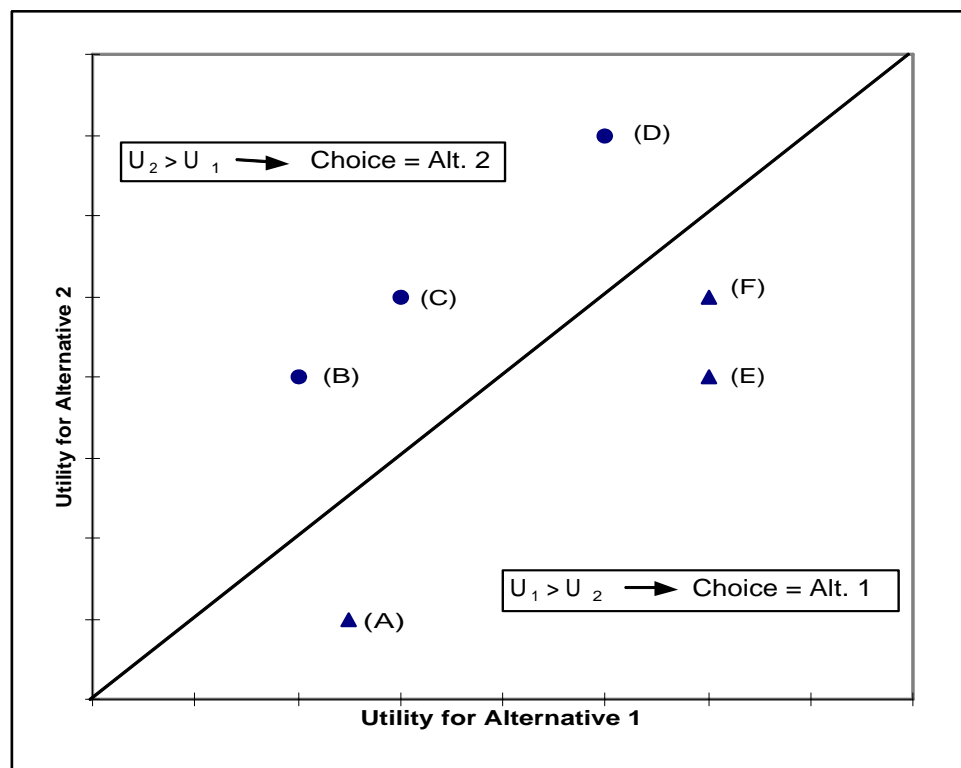


Figure 3.1 Illustration of Deterministic Choice

There are three primary sources of error in the use of deterministic utility functions. First, the individual may have incomplete or incorrect information or misperceptions about the attributes of some or all of the alternatives. As a result, different individuals, each with different information or perceptions about the same alternatives are likely to make different choices. Second, the analyst or observer has different or incomplete information about the same attributes relative to the individuals and an inadequate understanding of the function the individual uses to evaluate the utility of each alternative. For example, the analyst may not have good measures of the reliability of a particular transit service, the likelihood of getting a seat at a particular time of day or the likelihood of finding a parking space at a suburban rail station. However, the traveler, especially if he/she is a regular user, is likely to know these things or to have opinions about them. Third, the analyst is unlikely to know, or account for, specific circumstances of the individual's travel decision. For example, an individual's choice of mode for the work trip may depend on whether there are family visitors or that another family member has a special travel need on a particular day. Using models which do not account for and incorporate this lack of information results in the apparent behavioral inconsistencies described above. While human behavior may be argued to be inconsistent, it can also be argued that the inconsistency is only apparent and can be attributed to the analyst's lack of knowledge regarding the individual's decision making process. Models that take account of this lack of information on the part of the analyst are called random utility or probabilistic choice models.

3.3 Probabilistic Choice Theory

If analysts thoroughly understood all aspects of the internal decision making process of choosers as well as their perception of alternatives, they would be able to describe that process and predict mode choice using deterministic utility models. Experience has shown, however, that analysts do not have such knowledge; they do not fully understand the decision process of each individual or their perceptions of alternatives and they have no realistic possibility of obtaining this information. Therefore, mode choice models should take a form that recognizes and

accommodates the analyst's lack of information and understanding. The data and models used by analysts describe preferences and choice in terms of probabilities of choosing each alternative rather than predicting that an individual will choose a particular mode with certainty. Effectively, these probabilities reflect the population probabilities that people with the given set of characteristics and facing the same set of alternatives choose each of the alternatives.

As with deterministic choice theory, the individual is assumed to choose an alternative if its utility is greater than that of any other alternative. The probability prediction of the analyst results from differences between the estimated utility values and the utility values used by the traveler. We represent this difference by decomposing the utility of the alternative, from the perspective of the decision maker, into two components. One component of the utility function represents the portion of the utility observed by the analyst, often called the deterministic (or observable) portion of the utility. The other component is the difference between the unknown utility used by the individual and the utility estimated by the analyst. Since the utility used by the decision maker is unknown, we represent this difference as a random error. Formally, we represent this by:

$$U_{it} = V_{it} + \varepsilon_{it} \quad 3.2$$

where U_{it} is the true utility of the alternative i to the decision maker t , (U_{it} is equivalent to $U(X_i, S_t)$ but provides a simpler notation),
 V_{it} is the deterministic or observable portion of the utility estimated by the analyst, and
 ε_{it} is the error or the portion of the utility unknown to the analyst.

The analyst does not have any information about the error term. However, the total error which is the sum of errors from many sources (imperfect information, measurement errors, omission of modal attributes, omission of the characteristics of the individual that influence his/her choice decision and/or errors in the utility function) is represented by a random variable. Different assumptions about the distribution of the random variables associated with the utility of each

alternative result in different representations of the model used to describe and predict choice probabilities. The assumptions used in the development of logit type models are discussed in the next Chapter.

3.4 Components of the Deterministic Portion of the Utility Function

The deterministic or observable portion (often called the systematic portion) of the utility of an alternative is a mathematical function of the attributes of the alternative and the characteristics of the decision maker. The systematic portion of utility can have any mathematical form but the function is most generally formulated as additive to simplify the estimation process. This function includes unknown parameters which are estimated in the modeling process. The systematic portion of the utility function can be broken into components that are (1) exclusively related to the attributes of alternatives, (2) exclusively related to the characteristics of the decision maker and (3) represent interactions between the attributes of alternatives and the characteristics of the decision maker. Thus, the systematic portion of utility can be represented by:

$$V_{t,i} = V(S_t) + V(X_i) + V(S_t, X_i) \quad 3.3$$

where V_{it} is the systematic portion of utility of alternative i for individual t ,
 $V(S_t)$ is the portion of utility associated with characteristics of individual t ,
 $V(X_i)$ is the portion of utility of alternative i associated with the attributes of alternative i , and
 $V(S_t, X_i)$ is the portion of the utility which results from interactions between the attributes of alternative i and the characteristics of individual t .

Each of these utility components is discussed separately.

3.4.1 Utility Associated with the Attributes of Alternatives

The utility component associated exclusively with alternatives includes variables that describe the attributes of alternatives. These attributes influence the utility of each alternative for all people in the population of interest. The attributes considered for inclusion in this component are service attributes which are measurable and which are expected to influence people's preferences/choices among alternatives. These include measures of travel time, travel cost, walk access distance, transfers required, crowding, seat availability, and others. For example:

- Total travel time,
- In-vehicle travel time,
- Out-of-vehicle travel time,
- Travel cost,
- Number of transfers (transit modes),
- Walk distance and
- Reliability of on time arrival.

These measures differ across alternatives for the same individual and also among individuals due to differences in the origin and destination locations of each person's travel. For example, this portion of the utility function could look like:

$$V(X_i) = \gamma_1 \times X_{i1} + \gamma_2 \times X_{i2} + \cdots + \gamma_k \times X_{ik} \quad 3.4$$

where γ_k is the parameter which defines the direction and importance of the effect of attribute k on the utility of an alternative and

X_{ik} is the value of attribute k for alternative i .

Thus, this portion of the utility of each alternative, i , is the weighted sum of the attributes of alternative i . A specific example for the Drive Alone (DA), Shared Ride (SR), and Transit (TR) alternatives is:

$$V(X_{DA}) = \gamma_1 \times TT_{DA} + \gamma_2 \times TC_{DA} \quad 3.5$$

$$V(X_{SR}) = \gamma_1 \times TT_{SR} + \gamma_2 \times TC_{SR} \quad 3.6$$

$$V(X_{TR}) = \gamma_1 \times TT_{TR} + \gamma_2 \times TC_{TR} + \gamma_3 \times FREQ_{TR} \quad 3.7$$

where TT_i is the travel time for mode i ($i = DA, SR, TR$) and

TC_i is the travel cost for mode i , and

$FREQ_{TR}$ is the frequency for transit services.

Travel time and travel cost are generic; that is, they apply to all alternatives; frequency is specific to transit only. The parameters, γ_k , are identical for all the alternatives to which they apply.

This implies that the utility value of travel time and travel cost are identical across alternatives.

The possibility that travel time may be more onerous on Transit than by Drive Alone or Shared Ride could be tested by reformulating the above models to:

$$V(X_{DA}) = \gamma_{11} \times TT_{DA} + \gamma_{12} \times 0 + \gamma_2 \times TC_{DA} \quad 3.8$$

$$V(X_{SR}) = \gamma_{11} \times TT_{SR} + \gamma_{12} \times 0 + \gamma_2 \times TC_{SR} \quad 3.9$$

$$V(X_{TR}) = \gamma_{11} \times 0 + \gamma_{12} \times TT_{TR} + \gamma_2 \times TC_{TR} + \gamma_3 \times FREQ_{TR} \quad 3.10$$

That is, two distinct parameters would be estimated for travel time; one for travel time by DA and SR, γ_{11} , and the other for travel time by TR, γ_{12} . These parameters could be compared to determine if the differences are statistically significant or large enough to be important.

3.4.2 Utility ‘Biases’ Due to Excluded Variables

It has been widely observed that decision makers exhibit preferences for alternatives which cannot be explained by the observed attributes of those alternatives. These preferences are described as alternative specific preference or bias; they measure the average preference of individuals with different characteristics for an alternative relative to a ‘reference’ alternative. As will be shown in CHAPTER 4, the selection of the reference alternative does not influence the interpretation of the model estimation results. In the simplest case, we assume that the bias is the same for all decision makers. In this case, this portion of the utility function would be:

$$Bias_i = \beta_{i0} \times ASC_i \quad 3.11$$

where β_{i0} represents an increase in the utility of alternative i for all choosers and

ASC_i is equal to one for alternative i and zero for all other alternatives.

More detailed observation generally indicates that people with different personal and family characteristics have different preferences among sets of alternatives. For example, members of high income households are less likely to choose transit alternatives than low income individuals, all other things being equal. Similarly, members of households with fewer automobiles than workers are more likely to choose transit alternatives. Thus, it is useful to consider that the bias may differ across individuals as discussed in the next section.

3.4.3 Utility Related to the Characteristics of the Decision Maker

The differences in ‘bias’ across individuals can be represented by incorporating personal and household variables in mode choice models. Variables commonly used for this purpose include:

- Income of the traveler’s household,
- Sex of traveler,
- Age of traveler,
- Number of automobiles in traveler’s household,
- Number of workers in the traveler’s household, and
- Number of adults in the traveler’s household.

In some cases, these variables are combined. For example, the number of automobiles may be divided by the number of workers to indicate the availability of automobiles to each household member. This approach results in modification of the bias portion of the utility function to look like:

$$\beta_{i0} \times ASC_i + \beta_{i1} \times S_{1t} + \beta_{i2} \times S_{2t} + \cdots + \beta_{iM} \times S_{Mt} \quad 3.12$$

where β_{im} is the parameter which defines the direction and magnitude of the incremental bias due to an increase in the m^{th} characteristic of the decision maker ($m = 0$ represents the parameter associated with the alternative specific constant) and

S_{mt} is the value of the m^{th} characteristics for individual t .

For the case of three alternatives, (Drive Alone, Shared Ride and Transit), the decision maker components of the utility functions are:

$$V(S_{DA}) = \beta_{DA,0} \times 1 + \beta_{DA,1} \times Inc_t + \beta_{DA,2} \times NCar_t \quad 3.13$$

$$V(S_{SR}) = \beta_{SR,0} \times 1 + \beta_{SR,1} \times Inc_t + \beta_{SR,2} \times NCar_t \quad 3.14$$

$$V(S_{TR}) = \beta_{TR,0} \times 1 + \beta_{TR,1} \times Inc_t + \beta_{TR,2} \times NCar_t \quad 3.15$$

where β_{i0} is the modal bias constant for mode i ($i = DA, SR, TR$),

Inc_t is the household income of the traveler,

$NCar_t$ is the number of cars in the traveler's household, and

β_{i1}, β_{i2} are mode specific parameters on income and cars, respectively, for mode i ($i = DA, SR, TR$)

It is important to recognize that the parameters that describe the effect of traveler characteristics differ across alternatives while the variables are identical across alternatives for each individual.

3.4.4 Utility Defined by Interactions between Alternative Attributes and Decision Maker Characteristics

The final component of utility takes into account differences in how attributes are evaluated by different decision makers. One effect of income, described earlier, is to increase the preference for travel by private automobile. This representation implies that the preference for DA and SR increase with income, independent of the attributes of travel by each alternative. Another way to represent the influence of income is that increasing income reduces the importance of monetary cost in the evaluation of modal alternatives. The idea that high income travelers place less importance on cost can be represented by dividing the cost of travel of an alternative by annual income or some function of annual income of the traveler or his/her household. Another interaction effect might be that different travelers evaluate travel time differently. For example, one might argue that because women commonly take increased responsibility for home maintenance and child care, they are likely to evaluate increased travel time to work more

negatively than men. This could be represented by adding a variable to the model which represents the product of a dummy variable for female (one if the traveler is a woman and zero otherwise) times travel time, as illustrated below in the utility equations for a three alternative mode choice example (Drive Alone, Shared Ride, and Transit) using the same notation described previously.

$$V(X_{DA}, S_i) = \gamma_1 \times TT_{DA} + \gamma_2 \times TT_{DA} \times Fem + \gamma_3 \times TC_{DA} \quad 3.16$$

$$V(X_{SR}, S_i) = \gamma_1 \times TT_{SR} + \gamma_2 \times TT_{SR} \times Fem + \gamma_3 \times TC_{SR} \quad 3.17$$

$$V(X_{TR}, S_i) = \gamma_1 \times TT_{TR} + \gamma_2 \times TT_{TR} \times Fem + \gamma_3 \times TC_{TR} + \gamma_4 \times FREQ_{TR} \quad 3.18$$

In this example, γ_1 represents the utility value of one minute of travel time to men and γ_2 represents the additional utility value of one minute of travel time to women. Thus, the total utility value of one minute of travel time to women is $\gamma_1 + \gamma_2$. In this case, γ_1 , is expected to be negative indicating that increased travel time reduces the utility of an alternative. γ_2 may be negative or positive, indicating that women are more or less sensitive to increases in travel time.

3.5 Specification of the Additive Error Term

As described in section 3.3, the utility of each alternative is represented by a deterministic component, which is represented in the utility function by observed and measured variables, and an additive error term, ε_i which represents those components of the utility function which are not included in the model. In the three alternative examples used above, the total utility of each alternative can be represented by:

$$U_{DA} = V(S_t) + V(X_{DA}) + V(S_t, X_{DA}) + \varepsilon_{DA} \quad 3.19$$

$$U_{SR} = V(S_t) + V(X_{SR}) + V(S_t, X_{SR}) + \varepsilon_{SR} \quad 3.20$$

$$U_{TR} = V(S_t) + V(X_{TR}) + V(S_t, X_{TR}) + \varepsilon_{TR} \quad 3.21$$

where $V()$ represents the deterministic components of the utility for the alternatives, and

ε_i represents the random components of the utility, also called the error term.

The error term is included in the utility function to account for the fact that the analyst is not able to completely and correctly measure or specify all attributes that determine travelers' mode utility assessment. By definition, error terms are unobserved and unmeasured. A wide range of distributions could be used to represent the distribution of error terms over individuals and alternatives. If we assume that the error term for each alternative represents many missing components, each of which has relatively little impact on the value of each alternative, the central limit theorem suggests that the sum of these small errors will be distributed normally. This assumption leads to the formulation of the Multinomial Probit (MNP) probabilistic choice model. However, the mathematical complexity of the MNP model; which makes it difficult to estimate, interpret and predict; has limited its use in practice. An alternative distribution assumption, described in the next chapter, leads to the formulation of the multinomial logit (MNL) model.

CHAPTER 4: The Multinomial Logit Model

4.1 Overview Description and Functional Form

The mathematical form of a discrete choice model is determined by the assumptions made regarding the error components of the utility function for each alternative as described in section 3.5. The specific assumptions that lead to the Multinomial Logit Model are (1) the error components are extreme-value (or Gumbel) distributed, (2) the error components are identically and independently distributed across alternatives, and (3) the error components are identically and independently distributed across observations/individuals. We discuss each of these assumptions below.

The most common assumption for error distributions in the statistical and modeling literature is that errors are distributed normally. There are good theoretical and practical reasons for using the normal distribution for many modeling applications. However, in the case of choice models the normal distribution assumption for error terms leads to the Multinomial Probit Model (MNP) which has some properties that make it difficult to use in choice analysis³. The Gumbel distribution is selected because it has computational advantages in a context where maximization is important, closely approximates the normal distribution (see Figure 4.1 and Figure 4.2) and produces a closed-form⁴ probabilistic choice model.

³ These include numerical problems, because the MNP can only be calculated using multi-dimensional integration, and problems of interpretation. A special case of the MNP, when the error terms are distributed independently (no covariance) and identically (same variance), obtains estimation and prediction results that are very similar to those for the MNL model.

⁴ A model for which the probability can be calculated without use of numerical integration or simulation methods.

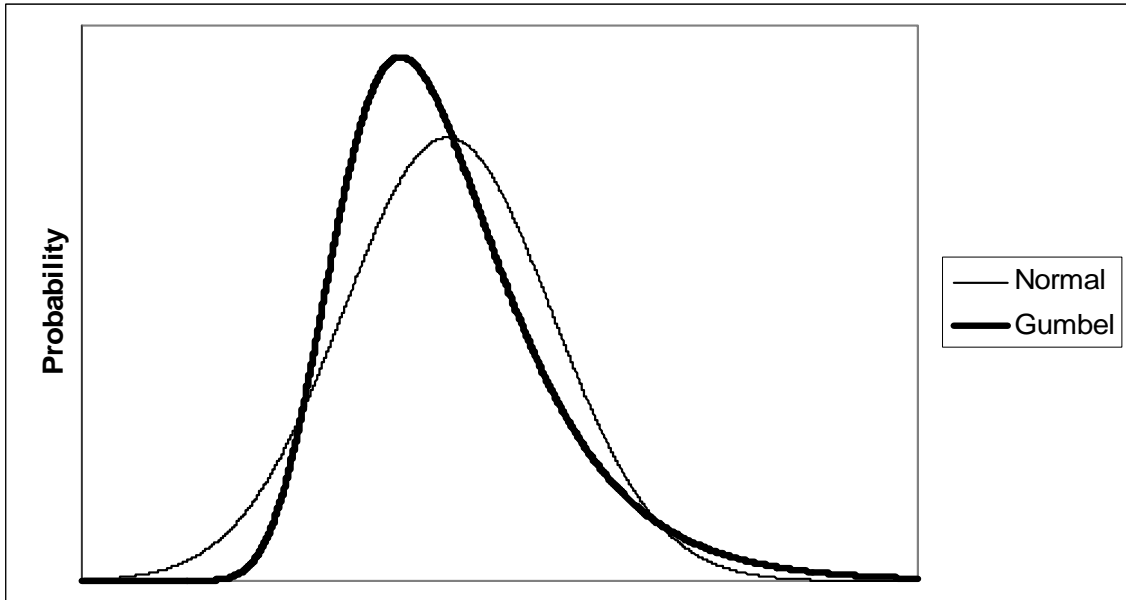


Figure 4.1 Probability Density Function for Gumbel and Normal Distributions (same mean and variance)

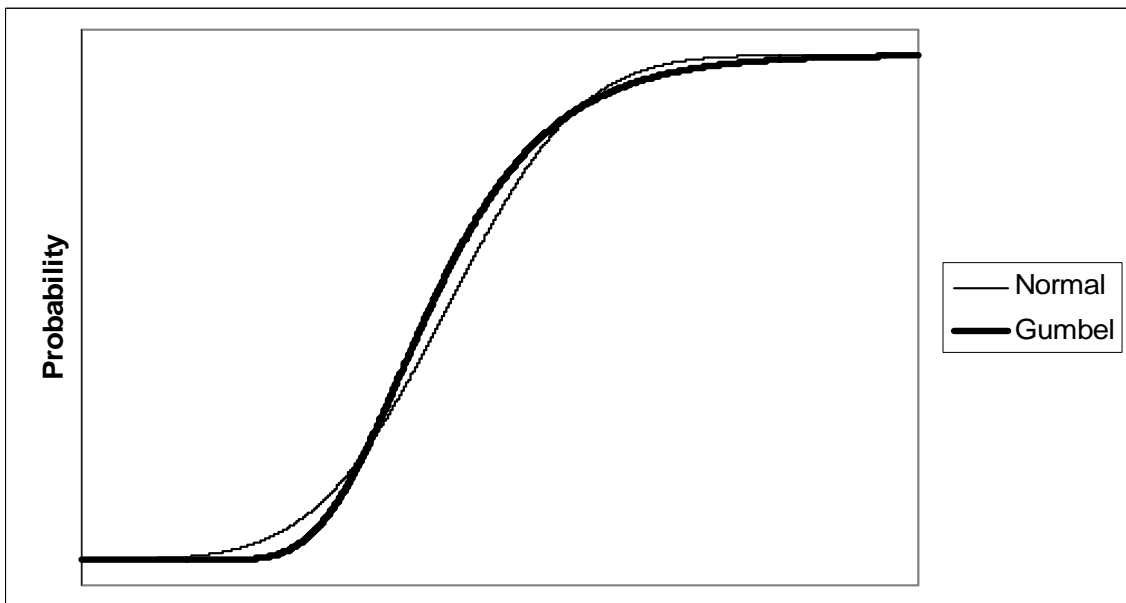


Figure 4.2 Cumulative Distribution Function for Gumbel and Normal Distribution with the Same Mean and Variance

The Gumbel has the following cumulative distribution and probability density functions:

$$F(\epsilon) = \exp\{-\exp[-\mu(\epsilon - \eta)]\} \quad 4.1$$

$$f(\epsilon) = \mu \times \{\exp[-\mu(\epsilon - \eta)]\} \times \exp\{-\exp[-\mu(\epsilon - \eta)]\} \quad 4.2$$

where μ is the scale parameter which determines the variance of the distribution and

η is the location (mode) parameter.

The mean and variance of the distribution are:

$$\text{Mean} = \eta + \frac{0.577}{\mu} \quad 4.3$$

$$\text{Variance} = \frac{\pi^2}{6\mu^2} \quad 4.4$$

The second and third assumptions state the location and variance of the distribution just as μ and σ^2 indicate the location and variance of the normal distribution. We will return to the discussion of the independence between/among alternatives in CHAPTER 8.

The three assumptions, taken together, lead to the mathematical structure known as the Multinomial Logit Model (MNL), which gives the choice probabilities of each alternative as a function of the systematic portion of the utility of all the alternatives. The general expression for the probability of choosing an alternative 'i' ($i = 1, 2, \dots, J$) from a set of J alternatives is:

$$\text{Pr}(i) = \frac{\exp(V_i)}{\sum_{j=1}^J \exp(V_j)} \quad 4.5$$

where $\text{Pr}(i)$ is the probability of the decision-maker choosing alternative i and

V_j is the systematic component of the utility of alternative j .

The exponential function is described in Figure 4.3 which shows the relationship between $\exp(V_i)$ and V_i . Note that $\exp(V_i)$ is always positive and increases monotonically with V_i .

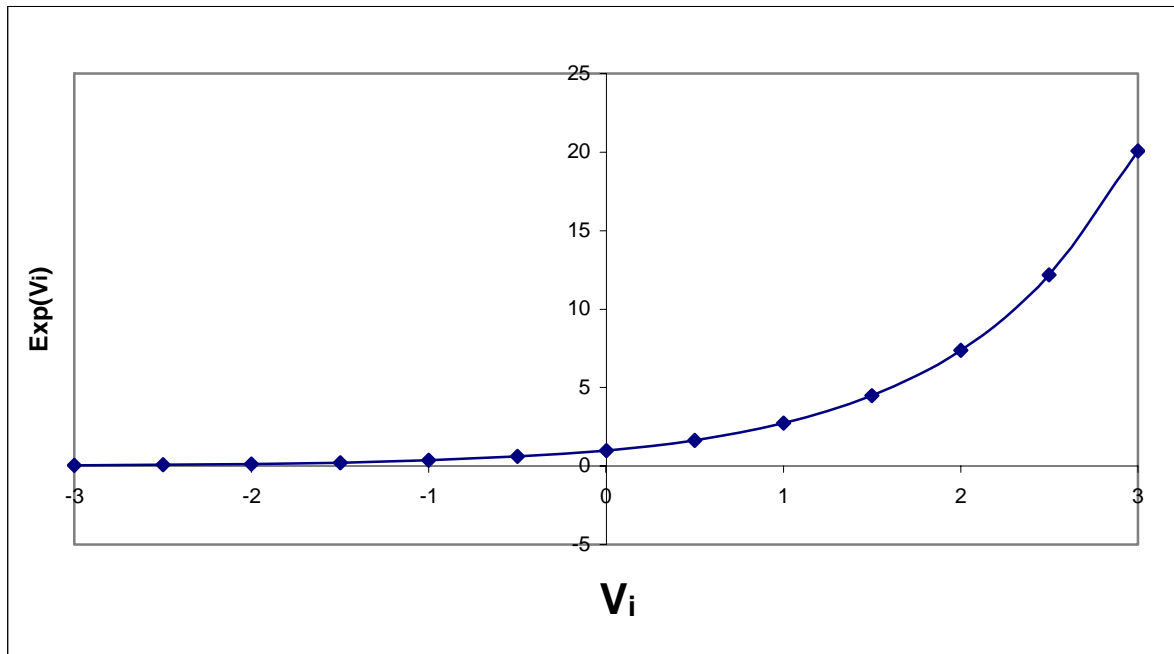


Figure 4.3 Relationship between V_i and $\exp(V_i)$

The multinomial logit (MNL) model has several important properties. We illustrate these for a case in which the decision maker has three available alternatives: Drive Alone (DA), Shared Ride (SR), and TRansit (TR). The probabilities of each alternative are given by modifying equation 4.5 for each alternative to obtain:

$$\Pr(DA) = \frac{\exp(V_{DA})}{\exp(V_{DA}) + \exp(V_{SR}) + \exp(V_{TR})} \quad 4.6$$

$$\Pr(SR) = \frac{\exp(V_{SR})}{\exp(V_{DA}) + \exp(V_{SR}) + \exp(V_{TR})} \quad 4.7$$

$$\Pr(TR) = \frac{\exp(V_{TR})}{\exp(V_{DA}) + \exp(V_{SR}) + \exp(V_{TR})} \quad 4.8$$

where $\Pr(DA)$, $\Pr(SR)$, and $\Pr(TR)$ are the probabilities of the decision-maker choosing drive alone, shared ride and transit, respectively, and V_{DA} , V_{SR} and V_{TR} are the systematic components of the utility for drive alone, shared ride, and transit alternatives, respectively. It is common to replace these three equations by a single general equation to represent the probability of any alternative and to simplify the equation by replacing the explicit summation in the denominator by the summation over alternatives as:

$$\Pr(i) = \frac{\exp(V_i)}{\exp(V_{DA}) + \exp(V_{SR}) + \exp(V_{TR})} \tag{4.9}$$

$$\Pr(i) = \frac{\exp(V_i)}{\sum_{j=DA,SR,TR} \exp(V_j)} \tag{4.10}$$

where i indicates the alternative for which the probability is being computed.

This formulation implies that the probability of choosing an alternative increases monotonically with an increase in the systematic utility of that alternative and decreases with increases in the systematic utility of each of the other alternatives. This is illustrated in Table 4-1 showing the probability of DA as a function of its own utility (with the utilities of other alternatives held constant) and in Table 4-2 as a function of the utility of other alternatives with its own utility fixed.

**Table 4-1 Probability Values for Drive Alone as a Function of Drive Alone Utility
(Shared Ride and Transit Utilities held constant)**

Case	V_{DA}	V_{SR}	V_{TR}	$\Pr(DA)$
1	-3.0	-1.5	-0.5	0.0566
2	-1.5	-1.5	-0.5	0.2119
3	0.0	-1.5	-0.5	0.5465
4	1.5	-1.5	-0.5	0.8438
5	3.0	-1.5	-0.5	0.9603

Table 4-2 Probability Values for Drive Alone as a Function of Shared Ride and Transit Utilities

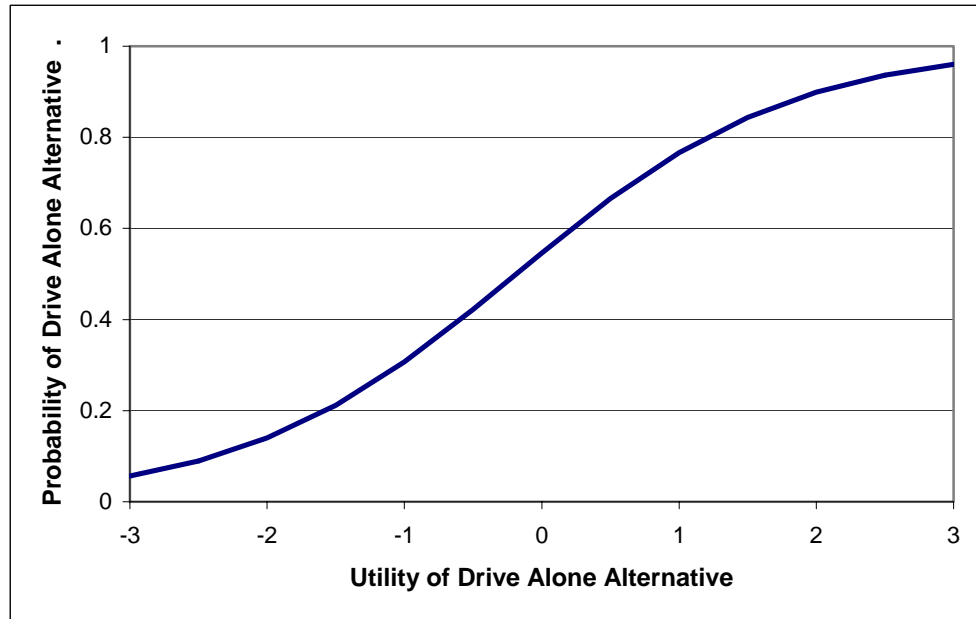
Case	V_{DA}	V_{SR}	V_{TR}	Pr(DA)
6	0.0	-1.5	-1.5	0.6914
7	0.0	-1.5	-1.0	0.6285
8	0.0	-1.5	-0.5	0.5465
9	0.0	-0.5	-1.5	0.5465
10	0.0	-0.5	-1.0	0.5065
11	0.0	-0.5	-0.5	0.4519

We use this three-alternative example to illustrate three important properties of the MNL: (1) its sigmoid or *S* shape, (2) dependence of the alternative choice probabilities on the differences in the systematic utility and (3) independence of the ratio of the choice probabilities of any pair of alternatives from the attributes and availability of other alternatives.

4.1.1 The Sigmoid or *S* shape of Multinomial Logit Probabilities

The *S* shape of the MNL probabilities is illustrated in Figure 4.4 where the probability of choosing Drive Alone is shown as a function of its own utility, with the utilities of the other alternatives held constant. The *S*-shape limits the probability range between zero when the utility of DA is very low, relative to other alternatives, and one when the utility of DA is very high, relative to other alternatives. This function has very gradual slope at extreme values of DA utility, relative to the other alternatives, and is much steeper when its utility reaches a value such that its choice probability is close to one-half. This implies that if the representative utility of one alternative is very low or very high, compared with the others, a small increase in the utility of this alternative will not substantially affect its probability of being chosen. The point at which an increase in the representative utility of an alternative has the greatest effect on its probability of being chosen (*i.e.*, the point of maximum slope along the curve) is when its representative

utility is equivalent to the combined utility of the other alternatives. When this is true, a small increase in the utility of one alternative can ‘tip the balance’ and induce a large increase in the probability of the alternative being chosen (Train, 1993).



4.1.2 The Equivalent Differences Property

A fundamental property of the multinomial logit and other choice models is that the choice probabilities of the alternatives depend only on the differences in the systematic utilities of different alternatives and not their actual values. This can be illustrated in two ways. First, we show that the choice probability equations are unchanged if the same incremental value, say ΔV , is added to the utility of each alternative. The original probabilities for the three alternatives in the example are given by:

$$\Pr(i) = \frac{\exp(V_i)}{\exp(V_{DA}) + \exp(V_{SR}) + \exp(V_{TR})} \quad 4.11$$

where i is the alternative for which the probabilities are being computed.

Adding ΔV to the systematic components of V_{DA} , V_{SR} and V_{TR} gives⁵:

$$\begin{aligned}
 \Pr(i) &= \frac{\exp(V_i + \Delta V)}{\exp(V_{DA} + \Delta V) + \exp(V_{SR} + \Delta V) + \exp(V_{TR} + \Delta V)} \\
 &= \frac{\exp(V_i) \times \exp(\Delta V)}{\exp(V_{DA}) \times \exp(\Delta V) + \exp(V_{SR}) \times \exp(\Delta V) + \exp(V_{TR}) \times \exp(\Delta V)} \\
 &= \frac{\exp(V_i) \times \exp(\Delta V)}{[\exp(V_{DA}) + \exp(V_{SR}) + \exp(V_{TR})] \times \exp(\Delta V)} \\
 &= \frac{\exp(V_i)}{[\exp(V_{DA}) + \exp(V_{SR}) + \exp(V_{TR})]}
 \end{aligned} \tag{4.12}$$

which is the same probability as before ΔV was added to each of the utilities. This result applies to any value of ΔV . We also illustrate this property through use of a numerical example for the three alternative choice problem used earlier. The following equations represent the case when the utility values for Drive Alone, Shared Ride and Transit equal -0.5, -1.5 and -3.0, respectively:

$$\Pr(DA) = \frac{\exp(-0.5)}{\exp(-0.5) + \exp(-1.5) + \exp(-3.0)} = 0.690 \tag{4.13}$$

$$\Pr(SR) = \frac{\exp(-1.5)}{\exp(-0.5) + \exp(-1.5) + \exp(-3.0)} = 0.254 \tag{4.14}$$

$$\Pr(TR) = \frac{\exp(-3.0)}{\exp(-0.5) + \exp(-1.5) + \exp(-3.0)} = 0.057 \tag{4.15}$$

Similarly, if the utility of each alternative is increased by one, the probabilities are:

$$\Pr(DA) = \frac{\exp(0.5)}{\exp(0.5) + \exp(-0.5) + \exp(-2.0)} = 0.690 \tag{4.16}$$

$$\Pr(SR) = \frac{\exp(-0.5)}{\exp(0.5) + \exp(-0.5) + \exp(-2.0)} = 0.254 \tag{4.17}$$

⁵ The exponent of a summation, $\exp(A + B)$, is equal to the product of the exponents of the elements in the sum, $\exp(A) \times \exp(B)$.

$$\Pr(TR) = \frac{\exp(-2.0)}{\exp(0.5) + \exp(-0.5) + \exp(-2.0)} = 0.057 \quad 4.18$$

As expected, the choice probabilities are identical to those obtained before the addition of the constant utility to each mode. The calculations supporting this comparison are shown in Table 4-3 and Table 4-4. Table 4-3 shows the computation of the choice probabilities based on the initial set of modal utilities and Table 4-4 shows the same computation after each of the utilities is increased by one⁶.

**Table 4-3 Numerical Example Illustrating Equivalent Difference Property:
Probability of Each Alternative Before Adding Delta**

Alternative	Utility		Exponent	Probability
	Expression	Value		
Drive Alone	-0.50	-0.50	0.607	0.690
Shared Ride	-1.50	-1.50	0.223	0.254
TRansit	-3.00	-3.00	0.050	0.057
			$\Sigma=0.879$	

**Table 4-4 Numerical Example Illustrating Equivalent Difference Property:
Probability of Each Alternative After Adding Delta (=1.0)**

Alternative	Utility		Exponent	Probability
	Expression	Value		
Drive Alone	-0.50 + 1.00	0.50	1.649	0.690
Shared Ride	-1.50 + 1.00	-0.50	0.607	0.254
TRansit	-3.00 + 1.00	-2.00	0.135	0.057
			$\Sigma=2.391$	

The expression for the probability equation of the logit model (equation 4.9) can also be presented in a different form which makes the equivalent difference property more apparent. For

⁶ We use tables to illustrate calculation of the utility values and MNL probabilities. The first column shows the specification expression with appropriate values for variables and parameters, the second shows the calculated utility value, the third shows the exponent of the utility including the sum of the exponents, and the fourth shows the probability values.

the drive alone alternative, this expression can be obtained by multiplying the numerator and denominator of the standard probability expression by $\exp(-V_{DA})$ as shown in the following equations.

$$\begin{aligned} \Pr(DA) &= \frac{\exp(V_{DA})}{\exp(V_{DA}) + \exp(V_{SR}) + \exp(V_{TR})} \times \frac{\exp(-V_{DA})}{\exp(-V_{DA})} \\ &= \frac{\exp(V_{DA}) \times \exp(-V_{DA})}{[\exp(V_{DA}) + \exp(V_{SR}) + \exp(V_{TR})] \times \exp(-V_{DA})} \\ &= \frac{\exp(0)}{\exp(0) + \exp(V_{SR} - V_{DA}) + \exp(V_{TR} - V_{DA})} \end{aligned} \quad 4.19$$

which simplifies to:

$$\Pr(DA) = \frac{1}{1 + \exp(V_{SR} - V_{DA}) + \exp(V_{TR} - V_{DA})} \quad 4.20$$

This formulation explicitly shows that the probability of the drive alone alternative is a function of the differences in systematic utility between the drive alone alternative and each other alternative. This can be applied to the general case for alternative i which can be represented in terms of the pairwise difference in its utility and the utility of each of the other alternatives by the following equation:

$$\Pr(i) = \frac{1}{1 + \sum_{j \neq i} \exp(V_j - V_i)} \quad \forall i \in J \quad 4.21$$

4.1.2.1

4.1.2.2 Implication of Constant Differences for Alternative Specific Constants and Variables

The constant difference property of logit models has an important implication for the specification of the utilities of the alternatives. Recall that the systematic portion of the utility of an individual, t , and alternative i is the sum of decision-maker related bias, mode attribute related utility, and interactions between these. That is:

$$V_{DA} = V_{DA}(S_t) + V(X_{DA}) + V(S_t, X_{DA}) \quad 4.22$$

$$V_{SR} = V_{SR}(S_t) + V(X_{SR}) + V(S_t, X_{SR}) \quad 4.23$$

$$V_{TR} = V_{TR}(S_t) + V(X_{TR}) + V(S_t, X_{TR}) \quad 4.24$$

Each term on the right hand side of each equation can be replaced by an explicit function of the relevant variables. For example, if the decision-maker preferences are a function of income, attribute based utility is a function of travel time and there are no interaction terms, the utility function becomes:

$$V_{DA} = \beta_{DA,0} + \beta_{DA,1} \times Income_t + \gamma \times TT_{DA} \quad 4.25$$

$$V_{SR} = \beta_{SR,0} + \beta_{SR,1} \times Income_t + \gamma \times TT_{SR} \quad 4.26$$

$$V_{TR} = \beta_{TR,0} + \beta_{TR,1} \times Income_t + \gamma \times TT_{TR} \quad 4.27$$

and the differences between pairs of alternatives for prediction of DA probability become:

$$V_{SR} - V_{DA} = (\beta_{SR,0} - \beta_{DA,0}) + (\beta_{SR,1} - \beta_{DA,1}) \times Income_t + \gamma \times (TT_{SR} - TT_{DA}) \quad 4.28$$

$$V_{TR} - V_{DA} = (\beta_{TR,0} - \beta_{DA,0}) + (\beta_{TR,1} - \beta_{DA,1}) \times Income_t + \gamma \times (TT_{TR} - TT_{DA}) \quad 4.29$$

It is not possible to estimate all of the constants; $\beta_{DA,0}$, $\beta_{SR,0}$ and $\beta_{TR,0}$; and all of the income parameters; $\beta_{DA,1}$, $\beta_{SR,1}$ and $\beta_{TR,1}$; in these equations because adding any algebraic value to each of the constants or to each of the income parameters does not cause any change in the probabilities of any of the alternatives. This phenomenon is common to all utility-based choice models and follows directly from the equivalent differences property discussed above. The solution to this problem is to place a single constraint on each set of parameters; in this case, the constants and the income parameters. Any constraint can be adopted for each set of parameters; however, the simplest and most widely used is to set the preference related parameters for one alternative, called the base or reference alternative, to zero and to re-interpret the remaining parameters to represent preference differences relative to the base alternative.

The selection of the reference alternative is arbitrary and does not affect the overall quality or interpretation of the model; however, the equations and the estimation results will appear to be different. For example, if we set TRansit as the reference alternative by setting $\beta_{TR,0}$ and $\beta_{TR,1}$ equal to zero, the utility functions become:

$$V_{DA,t} = \beta_{DA-TR,0} + \beta_{DA-TR,1} \times Inc_t + \gamma \times TT_{DA} \quad 4.30$$

$$V_{SR,t} = \beta_{SR-TR,0} + \beta_{SR-TR,1} \times Inc_t + \gamma \times TT_{SR} \tag{4.31}$$

$$V_{TR,t} = 0 + \gamma \times TT_{TR} \tag{4.32}$$

where the modified notation for the remaining constants and income parameters is used to emphasize that these parameters are ‘relative to the TRansit alternative.’ Alternatively, if we select Drive Alone as the reference alternative, we obtain:

$$V_{DA,t} = 0 + \gamma \times TT_{DA} \tag{4.33}$$

$$V_{SR,t} = \beta_{SR-DA,0} + \beta_{SR-DA,1} \times Income + \gamma \times TT_{SR} \tag{4.34}$$

$$V_{TR,t} = \beta_{TR-DA,0} + \beta_{TR-DA,1} \times Income + \gamma \times TT_{TR} \tag{4.35}$$

where the constants and income parameters are relative to the Drive Alone alternative.

These two models are equivalent as shown in Table 4-5 and Table 4-6 which correspond to the TRansit reference and Drive Alone reference examples, respectively, for an individual from a household with \$50,000 annual income and facing travel times of 30, 35 and 50 minutes for Drive Alone, Shared Ride and TRansit, respectively.

Table 4-5 Utility and Probability Calculation with TRansit as Base Alternative

Alternative	Utility		Exponent	Probability
	Expression	Value		
Drive alone	$1.1+0.008 \times 50^7 - 0.02 \times 30^8$	0.90	2.460	0.569
Shared ride	$0.8+0.006 \times 50 - 0.02 \times 35$	0.40	1.492	0.345
TRansit	$0.0+0.000 \times 50 - 0.02 \times 50$	-1.00	0.368	0.085
			$\Sigma=4.319$	

⁷ Income in thousands of dollars (\$000).

⁸ Time in minutes.

Table 4-6 Utility and Probability Calculation with Drive Alone as Base Alternative

Alternative	Utility		Exponent	Probability
	Expression	Value		
Drive alone	$0.0+0.000\times 50-0.02\times 30$	-0.60	0.549	0.569
Shared ride	$-0.3-0.002\times 50-0.02\times 35$	-1.10	0.333	0.345
TRansit	$-1.1-0.008\times 50-0.02\times 50$	-2.50	0.082	0.085
			$\Sigma=0.964$	

As expected, the resultant probabilities are identical in both cases. Table 4-7 shows that the differences in the alternative specific constants and income parameters between alternatives are the same for the TRansit base case and the Drive Alone base case.

Table 4-7 Changes in Alternative Specific Constants and Income Parameters

Alternative	TRansit as Base Alternative		Change in Parameters		Drive Alone as Base Alternative	
	Constant	Income	Constant	Income	Constant	Income
Drive Alone	1.1	0.008	-1.1	-0.008	0.0	0.000
Shared Ride	0.8	0.006	-1.1	-0.008	-0.3	-0.002
TRansit	0.0	0.000	-1.1	-0.008	-1.1	-0.008

4.2 Independence of Irrelevant Alternatives Property

One of the most widely discussed aspects of the multinomial logit model is its independence from irrelevant alternatives (IIA) property. The IIA property states that for any individual, the ratio of the probabilities of choosing two alternatives is independent of the presence or attributes of any other alternative. The premise is that other alternatives are irrelevant to the decision of choosing between the two alternatives in the pair. To illustrate this, consider a multinomial logit model for the choice among three intercity travel modes – automobile, rail, and bus. The probability of choosing automobile, rail and bus are:

$$\Pr(Auto) = \frac{\exp(V_{Auto})}{\exp(V_{Auto}) + \exp(V_{Bus}) + \exp(V_{Rail})} \quad 4.36$$

$$\Pr(Bus) = \frac{\exp(V_{Bus})}{\exp(V_{Auto}) + \exp(V_{Bus}) + \exp(V_{Rail})} \quad 4.37$$

$$\Pr(Rail) = \frac{\exp(V_{Rail})}{\exp(V_{Auto}) + \exp(V_{Bus}) + \exp(V_{Rail})} \quad 4.38$$

The ratios of each pair of probabilities are:

$$\frac{\Pr(Auto)}{\Pr(Bus)} = \frac{\exp(V_{Auto})}{\exp(V_{Bus})} = \exp(V_{Auto} - V_{Bus}) \quad 4.39$$

$$\frac{\Pr(Auto)}{\Pr(Rail)} = \frac{\exp(V_{Auto})}{\exp(V_{Rail})} = \exp(V_{Auto} - V_{Rail}) \quad 4.40$$

$$\frac{\Pr(Bus)}{\Pr(Rail)} = \frac{\exp(V_{Bus})}{\exp(V_{Rail})} = \exp(V_{Bus} - V_{Rail}) \quad 4.41$$

The ratios of probabilities for each pair of alternatives depend only on the attributes of those alternatives and not on the attributes of the third alternative and would remain the same regardless of whether that third alternative is available or not. This formulation can be generalized to any pair of alternatives by:

$$\frac{\Pr(i)}{\Pr(k)} = \frac{\exp(V_i)}{\exp(V_k)} = \exp(V_i - V_k) \quad 4.42$$

which, as before, is independent of the number or attributes of other alternatives in the choice set.

The IIA property has some important ramifications in the formulation, estimation and use of multinomial logit models. The independence of irrelevant alternatives property allows the addition or removal of an alternative from the choice set without affecting the structure or parameters of the model. The flexibility of applying the model to cases with different choice sets has a number of advantages. First, the model can be estimated and applied in cases where different members of the population (and sample) face different sets of alternatives. For example, in the case of intercity mode choice, individuals traveling between some city pairs may not have air service and/or rail service. Second, this property simplifies the estimation of the parameters in the multinomial logit model (as will be discussed later). Third, this property is

advantageous when applying a model to the prediction of choice probabilities for a new alternative.

On the other hand, the IIA property may not properly reflect the behavioral relationships among groups of alternatives. That is, other alternatives may not be irrelevant to the ratio of probabilities between a pair of alternatives. In some cases, this will result in erroneous predictions of choice probabilities. An extreme example of this problem is the classic “red bus/blue bus paradox.”

4.2.1 The Red Bus/Blue Bus Paradox

Consider the case of a commuter who has a choice of going to work by auto or taking a blue bus. Assume that the attributes of the auto and the blue bus are such that the probability of choosing auto is two-thirds and blue bus is one-third so the ratio of their choice probabilities is 2:1. Now suppose that a competing bus operator introduces red bus service (the bus is painted red, rather than blue) on the same route, operating the same vehicle type, using the same schedule and serving the same stops as the blue bus service. Thus, the only difference between the red and blue bus services is the color of the buses.

The most reasonable expectation, in this case, is that the same share of people will choose auto and bus and that the bus riders will split equally between the red and blue bus services. That is, the addition of the red bus to the commuters’ choice set should have no, or very little, effect on the share of commuters choosing auto since this change does not affect the relative quality of drive alone and bus. Therefore, we expect choice probabilities following the initiation of red bus service to be auto, two-thirds; blue bus, one-sixth and red bus, one-sixth. However, due to the IIA property, the multinomial logit model will maintain the relative probability of auto and blue bus as 2:1. If we assume that people are indifferent to color of their transit vehicle, the two bus services will have the same representative utility and consequently, their relative probabilities will be 1:1 and the share probabilities for the three alternatives will be: $\Pr(\text{Auto}) = \frac{1}{2}$, $\Pr(\text{Blue Bus}) = \frac{1}{4}$, and $\Pr(\text{Red Bus}) = \frac{1}{4}$. That is, the probability (share) of people choosing auto will decline from two-thirds to one half as a result of introducing an

alternative which is identical to an existing alternative⁹. The red bus/blue bus paradox provides an important illustration of the possible consequences of the IIA property. Although this is an extreme case; the IIA property can be a problem in other, less extreme cases.

4.3 Example: Prediction with Multinomial Logit Model

We illustrate the application of multinomial logit models with different specifications in the context of mode choice analysis. Consider a commute trip by an individual who has three available modes in the choice set: drive alone, carpool, and bus. The examples in this section illustrate the manner in which different utility specifications and the estimated parameters associated with them are used to predict choice probabilities based on characteristics of the traveler (decision-maker) and attributes of the alternatives. These examples progress from the simplest models to moderately complex models.

Example 1 -- Constants Only Model

The simplest specification of the multinomial logit model is the ‘constants only’ model, in which the utility of each alternative has a fixed value for all decision-makers. Typically, the alternative specific constants are considered to represent the average effect of all factors that influence the choice but are not included in the utility specification. For example, factors such as comfort, safety, privacy and reliability may be excluded due to the difficulty associated with their measurement. In the constants only model, it is implicitly assumed that the constants reflect the average effects of all the variables affecting the choice decision, since no variables are included explicitly in the utility specification. If these constants are 0.0, -1.6 and -1.8 for drive alone, shared ride and transit, respectively, the probability calculation is as shown in Table 4-8

⁹ This example ignores the increase in bus service frequency which might increase the probability of persons choosing bus; however, the increase is unlikely to be of the magnitude suggested by the IIA property.

Table 4-8 MNL Probabilities for Constants Only Model

Alternative	Utility		Exponent	Probability
	Expression	Value		
Drive Alone	0.0	0.0	1.0000	0.7314
Shared Ride	-1.60	-1.60	0.2019	0.1477
Transit	-1.80	-1.80	0.1653	0.1209
			1.3672	

As expected, Drive Alone has the highest probability, followed by Shared Ride, and TRansit.

Example 2 -- Including Mode Related Variables - Travel Time and Travel Cost

Two key attributes that influence choice of mode are travel time and travel cost. We include these variables in the deterministic component of the utility function of each mode with the parameter for time (in minutes) equal to -0.045 and for cost (in cents) equal to -0.004 for all three modes, Table 4.9. This implies that a minute of travel time (or a cent of cost) has the same marginal disutility regardless of the mode; such variables are referred to as *generic* variables. The negative signs of the travel time and travel cost coefficients imply that the utility of a mode and the probability that it will be chosen decreases as the travel time or travel cost of that mode increases. Positive coefficients would be inconsistent with our understanding of travel behavior and therefore any specification which results in a positive sign for travel time or travel cost should be rejected. Such counter-intuitive results are most likely due to an incorrect or inadequate model specification; however, it is possible that the data from any particular sample leads to such counter-intuitive results.

The inclusion of travel time and travel cost variables induces a change in the alternative specific constants, to -1.865 for shared ride and -0.650 for transit, as the effect of excluding these time and cost variables is removed from the constants. Such changes in alternative specific

constants, as a result of the introduction of new variables or the elimination of included variables preserve the sample shares¹⁰ and are expected.

To illustrate the application of the multinomial logit model for the above utility equation, we assume travel time and travel cost values as follows:

<u>Mode</u>	<u>Travel Time</u>	<u>Travel Cost</u>
Drive Alone	25 minutes	\$1.75
Shared Ride	28 minutes	\$0.75
TRansit	55 minutes	\$1.25

The utilities and probabilities are calculated as shown in Table 4-9.

Table 4-9 MNL Probabilities for Time and Cost Model

Alternative	Utility		Exponent	Probability
	Expression	Value		
Drive Alone	$-0.045 \times 25 - 0.004 \times 175$	-1.825	0.1612	0.7314
Shared Ride	$-1.865 - 0.045 \times 28 - 0.004 \times 75$	-3.425	0.0325	0.1477
TRansit	$-0.650 - 0.045 \times 55 - 0.004 \times 125$	-3.625	0.0266	0.1209
			0.2204	

This specification can be refined further by decomposing travel time into its two major components: (1) in-vehicle travel time, and (2) out-of-vehicle travel time. In-vehicle time (IVT) is defined as the time spent inside the vehicle, and out-of-vehicle time (OVT) is the time not spent inside the vehicle (including access time, waiting time, and egress time). There is an abundance of empirical evidence that travelers are much more sensitive to out-of-vehicle time than to in-vehicle time and therefore a minute of out-of-vehicle time will generate a higher disutility than a minute of in-vehicle time. This will be reflected in the modal utilities by a larger negative coefficient on out-of-vehicle time than on in-vehicle time. Introduction of this refinement will usually result in a less negative parameter for in vehicle time and a more

¹⁰ In this case, we show the preservation of probabilities for the individual. In general, the individual probabilities are expected to change.

negative parameter for out of vehicle time than for total time; say -0.031 and -0.062, respectively. If the travel times are split as follows:

<u>Mode</u>	<u>IVT</u>	<u>OVT</u>	<u>Travel Cost</u>
Drive Alone	21 minutes	4 minutes	\$1.75
Shared Ride	23 minutes	5 minutes	\$0.75
Bus	25 minutes	30 minutes	\$1.25

the new systematic utilities and choice probabilities are as computed in Table 4-10:

Table 4-10 MNL Probabilities for In and Out of Vehicle Time and Cost Model

Alternative	Utility		Exponent	Probability
	Expression	Value		
Drive Alone	$-0.031 \times 21 - 0.062 \times 4 - 0.004 \times 175$	-1.599	0.202	0.773
Shared Ride	$-1.90 - 0.031 \times 23 - 0.062 \times 5 - 0.004 \times 75$	-3.223	0.040	0.152
TRansit	$-0.80 - 0.031 \times 25 - 0.062 \times 30 - 0.004 \times 125$	-3.935	0.020	0.075
			0.261	

Example 3 -- Including Decision-Maker Related Biases - Income

The preceding examples do not include any characteristics of the traveler in the modal utilities. However, we know that choice probabilities of the available modes also depend on characteristics of the traveler, such as his/her income. Economic theory and empirical evidence suggests that higher income travelers are less likely to choose transit than drive alone or carpool. We can incorporate this behavior in the model by including an alternative specific income variable in the utility of up to two of the alternatives; in this case, we include income in the transit alternative with a negative coefficient. That is, everything else held constant, the utility of transit decreases as the income of the traveler increases. Consequently, a higher income traveler will have a lower probability of choosing transit than a lower income traveler. The absence of an alternative specific parameter for the carpool alternative implies that the choice of carpool, relative to drive alone, is unaffected by a traveler’s income. The alternative specific constant of the transit utility changes substantially from the preceding example as it no longer reflects the

average effect of excluding income from the transit utility specification. The calculation of utilities and probabilities for this model for a person from a household with \$50,000 annual income is shown in Table 4-11.

Table 4-11 MNL Probabilities for In and Out of Vehicle Time, Cost and Income Model

Alternative	Utility		Exponent	Probability
	Expression	Value		
Drive Alone	$-0.031 \times 21 - 0.062 \times 4 - 0.004 \times 175$	-1.599	0.202	0.780
Shared Ride	$-1.90 - 0.031 \times 23 - 0.062 \times 5 - 0.004 \times 75$	-3.223	0.040	0.154
TRansit	$-0.50 - 0.031 \times 25 - 0.062 \times 30 - 0.004 \times 125$ -0.0087×50	-4.070	0.017	0.066
			0.259	

The probability of choosing transit is smaller for this traveler than would have been predicted using the model reported in the preceding example. This model will give decreasing transit probabilities for higher income travelers and increasing transit probabilities for lower income travelers. That is, the lower the traveler’s income, the greater his/her probability of choosing the least expensive mode of travel (transit), an intuitive and reasonable result.

Example 4 – Interaction of Mode Attributes and Decision-Maker Related Biases

An alternative method of including income in the utility specification is to use income as a deflator of cost by forming a variable by dividing cost by income. This formulation reflects the rationale that cost becomes a less important factor in the choice of a travel mode as the income of the traveler increases. The revised utility functions and calculations are shown in Table 4-12 using the values for the modal attributes and income as used in preceding example.

Table 4-12 MNL Probabilities for In and Out of Vehicle Time, and Cost/Income Model

Alternative	Utility		Exponent	Probability
	Expression	Value		
Drive Alone	$-0.031 \times 21 - 0.062 \times 4 - 0.153 \times (175/50)$	-1.435	0.238	0.763
Shared Ride	$-1.90 - 0.031 \times 23 - 0.062 \times 5 - 0.153 \times (75/50)$	-3.153	0.043	0.137
TRansit	$-0.45 - 0.031 \times 25 - 0.062 \times 30 - 0.153 \times (125/50)$	-3.468	0.031	0.100
			0.312	

This specification of income in the utility function also results in lower income travelers predicted to have higher probability of choosing transit; it also suggests that such travelers will increase their probability of choosing carpool, the least expensive mode. The reader should compute the probabilities for different income values and verify the response pattern.

4.4 Measures of Response to Changes in Attributes of Alternatives

Choice probabilities in logit models are a function of the values of the attributes that define the utility of the alternatives; therefore, it is useful to know the extent to which the probabilities change in response to changes in the value of those attributes. For example, in a traveler's mode choice decision, an important question is to what extent the probability of choosing a mode (rail, for example) will decrease/increase, if the fares of that mode are increased by a certain amount. Similarly, a transit agency may want to know the gain in ridership that is likely to occur in response to service improvements (increased frequency). This section describes various aspects of understanding and quantifying the response to changes in attributes of alternatives.

4.4.1 Derivatives of Choice Probabilities

One measure for evaluating the response to changes is to calculate the *derivatives* of the choice probabilities of each alternative with respect to the variable in question. Usually, one is concerned about the change in probability of an alternative, P_i , with respect to the change in attributes of that alternative X_i . This measure, the *direct derivative*, is computed by

differentiating P_i with respect to X_{ik} , the k^{th} attribute of alternative i . The mathematical expression for the *direct derivative* of P_i with respect to X_{ik} is:

$$\frac{\partial P_i}{\partial X_{ik}} = \left(\frac{\partial V_i}{\partial X_{ik}} \right) \times (P_i) \times (1 - P_i) \quad 4.43$$

where V_i is the utility of the alternative¹¹

Typically the utility function is specified to be linear in parameters; that is:

$$V_i = \beta_o + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \dots + \beta_K X_{Ki} \quad 4.44$$

In this case, the expression for the *direct derivative* of P_i with respect to X_{ik} reduces to:

$$\frac{\partial P_i}{\partial X_{ki}} = \beta_k \times (P_i) \times (1 - P_i) \quad 4.45$$

where β_k is the coefficient of attribute k .

The value of the derivative is largest at $P_i = 1/2$ and becomes smaller as P_i approaches zero or one. This implies that the magnitude of the response to a change in an attribute will be greatest when the choice probability for the alternative under consideration is 0.5 and this response diminishes as the probability approaches zero or one. The direct derivative is simply the slope of the logit model probability curve illustrated in Figure 4.4 and that its mathematical properties are consistent with the qualitative discussion of the S-shape of the logit probability curve in section 4.1.1. The sign of the derivative is the same as the sign of the parameter describing the impact of X_{ik} in the utility of alternative i . Thus, an increase in X_{ik} will increase (decrease) P_i if β_{ik} is positive (negative).

Often it is important to understand how the choice probability of other alternatives changes in response to a given change in the attribute level of the action alternative. This

¹¹ Interested readers are referred to Train (1986) for a derivation and proof.

measure, termed the *cross derivative*, is obtained by computing the derivative of the choice probability of an alternative, P_j , with respect to the attribute of the changed alternative, X_{ik} .

This *cross derivative* for linear utility functions is:

$$\frac{\partial P_j}{\partial X_{ik}} = -\beta_k \times (P_i) \times (P_j) \quad \forall i \neq j \quad 12 \quad 4.46$$

- where β_{jk} is the coefficient of the k^{th} attribute of alternative j ,
- P_i is the probability of alternative i , and
- P_j is the probability of alternative j .

In this case, the sign of the derivative is opposite to the sign of the parameter describing the impact of X_{ik} on the utility of alternative i . Thus an increase in X_{ik} will decrease (increase) the probability of choosing alternative, P_j , if the parameter β_k is positive (negative).

It is useful to recognize that the sum of the derivatives over all the alternatives must be equal to zero. That is,

$$\begin{aligned} \sum_{\forall j} \frac{\partial P_j}{\partial X_{ik}} &= \frac{\partial P_i}{\partial X_{ik}} + \sum_{\forall j \neq i} \frac{\partial P_j}{\partial X_{ik}} \\ &= \beta_k P_i (1 - P_i) - \sum_{\forall j \neq i} \beta_k P_i P_j \\ &= \beta_k P_i (1 - P_i) - \beta_k P_i \sum_{\forall j \neq i} P_j \\ &= \beta_k P_i (1 - P_i) - \beta_k P_i (1 - P_i) = 0 \end{aligned} \quad 4.47$$

This is as expected. Since the sum of all probabilities is fixed at one, the sum of the derivatives of the probability due to a change in any attribute of any alternative must be equal to zero.

4.4.2 Elasticities of Choice Probabilities

Elasticity is another measure that is used to quantify the extent to which the choice probabilities of each alternative will change in response to the changes in the value of an attribute. In general,

12 As before, interested readers are referred to Train (1986) for derivation and proof.

elasticity is defined as the percentage change in the response variable with respect to a one percent change in an explanatory variable. In the context of logit models, the response variable is the choice probability of an alternative, such as P_i , and the explanatory variable is the attribute X_{ik} . Elasticities are different from derivatives in that elasticities are normalized by the variable units. To clearly illustrate the concept of elasticity, let us consider that P_{i1} and P_{i2} are choice probabilities of an alternative i at attribute levels X_{i1} and X_{i2} , respectively. In this case, the elasticity is the proportional change in the probability divided by the proportional change in the attribute under consideration:

$$\begin{aligned} \text{Elasticity} &= \frac{\text{Percentage Change in Probability}}{\text{Percentage Change in Attribute}} \\ &= \frac{(P_2 - P_1)/P_1}{(X_2 - X_1)/X_1} = \frac{\Delta P/P_1}{\Delta X/X_1} \end{aligned} \quad 4.48$$

There is some ambiguity in the computation of this elasticity measure in terms of whether it should be normalized using the original probability-attribute combination (P_{i1}, X_{i1}) or the new probability-attribute combination (P_{i2}, X_{i2}) . A compromise approach is to compute the elasticity relative to the mid-point of both sets of variables, yielding a measure called the *arc elasticity*. The expression for arc elasticity is:

$$\text{Arc Elasticity} = \frac{\left(\frac{P_2 - P_1}{(P_1 + P_2)/2} \right)}{\left(\frac{X_2 - X_1}{(X_1 + X_2)/2} \right)} = \frac{\left(\frac{\Delta P}{(P_1 + P_2)/2} \right)}{\left(\frac{\Delta X}{(X_1 + X_2)/2} \right)} \quad 4.49$$

Estimates of elasticity will differ depending on whether the normalization is at the start value, final value or mid-point values. This confusion can be avoided by computing elasticities for very small changes. When the elasticity is computed for infinitesimally small changes, the elasticity obtained is termed the *point elasticity*. The expression for point elasticity is given as a function of the derivatives discussed earlier (section 4.4.1):

$$\eta_X^P = \frac{\left(\frac{\partial P}{P}\right)}{\left(\frac{\partial X}{X}\right)} = \left(\frac{\partial P}{\partial X}\right) \times \left(\frac{X}{P}\right) \quad 4.50$$

We are interested in both *direct-* and *cross-elasticities* corresponding to the direct- and cross-derivatives discussed above. The direct elasticity measures the percent change in the choice probability of alternative, P_i with respect to a percent change in the attribute level (X_{ik}) of that alternative:

$$\eta_{X_{ik}}^{P_i} = \frac{\left(\frac{\partial P_i}{P_i}\right)}{\left(\frac{\partial X_{ik}}{X_{ik}}\right)} = \left(\frac{\partial P_i}{\partial X_{ik}}\right) \times \left(\frac{X_{ik}}{P_i}\right) \quad 4.51$$

We can substitute equation 4.45 into 4.51 to obtain the following expression for the direct elasticity

$$\eta_{X_{ik}}^{P_i} = \beta_k P_i (1 - P_i) \left(\frac{X_{ik}}{P_i}\right) = \beta_k X_{ik} (1 - P_i) \quad 4.52$$

Thus, the direct elasticity is not only a function of the parameter value, β_k , for the attribute in the utility, but is also a function of the attribute level, X_{ik} , at which the elasticity is being computed.

Similarly, the cross-elasticity is defined as the proportional change in the choice probability of an alternative (P_j) with respect to a proportional change in some attribute of another alternative (X_{ik}). The expression for cross elasticity in a multinomial logit model is given by:

$$\eta_{X_{ik}}^{P_j} = \frac{\left(\frac{\partial P_j}{P_j}\right)}{\left(\frac{\partial X_{ik}}{X_{ik}}\right)} = \left(\frac{\partial P_j}{\partial X_{ik}}\right) \times \left(\frac{X_{ik}}{P_j}\right) \quad \forall j \neq i \quad 4.53$$

We substitute equation 4.46 into 4.53 to obtain the following expression for cross elasticity of logit model probabilities:

$$\eta_{X_{ik}}^{P_j} = (-\beta_k P_i P_j) \left(\frac{X_{ik}}{P_j} \right) = -\beta_k X_{ik} P_i \quad 4.54$$

An important property of MNL models is that the cross-elasticities are the same for all the other alternatives in the choice set. This property of the multinomial logit model is another manifestation of the IIA property discussed in the previous section.

The elasticity expressions derived above are based on the assumption that the independent variables for which the elasticities are being derived are continuous variables (*e.g.*, travel time and travel cost). However, if the variable of interest is discrete in nature (*e.g.*, number of automobiles in a household), it is not differentiable by definition. For this reason, derivatives can not be derived for ordinal or categorical variables. An alternative is to calculate the incremental change in each probability with respect to a one unit change in an ordered variable or a category shift for categorical variables and to use the differences to compute arc elasticities.

4.5 Measures of Responses to Changes in Decision Maker Characteristics

Choice probabilities in logit models are also a function of the values of the characteristics of the decision maker (traveler). Therefore, it is equally useful to know the extent to which the probabilities of alternatives change in response to changes in the value of these characteristics. For example, in a traveler's mode choice decision, an important question is to what extent the probability of choosing a mode will decrease/increase, if the income of the traveler changes by a certain amount. This section formulates both the derivatives and elasticity equations for evaluating such responses and thereby provides understanding and quantifies the response to changes in the characteristics of travelers.

4.5.1 Derivatives of Choice Probabilities

Derivatives indicate the change in probability of each alternative in the choice set per unit change in a characteristic of the traveler. The analysis approach is similar to that employed in

section 4.4. The important difference is that in section 4.4, we were assessing the impact of a change in probability of an alternative in response to a change in an attribute of a single alternative; either the same alternative (direct-derivative and direct-elasticity) or another alternative (cross-derivative and cross-elasticity). In the case of traveler characteristics, those characteristics may appear in alternative specific form in all alternatives (except for one reference alternative). Thus, we are considering what, in effect, becomes a combination of one direct response and multiple cross responses. Consider, for example, the probability of choosing alternative i in response to a change in income, specific to alternative i . That is,

$$\frac{\partial P_i}{\partial Inc_i} = (1 - P_i) P_i \beta_{Inc_i} \quad 4.55$$

However, since an identical change in income will occur for all alternatives in which income appears as an alternative specific variable, we consider the cross-derivative of the probability of choosing alternative i in response to a change in income, specific to alternative j . That is,

$$\frac{\partial P_i}{\partial Inc_j} = -P_i P_j \beta_{Inc,j} \quad 4.56$$

The corresponding sum over all alternatives $j \neq i$ is

$$\sum_{j \neq i} \frac{\partial P_i}{\partial Inc_j} = -P_i \sum_{j \neq i} P_j \beta_{Inc,j} \quad 4.57$$

and the sum over all alternatives including i is

$$\begin{aligned}
 \frac{\partial P_i}{\partial Inc} &= P_i(1 - P_i)\beta_{Inc,i} - \sum_{j \neq i} P_i P_j \beta_{Inc,j} \\
 &= P_i \beta_{Inc,j} - \sum_{\forall j} P_i P_j \beta_{Inc,j} \\
 &= P_i \left[\beta_{Inc,i} - \sum_{\forall j} P_j \beta_{Inc,j} \right] \\
 &= P_i \left[\beta_{Inc,i} - \bar{\beta}_{Inc} \right]
 \end{aligned}
 \tag{4.58}$$

where $\bar{\beta}_{Inc}$ is the probability weighted average of the alternative specific income parameters

That is, the derivative of the probability with respect to a change in income is equal to the probability times the amount by which the income coefficient for that alternative exceeds the probability weighted average income coefficient over all alternatives¹³.

It should be apparent that the sum over all alternatives of the income derivatives must be equal to one to ensure that the total probability over all alternatives is unchanged by any change in income. This can be shown by summing equation 4.58 over all alternatives in the choice set.

4.5.2 Elasticities of Choice Probabilities

Elasticity is another measure that can be used to quantify the extent to which the choice probabilities are influenced by changes in a variable; in this case, a variable that describes the characteristics of the traveler. In this case, the elasticity of the probability of alternative i to a change in income is given by

$$\eta_{Inc}^{P_i} = \left(\beta_{Inc,i} - \bar{\beta}_{Inc} \right) \times Inc \tag{4.59}$$

As before, the elasticity represents the proportional change in probability of an alternative to a proportional change in the explanatory variable.

¹³ This comparison is independent of which alternative is chosen as the reference alternative since the effect of changing the reference alternative will be to increase each parameter by the same value.

4.6 Model Estimation: Concept and Method

Logit model development consists of formulating model specifications and estimating numerical values of the parameters for the various attributes specified in each utility function by fitting the models to the observed choice data. The critical elements of this process become the selection of a preferred specification based on statistical measures and judgment. Under some circumstances, the model developer may impose constraints on the estimation to ensure desired relationships with respect to the relative value of different variables.

4.6.1 Graphical Representation of Model Estimation

We illustrate the basic concepts of model estimation using a binary choice model with two variables and no constant. Consider that the only two modes available to a traveler are auto and bus and the deterministic component of the utility function for the two modes is defined as follows:

$$V_{Auto} = \beta_1 \times TravelTime_{Auto} + \beta_2 \times TravelCost_{Auto} \quad 4.60$$

$$V_{Bus} = \beta_1 \times TravelTime_{Bus} + \beta_2 \times TravelCost_{Bus} \quad 4.61$$

Let us assume that in this example, bus has higher travel time and lower cost than car. For a traveler, the choice of travel mode will depend on his/her relative valuation of travel time and travel cost. A traveler who values time much more than cost will have a higher utility for car; whereas, a traveler who values time less than cost will have a higher utility for bus. This concept is illustrated in Figure 4.5 which shows the time and cost values for auto and bus iso-utility line¹⁴ for both types of travelers. When the iso-utility lines are steep enough to ensure that some iso-utility line is below bus and above car (case 1) implying a low value of time, bus will be the chosen mode. Conversely, if the iso-utility lines are flat enough to ensure that some iso-utility line is below car and above bus (case 2) implying a high value of time, car will be the chosen mode.

¹⁴ Iso-utility lines connect all points for which the values of time and cost result in the same utility. The slope of these lines equals the inverse of the value of time, β_1 / β_2 .

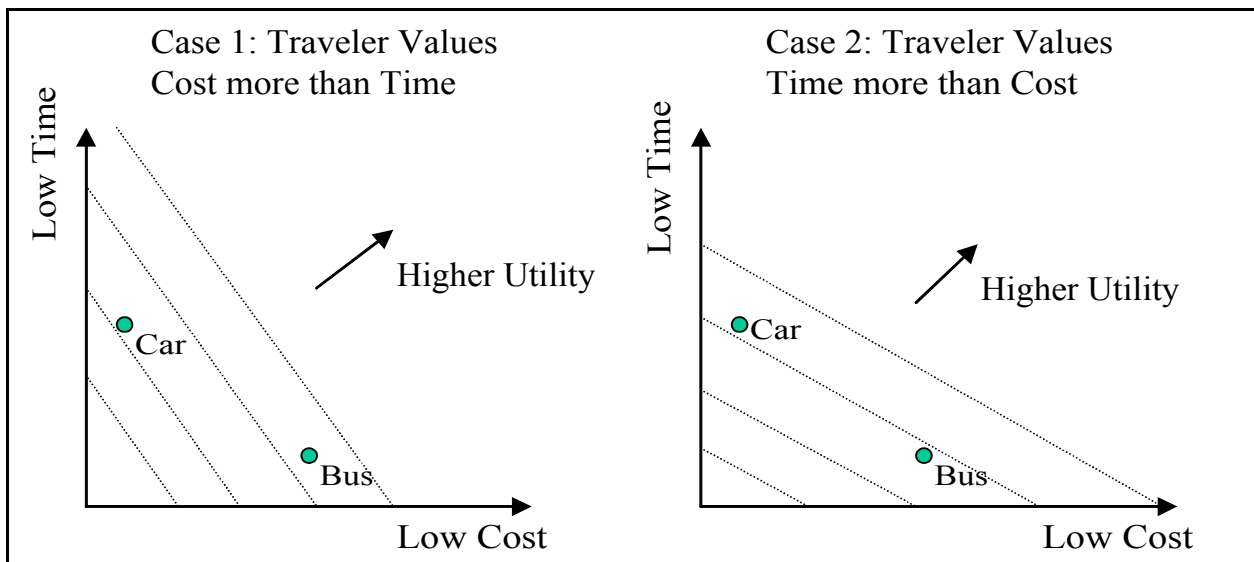


Figure 4.5 Iso-Utility Lines for Cost-Sensitive versus Time-Sensitive Travelers

The concept of iso-utility lines can be used to graphically describe the estimation of model parameters using observed choice data. This is illustrated in Figure 4.6 which shows the observed choice data for two travelers where each traveler has a choice between bus and car. The first traveler chooses bus (Bus 1) whereas the second traveler chooses car (Car 2). The objective is to find an iso-utility line with a slope that is steep enough to place Bus 1 above Car 1 and flat enough to place Bus 2 below Car 2. The figure shows that there many iso-utility lines with different slopes that satisfy the above conditions. However, even with only two observations, the range of the slopes of these lines is limited. The addition of more choice observation will further reduce the possible range of iso-utility lines. If all choosers are governed by the strict utility equations 4.60 and 4.61; additional observations will narrow the range of results to any satisfactory level of precision.

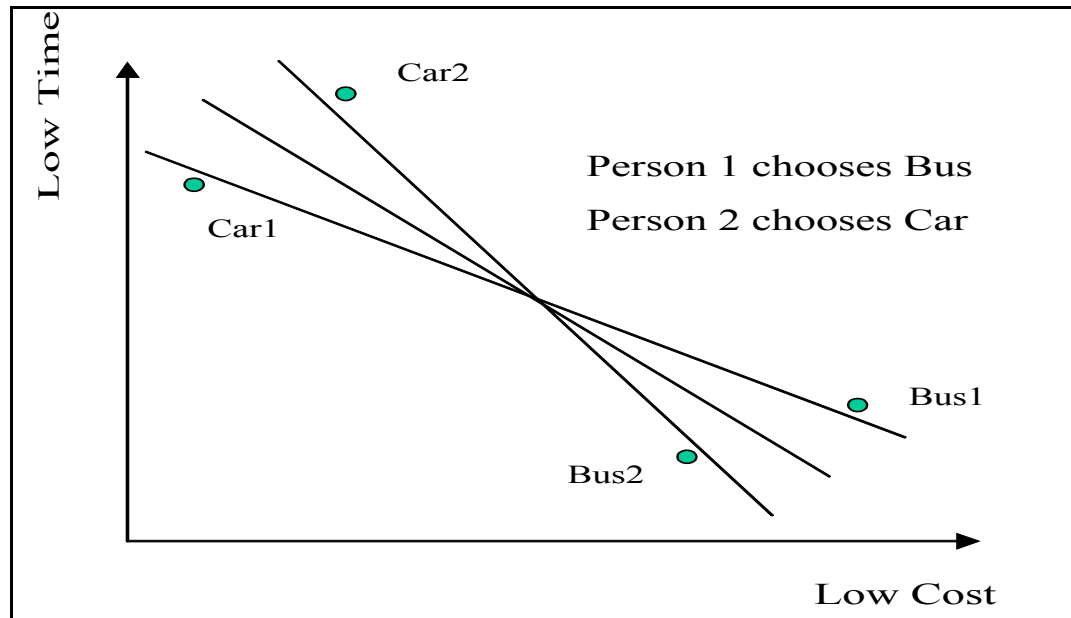


Figure 4.6 Estimation of Iso-Utility Line Slope with Observed Choice Data

However, as discussed previously, we expect that the analyst will not know all the variables that influence the traveler's choice and/or will measure some variables differently than the user. Thus, it is unlikely that the observations can be separated by a single boundary. In this case, it becomes necessary to use an estimation method which scores different estimation results in terms of how well they identify the chosen alternatives. This is accomplished by using maximum likelihood estimation methods. The maximum likelihood method consists of finding model parameters which maximize the likelihood (posterior probability) of the observed choices conditional on the model. That is, to maximize the likelihood that the sample was generated from the model with the selected parameter values.

4.6.2 Maximum Likelihood Estimation Theory

The procedure for maximum likelihood estimation involves two important steps: 1) developing a joint probability density function of the observed sample, called the *likelihood function*, and 2) estimating parameter values which maximize the likelihood function. The likelihood function for a sample of ' T ' individuals, each with ' J ' alternatives is defined as follows:

$$L(\beta) = \prod_{\forall t \in T} \prod_{\forall j \in J} (P_{jt}(\beta))^{\delta_{jt}} \tag{4.62}$$

where $\delta_{jt} = 1$ is chosen indicator (=1 if j is chosen by individual t and 0, otherwise) and

P_{jt} is the probability that individual t chooses alternative j .

The values of the parameters which maximize the likelihood function are obtained by finding the first derivative of the likelihood function and equating it to zero. Since the log of a function yields the same maximum as the function and is more convenient to differentiate, we maximize the *log-likelihood function* instead of the likelihood function itself. The expressions for the log-likelihood function and its first derivative are shown in equations 4.63 and 4.64 respectively:

$$LL(\beta) = \text{Log}(L(\beta)) = \sum_{\forall t \in T} \sum_{\forall j \in J} \delta_{jt} \times \ln(P_{jt}(\beta)) \tag{4.63}$$

$$\frac{\partial(LL)}{\partial \beta_k} = \sum_{\forall t \in T} \sum_{\forall j \in J} \delta_{jt} \times \frac{1}{P_{jt}} \times \frac{\partial P_{jt}(\beta)}{\partial \beta} \quad \forall k \tag{4.64}$$

Further development of the derivative requires representation of the probability function, P_{jt} , expanded from the version that appears in Equation 4.5 is

$$P_{jt} = \frac{\exp(X'_{jt}\beta)}{\sum_{j'} \exp(X'_{j't}\beta)} \tag{4.65}$$

and the first derivative with respect to each element of β is

$$\frac{\partial P_{jt}}{\partial \beta_k} = P_{jt} \left(X'_{jkt} - \sum_{j'} P_{j't} X'_{j'kt} \right) \quad \forall k \tag{4.66}^{15}$$

Substituting Equation 4.66 into Equation 4.64 gives

¹⁵ Dependence of $P_{j't}$ on β is made implicit to simplify notation.

$$\begin{aligned} \frac{\partial(LL)}{\partial\beta_k} &= \sum_{\forall t \in T} \sum_{\forall j \in J} \delta_{jt} \left(X'_{jt} - \sum_{j't} P_{j't} X'_{j't} \right) \\ &= \sum_{\forall t \in T} \sum_{\forall j \in J} (\delta_{jt} - P_{j't}) X'_{jt} \quad \forall k \end{aligned} \tag{4.67}$$

for the derivative of the log-likelihood with respect to β_k . The maximum likelihood is obtained by setting Equation 4.67 equal to zero and solving for the best values of the parameter vector, $\hat{\beta}$. We can be sure this is the solution for a maximum value provided that the second derivative is negative definite. In this case, the second derivative of the log-likelihood with respect to β is

$$\frac{\partial^2(LL)}{\partial\beta\partial\beta'} = \sum_{\forall t \in T} \sum_{\forall j \in J} -P_{j't} (X'_{jt} - \bar{X}_t)(X'_{jt} - \bar{X}_t)' \tag{4.68}$$

is negative definite for all values of β . Equations 4.67 and 4.68 are used to solve the maximum likelihood problem using a variety of available algorithms. In most practical problems, this involves significant computations and specialized computer programs to find the desired solution.

The following example illustrates the application of maximum likelihood to estimate logit model parameters.

4.6.3 Example of Maximum Likelihood Estimation

Suppose a logit model of binary choice between car and bus is to be estimated. For simplicity, let us assume the utility specification includes only the travel time variable and that the deterministic portion of the utility function for the two modes is defined as follows:

$$V_{Auto} = \beta_1 \times Travel\ Time_{Auto} \tag{4.69}$$

$$V_{Bus} = \beta_1 \times Travel\ Time_{Bus} \tag{4.70}$$

Suppose that the estimation sample consists of observations of mode choice of only three individuals. The modal travel times and observed choice for each of the three individuals in the sample is as follows:

<u>Individual #</u>	<u>Auto Travel Time</u>	<u>Bus Travel Time</u>	<u>Chosen Mode</u>
1	30 minutes	50 minutes	Car (mode 1)
2	20 minutes	10 minutes	Car (mode 1)
3	40 minutes	30 minutes	Bus (mode 2)

According to the logit model, the probabilities for the observed mode for each individual are:

$$Individual\ 1\ (P_{11}) = \frac{\exp(30\beta)}{\exp(50\beta) + \exp(30\beta)} = \frac{1}{1 + \exp(20\beta)} \quad 4.71$$

$$Individual\ 2\ (P_{12}) = \frac{\exp(20\beta)}{\exp(10\beta) + \exp(20\beta)} = \frac{1}{1 + \exp(-10\beta)} \quad 4.72$$

$$Individual\ 3\ (P_{23}) = \frac{\exp(30\beta)}{\exp(30\beta) + \exp(40\beta)} = \frac{1}{1 + \exp(10\beta)} \quad 4.73$$

The log-likelihood expression for this sample will be as follows:

$$\begin{aligned} LL &= \sum_{j=1,2} \sum_{t=1,2,3} \delta_{jt} \times \ln(P_{jt}) \\ &= 1 \times \ln(P_{11}) + 0 \times \ln(P_{21}) + 1 \times \ln(P_{12}) \\ &\quad + 0 \times \ln(P_{22}) + 0 \times \ln(P_{13}) + 1 \times \ln(P_{23}) \\ &= \ln(P_{11}) + \ln(P_{12}) + \ln(P_{23}) \end{aligned} \quad 4.74$$

A maximum likelihood estimator finds the value of the parameter β which maximizes the *log-likelihood* value. This solution is obtained by setting the first derivative of the log-likelihood function equal to zero, and solving for $\hat{\beta}$. However, for this illustrative example, we can plot the graph for log-likelihood (or likelihood) as a function of β to find the point where the maximum occurs. Figure 4.7 shows the graph of likelihood and log-likelihood as a function of β . It can be seen that the point where the maximum occurs is identical for both the likelihood and the log-likelihood functions, *i.e.*, at $\beta = -0.076$. This value is called the maximum likelihood estimate of β .

Practical applications of logit models include multiple parameters and many observations. Consequently, the maximum likelihood estimates for the parameters cannot be found graphically. Specialized software packages are available for this purpose.

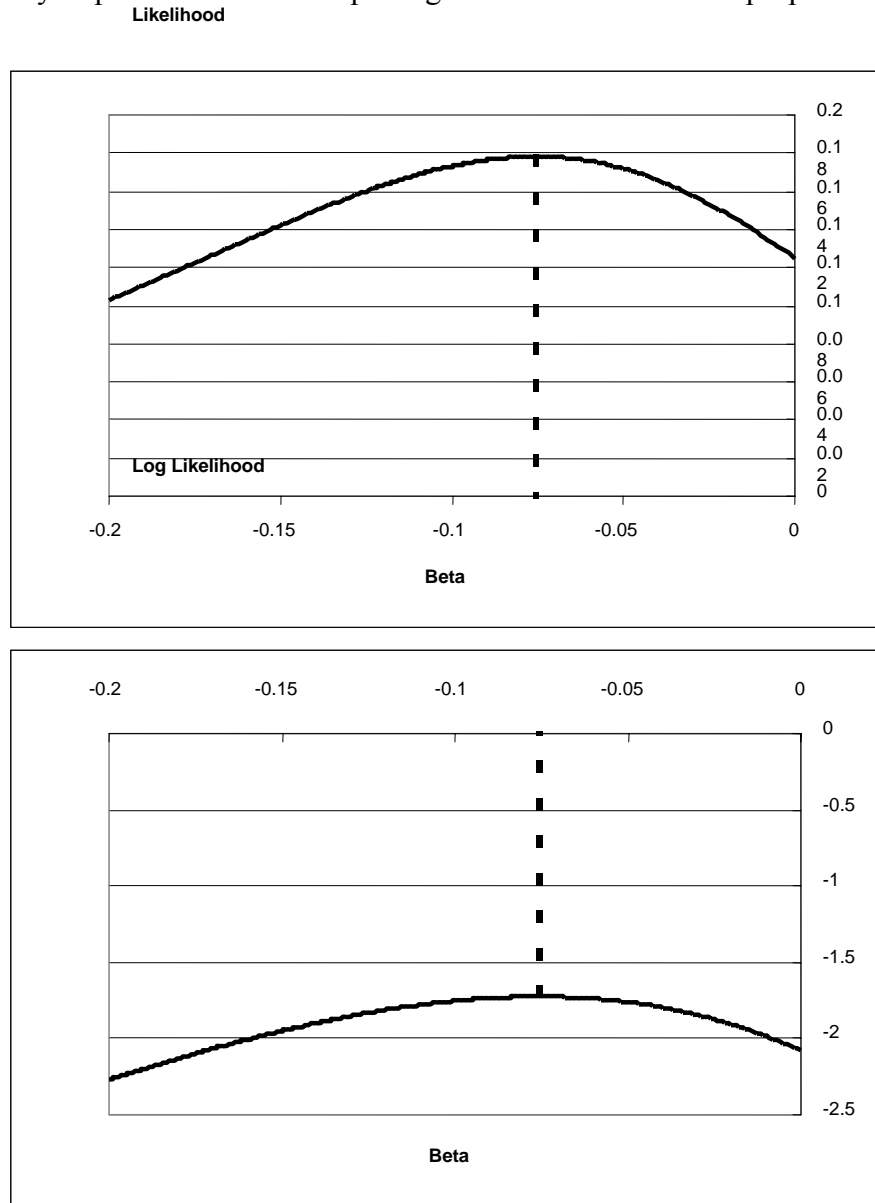


Figure 4.7 Likelihood and Log-likelihood as a Function of a Parameter Value

CHAPTER 5: Data Assembly and Estimation of Simple Multinomial Logit Model

5.1 Introduction

This chapter describes the estimation of basic specifications for the multinomial logit (MNL) model including the collection and organization of data required for model estimation. The chapter is organized as follows: Section 5.2 presents an overview of the data required to estimate mode choice models. Section 5.3 reviews data collection approaches for obtaining traveler and trip- related data. Section 5.4 discusses the methods for collecting data which describes the availability and service characteristics of the various modal alternatives. Section 5.5 illustrates two different data structures used by software packages for estimation of MNL and NL models. Section 5.6 describes the data used to estimate work mode choice model in the San Francisco Bay Region. Section 5.7 describes preliminary estimation results for mode choice to work based on this data and interpret these results in terms of judgment, descriptive measures and statistical tests. CHAPTER 6 extends this example to the estimation of more sophisticated models. CHAPTER 7 develops a parallel example for shop/other mode choice in the San Francisco Bay Region. CHAPTER 9 extends the examples from previous chapters and explores nested logit models for work and shop/other trips in the San Francisco Bay Region. Additional examples based on data collected in different urban regions are presented in the appendices.

5.2 Data Requirements Overview

The first step in the development of a choice model is to assemble data about traveler choice and the variables believed to influence that choice process. In the context of travel mode choice, such data include:

- Traveler and trip related variables that influence the travelers' assessment of modal alternatives (*e.g.*, income, automobile ownership, trip purpose, time of day of travel, origin and destination of trip, and travel party size),

- Mode related variables describing each alternative available to the traveler (e.g., travel time, travel cost and service frequency for carrier modes) and
- The observed or reported mode choice of the traveler (the “dependent” or “endogenous” variable).

The first two categories of variables are selected to describe the factors which influence each decision maker’s choice of an alternative. These independent or exogenous variables are likely to differ across trip purpose. The commonly used explanatory variables in mode choice models include:

- Traveler (Decision Maker) Related Variables
 - Income of traveler or traveler's household,
 - Number of automobiles in traveler's household,
 - Number of workers in traveler's household,
 - Sex of the traveler,
 - Age group of the traveler,
 - Functions of these variables such as number of autos divided by number of workers and
- Trip Context Variables
 - Trip purpose,
 - Employment density at the traveler's workplace,
 - Population density at the home location and
 - Dummy variable indicating whether the traveler's workplace is in the Central Business District (CBD).

- Mode (Alternative) Related Variables
 - Total travel time,
 - In-vehicle travel time,
 - Out-of-vehicle travel time,
 - Walk time,
 - Wait time,
 - Number of transfers
 - Transit headway and
 - Travel cost.
- Interaction of Mode and Traveler or Trip Related Variables
 - Travel cost divided by household income,
 - Travel time or cost interacted with sex or age group of traveler, and
 - Out-of-vehicle time divided by total trip distance.

5.3 Sources and Methods for Traveler and Trip Related Data Collection

Traveler and trip related data (including the actual mode choice of the traveler) needed for estimation of mode choice models are generally obtained by surveying a sample of travelers from the population of interest. This section discusses the types of surveys that may be used to obtain traveler and/or trip-related information (Section 5.3.1) and associated sample design considerations (Section 5.3.2).

5.3.1 Travel Survey Types

There are several types of travel surveys. The most common of these are household, workplace and intercept surveys.

Household Travel Surveys involve contacting respondents in their home and collecting information regarding their household characteristics (*e.g.*, number of members in household, automobile ownership, *etc.*), their personal characteristics (such as income, work status, *etc.*) and

the travel decisions made in the recent past (*e.g.*, number of trips, mode of travel for each trip, *etc.*). Historically, most household traveler surveys were conducted through personal interviews in the respondent's home. Currently, most household travel surveys are conducted using telephone or mail-back surveys, or a combination of both. It is common practice to include travel diaries as a part of the household travel survey. Travel diaries are a daily log of all trips (including information about trip origin and destination, start and end time, mode of travel, purpose at the origin and destination, *etc.*) made by each household member during a specified time period. This information is used to develop trip generation, trip distribution, and mode choice models for various trip purposes. Recently, travel diaries have been extended to include detailed information about the activities engaged in at each stop location and at home to provide a better understanding of the motivation for each trip and to associate trips of different purposes with different members of the household. Also, in some cases, diaries have been collected repeatedly from the same 'panel' of respondents to understand changes in their behavior over time.

Workplace Surveys involve contacting respondents at their workplace. The information collected is similar to that for household surveys but focuses exclusively on the traveler working at that location and on his/her work and work-related trips. Such surveys are of particular interest in understanding work commute patterns of individuals and in designing alternative commuter services.

Destination Surveys involve contacting respondents at other destinations. Similar information is collected as for workplace surveys but the objective is to learn more about travel to other types of destinations and possibly to develop transportation services which better serve such destinations.

Intercept Surveys "intercept" potential respondents during their travel. The emphasis of the survey is on collecting information about the specific trip being undertaken by the traveler. Intercept surveys are commonly used for intercity travel studies due to the high cost of identifying intercity travelers through home-based or work-based surveys. In intercept surveys, travelers are intercepted at a roadside rest area for highway travel and on board carriers (or at carrier terminals) for other modes of travel. The traveler is usually given a brief survey (paper or

interview) for immediate completion or future response and/or recruited for a future phone survey. A variant of highway intercept surveys is to record the license plate of vehicles and subsequently contact the owners of a sample of vehicles to obtain information on the trip that was observed. Intercept surveys can be used to cover all available modes or they can be used to enrich a household or workplace survey sample by providing additional observations for users of infrequently used modes since few such users are likely to be identified through household or workplace surveys.

5.3.2 Sampling Design Considerations

The first issue to consider in sample design is the population of interest in the study. Obviously, the population of interest will depend on the purpose of the study. In the context of urban work mode choice analysis, the population of interest would be all commuters in the urban region. In the context of non-business intercity mode choice, the population of interest would be all non-business intercity travelers in the relevant corridor. However, most surveys are designed to address a number of current or potential analysis and decision issues. Thus, the population of interest is selected based on the full range of such issues and is likely to include a wide range of households and household members each undertaking a variety of trips.

After identifying the population of interest, the next step in sample design is to determine the unit for sampling. The sampling units should be mutually exclusive and collectively exhaust the population. Thus, if the population of interest is commuters in an urban region, the sampling unit could be firms in the area (whose employees collectively represent the commuting population). Within each firm, all or a sample of workers could be interviewed. However, if the survey is addressed to multiple urban travel issues, the population of interest is likely to be all individuals or households resident in the urban region. A comprehensive list of all sampling units constitutes the sampling frame. This may be a directory of firms or households obtained from a combination of public and private sources such as local or regional commerce departments, business directories, tourism offices, utilities, *etc.* Of course, the sampling frame may not always completely represent the population of interest. For example, some commuters

may be employed by firms outside the urban region and some households may not have telephones or other utility connections.

Once the sampling frame is determined, a sampling design is used to select a sample of cases. The most common class of sampling procedures is probability sampling, where each sampling unit has a pre-determined probability of being selected into the survey sample.

A probability sampling procedure in which each sampling unit has an equal probability of being selected is simple random sampling. In this method, the sampling units are selected randomly from the sampling frame. This would apply to household sampling, for example. An alternative approach for work place sampling would be to sample each employer to further sample workers for those employers sampled. This is referred to as a two-stage sample. In either case, the result is an unbiased sample which is representative of the population of interest. Simple random sampling offers the advantage of being easy to understand, communicate, and implement in the field, making it less prone to errors. However, a random sample may not adequately represent some population segments of interest.

This problem can be addressed by using stratified random sampling. This entails partitioning the sampling frame into several mutually exclusive and collectively exhaustive segments (or strata) based on one or more stratification variables, followed by random sampling within each stratum. Thus, households in a region may be stratified by location within the region; city vs. suburban residence; or income. Similarly, employment firms in a region may be stratified by the number of employees or the type of business. Stratified random sampling may be useful in cases where it is important to understand the characteristics of certain subpopulations as well as the overall population. For example, it might be useful to study the mode choice behavior of employees working for very small employers in a region and compare their behavior with those of employees working for large employers. But if the number of small employers in a region is a small proportion of all employers, they may not be well-represented in a simple random sample. In such a case, the analyst might stratify the sampling frame to ensure adequate representation of very small employers in the survey sample. Stratified random sampling can also be less expensive than simple random sampling. This can occur, for example, because per-person sampling is less expensive if a few large employers are targeted rather than

several smaller ones. Finally, increased dispersion in the range of important independent variables can be increased through “clever” identification of alternative strata based on income, household size, number of vehicles owned or number of workers providing benefits in estimation efficiency.

The use of stratified random sampling does not present any new problems in estimation as long as the stratification variable(s) is (are) exogenous to the choice process. However, in some situations, one may want to use the discrete choice of interest (the dependent or endogenous variable) as the stratification variable. For example, if the bus mode is rarely chosen for urban or intercity travel, a random sampling procedure or even an exogenously stratified sample might not provide sufficient observations of bus riders to understand the factors that affect the choice of the bus mode. In this case, bus riders can be intentionally over sampled by interviewing them at bus stops or on board the buses. Such choice based sampling may also be motivated by cost considerations. Special techniques are required to estimate model parameters using choice based sampling methods.

A further problem is that it may be difficult to obtain a random sample as it may not be possible to get an up to date and complete list of potential respondents from which to select a sample. This is the primary reason for using "deterministic rules" of sampling. In systematic sampling, sample units are drawn by deterministic rather than random rules. For example, rather than drawing a random sample of 5 percent all households in a region from a published directory, a systematic sampling plan may pick out every 20th household in the directory. As long as there is no inherent bias in setting up the deterministic rule, the systematic sampling plan is essentially equivalent to random sampling, and the choice of one over the other may be a matter of operational convenience.

In addition to the sampling approaches discussed above, combinations of the approaches can also be used. For example, enriched sampling uses a combination of household or workplace sampling with choice based sampling to increase the number of transit users in the study data. As with choice based sampling, special estimation techniques are required to offset the biases associated with this sampling approach.

Another issue in survey sampling is the sample size to be used. The size of the sample required to adequately represent the population is a function of the level of statistical accuracy and confidence desired from the survey. The precision of parameter estimates and the statistical validity of estimation results improve with sample size. However, the cost of the survey also increases with sample size and in many cases it is necessary to restrict the sample size to ensure that the cost of the survey remains within budgetary constraints. The decision about sample size requires a careful evaluation of the need for adequate data to satisfy study objectives *vis-à-vis* budgetary constraints. Interested readers are referred to Ben-Akiva and Lerman (1985; Chapter 8) or Börsch-Supan (1987) for a more comprehensive discussion of sampling methods and sample size issues.

5.4 Methods for Collecting Mode Related Data

Surveys can be used to collect information describing the trip maker and his or her household, the context of the trip (purpose, time of day, frequency of travel, origin, and destination), the chosen mode and the respondent's perception of travel service. However, objective data about modal service (mode availability and level of service) must be obtained from other sources. Modal data is usually generated from simulation of network service characteristics including carrier schedules and fare and observed volumes, travel times and tolls on roadway links.

Network analysis provides the zone-to-zone in-vehicle travel times for the highway (non-transit) and transit modes (the urban area is divided into several traffic analysis zones for travel demand analysis purposes). The highway out-of-vehicle time by the highway mode is assigned a nominal value to reflect walk access/egress to/from the car. The transit out-of-vehicle time is based on transit schedules, the presence or absence of transfers, and location of bus stops *vis-à-vis* origin and destination locations. The highway distance of travel is obtained from the network structure and a per-mile vehicle operation cost is applied to obtain highway zone-to-zone driving costs. Parking costs are obtained from per-hour parking rates at the destination zone multiplied by half the estimated duration of the activity pursued at the destination zone (allowing half the parking cost to be charged to the incoming and departing trips) and toll costs can be

identified from the network. Transit cost is determined from the actual fare for travel between zones plus any access costs to/from the transit station from/to the origin/destination. The highway in-vehicle times, out-of-vehicle times, and costs are used as the relevant values for driving alone and these values are modified appropriately to reflect increased in-vehicle times and decreased driving/parking costs for the shared ride modes. The travel times for non-motorized modes (such as bike and walk) are obtained from the zone-to-zone distance and an assumed walk/bike speed. Finally, the appropriate travel times and cost between zones is appended to each trip in the trip file based on the origin and destination zones of the trip.

5.5 Data Structure for Estimation

The data collected from the various sources described in the previous sections must be assembled into a single data set to support model estimation. This can be accomplished using a variety of spreadsheet, data base, statistical software or user prepared programs. The structure of the resultant data files must satisfy the format requirements of the software packages designed for choice model estimation. The commonly used software packages for discrete choice model estimation require the data to be structured in one of two formats: a) the trip format or b) the trip-alternative format. These are commonly referred to as IDCASE (each record contains all the information for mode choice over alternatives for a single trip) or IDCASE-IDALT (each record contains all the information for a single mode available to each trip maker so there is one record for each mode for each trip).

In the trip format, each record provides all the relevant information about an individual trip, including the traveler/trip related variables, mode related variables for all available modes and a variable indicating which alternative was chosen. In the trip-alternative format, each record includes information on the traveler/trip related variables, the attributes of that modal alternative, and a choice variable that indicates whether the alternative was or was not chosen¹⁶.

The two data structures are illustrated in Figure 5.1 for four observations (individuals) with up to three modal alternatives available. The first column in both formats is the trip

¹⁶ The data formats required vary for different programs and are documented in the relevant software.

number. In the trip (or IDCASE) format, this is followed by traveler/trip related variables (e.g., income), the level of service (time and cost in the figure) variables associated with each alternative and a variable that indicates the chosen alternative. In the trip-alternative (or IDCASE-IDAIt) format, the second column generally identifies the alternative number with which that record is associated. Additional columns will generally include a 0-1 variable indicating the chosen alternative¹⁷, the number of alternatives available, traveler/trip related variables mode related attributes for the alternative with which the record is associated. Data is displayed in these two formats in Figure 5.1.

¹⁷ When the same behavior is observed repeatedly for a case, the chosen alternative will be replaced by the frequency with which each alternative is chosen.

Data Layout Type I: Trip Format									
<i>Trip</i>		<i>Alternative 1</i>		<i>Alternative 2</i>		<i>Alternative 3</i>		<i>Alternative</i>	
<i>Number</i>	<i>Income</i>	<i>Time</i>	<i>Cost</i>	<i>Time</i>	<i>Cost</i>	<i>Time</i>	<i>Cost</i>	<i>Chosen</i>	
1	30000	30	150	40	100	20	200	1	
2	30000	25	125	35	100	0	0	2	
3	40000	40	125	50	75	30	175	3	
4	50000	15	225	20	150	10	250	3	
Data Layout Type II: Trip-Alternative Format									
<i>Trip</i>	<i>Alternative</i>	<i>Number of</i>	<i>Income</i>	<i>Time</i>	<i>Cost</i>	<i>Alternative</i>			
<i>Number</i>	<i>Number</i>	<i>Alternatives</i>				<i>Chosen</i>			
1	1	3	30000	30	150	1			
1	2	3	30000	40	100	0			
1	3	3	30000	20	200	0			
2	1	2	30000	25	125	0			
2	2	2	30000	35	100	1			
3	1	3	40000	40	125	0			
3	2	3	40000	50	75	0			
3	3	3	40000	30	175	1			
4	1	3	50000	15	225	0			
4	2	3	50000	20	150	0			
4	3	3	50000	10	250	1			

Figure 5.1 Data Structure for Model Estimation

The unavailability of an alternative is indicated in the trip format by zeros for all attribute variables of the unavailable alternative and in the trip-alternative format by excluding the record for the unavailable alternative. Thus, the second individual in the sample of Figure 5.1 has only two alternatives available.

Either of the two data formats may be used to represent the information required for model estimation. The choice is based on the programming decisions of the software developer taking into account data storage and computational implications of each choice.

5.6 Application Data for Work Mode Choice in the San Francisco Bay Area

The examples used in previous chapters are based on simulated or hypothetical data to highlight fundamental concepts of multinomial logit choice models. Henceforth, our discussion of model specification and interpretation of results will be based on application data sets assembled from travel survey and other data collection to support transportation decision making in selected urban regions. Use of real data provides richer examples and “hands-on” experience in estimating mode choice models. The data used in this chapter was collected for the analysis of work trip mode choice in the San Francisco Bay Area in 1990.

The San Francisco Bay Area work mode choice data set comprises 5,029 home-to-work commute trips in the San Francisco Bay Area. The data is drawn from the San Francisco Bay Area Household Travel Survey conducted by the Metropolitan Transportation Commission (MTC) in the spring and fall of 1990 (see White and Company, Inc., 1991, for details of survey sampling and administration procedures). This survey included a one day travel diary for each household member older than five years and detailed individual and household socio-demographic information.

There are six work mode choice alternatives in the region: drive alone, shared-ride with 2 people, shared ride with 3 or more people, transit, bike, and walk¹⁸. The drive alone mode is available for a trip only if the trip-maker's household has a vehicle available and if the trip-maker has a driver's license. The shared-ride modes (with 2 people and with 3 or more people) are available for all trips. Transit availability is determined based on the residence and work zones of individuals. The bike mode is deemed available if the one-way home-to-work distance is less than 12 miles, while the walk mode is considered to be available if the one-way home to work distance is less than 4 miles (the distance thresholds to determine bike and walk availability are determined based on the maximum one-way distance of bike and walk-users, respectively).

Level of service data were generated by the Metropolitan Transportation Commission for each zone pair and for each mode. These data were appended to the home-based work trips

¹⁸ The estimation reported by the Metropolitan Transportation Commission (Travel Demand Models for the San Francisco Bay Area (BAYCAST-90): Technical Summary, Metropolitan Transportation Commission, Oakland, California, June 1997) includes drive alone, shared-ride with 2 people, shared ride with 3 or more people, transit with walk access, transit with auto access, bike and walk.

based on the origin and destination of each trip. The data includes traveler, trip related area and mode related variables for each trip including in-vehicle travel time, out-of-vehicle travel time, travel cost, travel distance, and a mode availability indicator.

Table 5-1 provides information about the availability and usage of each alternative and the average values of in-vehicle time, out-of-vehicle time and travel cost in the sample. Drive alone is available to most work commuters in the Bay Area and is the most frequently chosen alternative. The shared-ride modes are available for all trips (by construction) and together account for the next largest share of chosen alternatives. The combined total of drive alone and shared ride trips represent close to 85% of all work trips. Transit trips constitute roughly 10% of work trips, a substantially greater share than in most metropolitan regions in the U.S. The fraction of trips using non-motorized modes (walk and bike) constitutes a small but not insignificant portion of total trips.

Table 5-1 Sample Statistics for Bay Area Journey-to-Work Modal Data

Mode	Fraction of Sample with Mode Available	Market Share	Average IVTT (minutes)	Average OVTT (minutes)	Average Cost (1990 cents)
1. Drive Alone	94.6%	72.3%	21.0	3.8	176
2. Shared Ride (2)	100.0%	10.3%	25.0	3.9	89
3. Shared Ride (3+)	100.0%	3.2%	27.0	3.9	50
4. Transit	79.6%	9.9%	24.5	28.8	123
5. Bike	34.6%	1.0%	28.0	3.7	-
6. Walk	29.4%	3.3%	-	49.0	-

5.7 Estimation of MNL Model with Basic Specification

We use the San Francisco Bay Area data to estimate a multinomial logit work mode choice model using a basic specification which includes travel time, travel cost and household income as the explanatory variables. Travel time and travel cost represent mode related attributes; all other things being equal, a faster mode of travel is more likely to be chosen than a slower mode and a less expensive mode is more likely to be chosen than a costlier mode. Household income is included in the model with the expectation that travelers from high income households are more likely to drive alone than to use other travel modes.

The multinomial logit work trip mode choice models are estimated using ALOGIT, LIMDEP and ELM software, to illustrate differences in the outputs of these packages¹⁹, as well as a specially programmed module for Matlab. The data sets, input control files, and estimation output results for this a selection of other model specifications for both software packages are included in the CD-ROM supplied with this manual. The Matlab module code, command files and output for all specifications in the manual are also included in the CD-ROM.

The travel time (TT) and travel cost (TC) variables are specified as generic in this model. This implies that an increase of one unit of travel time or travel cost has the same impact on modal utility for all six modes. Household Income (Inc) is included as an alternative-specific variable. The drive alone mode is considered the base alternative for household income and the modal constants (see section 4.1.2.2 for a discussion of the need for a base alternative for these variables).

The following mode labels are used in the subsequent discussion and equations: DA (drive alone), SR2 (shared ride with 2 people), SR3+ (shared ride with 3 or more people), TR (transit), BK (bike) and WK (walk). The deterministic portion of the utility for these modes, based on the utility specification discussed above, may be written as:

¹⁹ Essentially identical estimation results are produced by these and a variety of other commercially available and programmer developed estimation procedures. However, some software applies simplifying assumptions that are not appropriate in every case. See Appendix A for additional information.

$$V_{DA} = \beta_1 \times TT_{DA} + \beta_2 \times TC_{DA} \quad 5.1$$

$$V_{SR2} = \beta_{SR2} + \beta_1 \times TT_{SR2} + \beta_2 \times TC_{SR2} + \gamma_{SR2} \times Inc \quad 5.2$$

$$V_{SR3+} = \beta_{SR3+} + \beta_1 \times TT_{SR3+} + \beta_2 \times TC_{SR3+} + \gamma_{SR3+} \times Inc \quad 5.3$$

$$V_{TR} = \beta_{TR} + \beta_1 \times TT_{TR} + \beta_2 \times TC_{TR} + \gamma_{TR} \times Inc \quad 5.4$$

$$V_{BK} = \beta_{BK} + \beta_1 \times TT_{BK} + \beta_2 \times TC_{BK} + \gamma_{BK} \times Inc \quad 5.5$$

$$V_{WK} = \beta_{WK} + \beta_1 \times TT_{WK} + \beta_2 \times TC_{WK} + \gamma_{WK} \times Inc \quad 5.6$$

The estimation results reported in this manual are obtained from software programmed in the Matlab language and included in the CD-Rom distributed with the manual. The output for the above model specification is shown in structured format in Table 5-2. The corresponding estimation results for this specification from various commercial packages; ALOGIT, LIMDEP and ELM; are reported in Appendix A. The outputs from these and other software package typically include, at least, the following estimation results:

- Parameter²⁰ names, parameter estimates, standard errors of these estimates and the corresponding t-statistics for each variable/parameter;
- Log-likelihood values at zero (equal probability model), constants only (market shares model) and at convergence and
- Rho-Squared and other indicators of goodness of fit.

In addition, different software reports a variety of other information either as part of the default output or as a user selected option. These include:

- The number of observations,
- The number of cases for which each alternative is available,
- The number of cases for which each alternative is chosen,
- The number of iterations required to obtain convergence, and
- The status of the convergence process at each iteration.

The value for the log-likelihood at zero and constants can be obtained for either software by estimating models without (zero) and with (constants) alternative specific constants²¹ and no

²⁰ Commonly, the name is selected to match the corresponding variable name.

²¹ The zero model can be estimated with constants included but restricted to zero, the constants model can be estimated with constants included and not restricted.

other variables. Further, the log-likelihood at zero can be calculated directly as

$$\sum_t \ln(1 / NAlt_t).$$

In the next three sections, we review the estimation results for the base model using informal judgment-based tests (section 5.7.1), goodness-of-fit measures (section 5.7.2), and statistical tests (section 5.7.3). These elements, taken together, provide a basis to evaluate each model and to compare models with different specifications.

Table 5-2 Estimation Results for Zero Coefficient, Constants Only and Base Models

Variables	Zero Coefficients Model	Constants Only Model	Base Model
Travel Cost (1990 cents)			-0.0049 (-20.6)
Total Travel Time (minutes)			-0.0513 (-16.6)
Income (1,000's of 1990 DOLLARS)			
Drive Alone (Base)			0.0
Shared Ride 2			-0.0022 (-1.4)
Shared Ride 3+			0.0004 (0.1)
Transit			-0.0053 (-2.9)
Bike			-0.0128 (-2.4)
Walk			-0.0097 (-3.2)
Mode Constants			
Drive Alone (base)		0.0	0.0
Shared Ride 2		-2.137 (-44.1)	-2.178 (-20.8)
Shared Ride 3+		-3.303 (-40.6)	-3.725 (-21.0)
Transit		-1.950 (-38.5)	-0.6709 (-5.1)
Bike		-3.334 (-23.1)	-2.376 (-7.8)
Walk		-2.040 (-23.9)	-0.2068 (-1.1)
Log-likelihood at Zero		-7309.601	-7309.601
Log-likelihood at Constant			-4132.916
Log-likelihood at Convergence	-7309.601	-4132.916	-3626.186
Rho-Squared w.r.t. Zero	NA	0.4346	0.5039
Rho-Squared w.r.t. Constants	NA	NA	0.1226

5.7.1 Informal Tests

A variety of informal tests can be applied to an estimated model. These tests are designed to assess the reasonableness of the implications of estimated parameters. The most common tests concern:

- The sign of parameters (do the associated variables have a positive or negative effect on the alternatives with which they are associated?),
- The difference (positive or negative) within sets of alternative specific variables (does the inclusion of this variable have a more or less positive effect on one alternative relative to another?) and
- The ratio of pairs of parameters (is the ratio between the parameters of the correct sign and in a reasonable range?).

5.7.1.1 Signs of Parameters

The most basic test of the estimation results is to examine the signs of the estimated parameters with theory, intuition and judgment regarding the expected impact of the corresponding variables. The estimated coefficients on the travel time and cost variables in Table 5-2 are negative, as expected, implying that the utility of a mode decreases as the mode becomes slower and/or more expensive. This, in turn, will reduce the choice probability of the corresponding mode.

5.7.1.2 Differences in Alternative Specific Variable Parameters across Alternatives

We often have expectations about the impact of decision-maker characteristics on different alternatives. For example, when analyzing mode choice, we expect a number of variables to be more positive for automobile alternatives, especially Drive Alone, than for other alternatives. These include income, automobile ownership, home ownership, single family dwelling unit, etc. Since DA is the reference alternative in these models, we expect negative parameters on all alternative specific income variables, with small values for the shared ride alternatives and larger values for other alternatives, to reflect our intuition that increasing income will be associated with decreased preference for all other alternatives relative to drive alone. All the estimated

alternative specific income parameters in Table 5-2 Estimation Results for Zero Coefficient, Constants Only and Base Models are negative, as expected, with the exception of the parameter for shared ride 3+. The positive sign for the shared ride 3+ parameter is counter-intuitive but very small and not significantly different from zero. An approach to address this problem is described in section 5.7.3.1.

Additional informal tests involve comparisons among the estimated income parameters. The differences in the magnitude of these alternative-specific income parameters indicate the relative impact of increasing income on the utility and, hence, the choice probability of each mode. The results reported in Table 5-2 show that an increase in income will have a larger negative effect on the utilities of the non-motorized modes (bike and walk) than on those of transit and shared ride 2 modes. It is important to understand that the change in the choice probability of that alternative is not determined by the sign of the income parameter for a particular alternative but on the value of the parameter relative to the weighted mean of the income parameters across all alternatives, as described in section 4.5. For example, in the base model specification (Table 5-2), an increase in income will increase the probability of choosing drive alone and shared ride 3+ (the alternatives with the most positive parameters²²) and will decrease the probability of choosing walk and bike (the alternatives with the smallest, most negative, parameters). However, the effect on shared ride 2 and transit is unclear; as income increases, they will lose some probability to drive alone and shared ride 3+ and gain some from walk and bike. The net effect will depend on the difference between the parameter for the alternative of interest and the individual choice probability weighted average of all the alternative specific income parameters (including zero for the base alternative) as illustrated in equation 4.58.

5.7.1.3 The Ratio of Pairs of Parameters

The ratio of the estimated travel time and travel cost parameters provides an estimate of the value of time implied by the model; this can serve as another important informal test for evaluating the reasonableness of the model. For example, in the Base Model reported in Table

²² If all the reported constants are negative, the alternative with the most positive parameter is drive alone at zero.

5-2, the implied value of time is $(-0.05134)/(-0.00492)$ or 10.4 cents per minute (the units are determined from the units of the time and cost variables used in estimation). This is equivalent to \$6.26 per hour which is much lower than the average wage rate in the area at the time of the survey, approximately \$21.20 per hour, suggesting that the estimated value of time may be too low. We revisit this issue in greater detail in the next chapter.

Similar ratios may be used to assess the reasonableness of the relative magnitudes of other pairs of parameters. These include out of vehicle time relative to in vehicle time, travel time reliability (if available) relative to average travel time, etc.

5.7.2 Overall Goodness-of-Fit Measures

This section presents a descriptive measure, called the rho-squared value (ρ^2) which can be used to describe the overall goodness of fit of the model. We can understand the formulation of this value in terms of the following figure which shows the scalar relationship among the log-likelihood values for a zero coefficients model (or the equally likely model), a constants only or market share model, the estimated model and a perfect prediction model. In the figure, $LL(0)$ represents the log-likelihood with zero coefficients (which results in equal likelihood of choosing each available alternative), $LL(C)$ represents the log-likelihood for the constants only model, $LL(\hat{\beta})$ represents the log-likelihood for the estimated model and $LL(*) = 0$ is the log-likelihood for the perfect prediction model.



Figure 5.2 Relationship between Different Log-likelihood Measures

The relationships among modeling results will always appear in the order shown provided the estimated model includes a full set of alternative specific constants. That is, the constants only model is always better than the equally likely model, and the estimated model with constants is

always better than the constants only model. The order of different estimated models will vary except that any model which is a restricted version of another model will be to the left of the unrestricted model.

The rho-squared (ρ_0^2) value is based on the relationship among the log-likelihood values indicated in Figure 5.2. It is simply the ratio of the distance between the reference model and the estimated model divided by the difference between the reference model and a perfect model. If the reference model is the equally likely model, the rho square with respect to zero, ρ_0^2 , is:

$$\rho_0^2 = \frac{LL(\hat{\beta}) - LL(0)}{LL(*) - LL(0)} \quad 5.7$$

Since the log-likelihood value for the perfect model is zero²³, the ρ_0^2 measure reduces to:

$$\rho_0^2 = 1 - \frac{LL(\hat{\beta})}{LL(0)} \quad 5.8$$

Similarly, the rho-square with respect to the constants only model is:

$$\begin{aligned} \rho_c^2 &= \frac{LL(\hat{\beta}) - LL(c)}{LL(*) - LL(c)} \\ &= 1 - \frac{LL(\hat{\beta})}{LL(c)} \end{aligned} \quad 5.9$$

By definition, the values of both rho-squared measures lie between 0 and 1 (this is similar to the R^2 measure for linear regression models). A value of zero implies that the model is no better than the reference model, whereas a value of one implies a perfect model; that is, every choice is predicted correctly.

The rho-squared values for the basic model specification in Table 5-2 are computed based on the formulae shown below as:

$$\rho_0^2 = 1 - \frac{(-3626.2)}{(-7309.6)} = 0.5039 \quad \rho_c^2 = 1 - \frac{(-3626.2)}{(-4132.9)} = 0.1226 \quad 5.10$$

²³ The perfect model “predicts” a probability of one for the chosen mode for each and every case so the contribution of each case to the log-likelihood function is $\ln(1) = 0$.

The rho-squared measures are widely used to describe the goodness of fit for choice models because of their intuitive formulation. The ρ_0^2 measures the improvement due to all elements of the model, including the fit to market shares, which is not very interesting for disaggregate analysis so it should not be used to assess models in which the sample shares are very unequal. The rho-squared measure with respect to the constant only model, ρ_c^2 , controls for the choice proportions in the estimation sample and is therefore a better measure to use for evaluating models.

A problem with both rho-squared measures is that there are no guidelines for a “good” rho-squared value. Consequently, the measures are of limited value in assessing the quality of an estimated model and should be used with caution even in assessing the relative fit among alternative specifications. It is preferable to use the log-likelihood statistic (which has a formal and convenient mechanism to test among alternative model specifications) to support the selection of a preferred specification among alternative specifications.

Another problem with the rho-squared measures is that they improve no matter what variable is added to the model independent of its importance. This directly results from the fact that the objective function of the model is being modeled with one or more additional degrees of freedom and that the same data that is used for estimation is used to assess the goodness of fit of the model. One approach to this problem is to replace the rho-squared measure with an adjusted rho-square measure which is designed to take account of these factors. The adjusted rho-squared for the zero model is given by:

$$\begin{aligned}\bar{\rho}_0^2 &= \frac{[LL(\hat{\beta}) - K] - LL(0)}{LL(*) - LL(0)} \\ &= 1 - \frac{LL(\hat{\beta}) - K}{LL(0)}\end{aligned}\tag{5.11}$$

where K is the number of degrees of freedom (parameters) used in the model.

The corresponding adjusted rho-squared for the constants only model is given by:

$$\begin{aligned}\bar{\rho}_C^2 &= \frac{LL(\hat{\beta}) - K - (LL(C) - K_{MS})}{LL(*) - (LL(C) - K_{MS})} \\ &= 1 - \frac{LL(\hat{\beta}) - K}{LL(C) - K_{MS}}\end{aligned}\tag{5.12}$$

where K_{ms} is the number of degrees of freedom (parameters) used in the constants only model.

5.7.3 Statistical Tests

Statistical tests may be used to evaluate formal hypotheses about individual parameters or groups of parameters taken together. In this section we describe each of these tests.

5.7.3.1 Test of Individual Parameters

There is sampling error associated with the model parameters because the model is estimated from only a sample of the relevant population (the relevant population includes all commuters in the Bay Area). The magnitude of the sampling error in a parameter is provided by the standard error associated with that parameter; the larger the standard error, the lower the precision with which the corresponding parameter is estimated. The standard error plays an important role in testing whether a particular parameter is equal to some hypothesized value, as we discuss next.

The statistic used for testing the null hypothesis that a parameter $\hat{\beta}_k$ is equal to some hypothesized value, β_k^* , is the asymptotic t-statistic, which takes the following form:

$$t - statistic = \frac{\hat{\beta}_k - \beta_k^*}{S_k}\tag{5.13}$$

where $\hat{\beta}_k$ is the estimate for the k^{th} parameter,
 β_k^* is the hypothesized value for the k^{th} parameter and
 S_k is the standard error of the estimate.

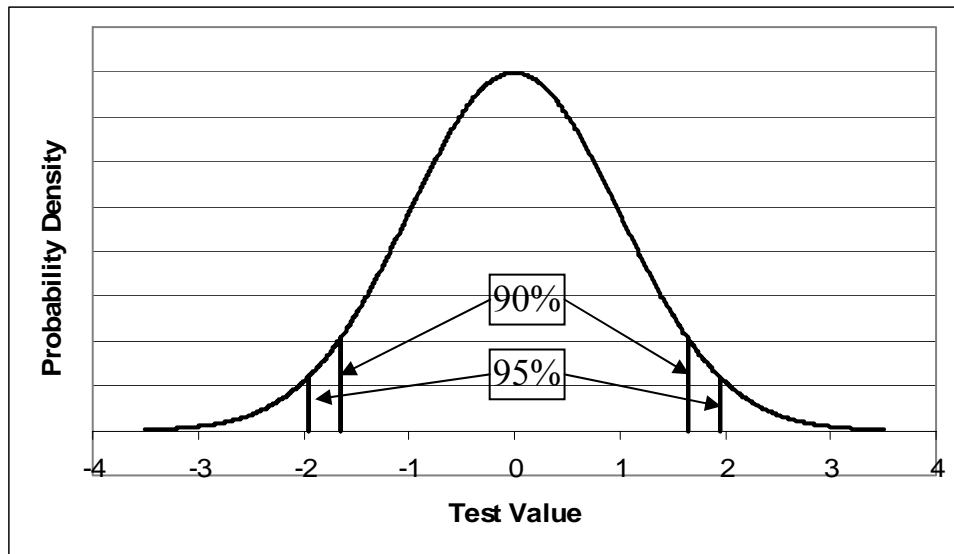
Sufficiently large absolute values of the *t*-statistic lead to the rejection of the null hypothesis that the parameter is equal to the hypothesized value. When the hypothesized value, β_k^* , is zero, the *t*-statistic becomes the ratio of the estimated parameter to the standard error. The default estimation output from most software packages includes the *t*-statistic for the test of the hypothesis that the true value is zero. The rejection of this null hypothesis implies that the corresponding variable has a significant impact on the modal utilities and suggests that the variable should be retained in the model. Low absolute values of the *t*-statistic imply that the variable does not contribute significantly to the explanatory power of the model and can be considered for exclusion. If it is concluded that the hypothesis is not rejected, the equality constraint can be incorporated in the model by creating a new variable $X_{kl} = X_k + X_l$.

The selection of a critical value for the *t*-statistic test is a matter of judgment and depends on the level of confidence with which the analyst wants to test his/her hypotheses. The critical *t*-values for different levels of confidence for samples sizes larger than 150 (which is the norm in mode choice analysis) are shown in Table 5-3. It should be apparent that the critical *t*-value increases with the desired level of confidence. Thus, one can conclude that a particular variable has no influence on choice (or equivalently that the true parameter associated with the variable is zero) can be rejected at the 90% level of confidence if the absolute value of the *t*-statistic is greater than 1.645 and at the 95% level of confidence if the *t*-statistic is greater than 1.960²⁴.

²⁴ Confidence levels commonly used are in the range of 90% to 99%.

Table 5-3 Critical t-Values for Selected Confidence Levels and Large Samples

Confidence Level	Critical t-value (two-tailed test)
90%	1.645
95%	1.960
99%	2.576
99.5%	2.810
99.9%	3.290

**Figure 5.3 t-Distribution Showing 90% and 95% Confidence Intervals**

We illustrate the use of the t-test by reviewing the t-statistics for each parameter in the initial model specification (Table 5-2). Both the travel cost and travel time parameters have large

absolute t-statistic values (20.6 and 16.6, respectively) which lead us to reject the hypothesis that these variables have no effect on modal utilities at a confidence level higher than 99.9%. Thus, these variables should be retained in the model. All of the other t-statistics, except for Income-Shared Ride 2, Income-Shared Ride 3+ and the walk constant are greater than 1.960 (95% confidence) supporting the inclusion of the corresponding variables. The t-statistics on the shared ride specific income variables are even less than 1.645 in absolute value (90% confidence), suggesting that the effect of income on the utilities of the shared ride modes may not differentiate them from the reference (drive alone) mode. Consequently, the analyst should consider removing these income variables from the utility function specifications for the shared ride modes. The case is particularly compelling for removal of the income shared ride 3+ variable since the t-statistic is very low and the parameter has a counter-intuitive sign. Another alternative would be to combine the two shared ride income parameters suggesting that income has differential effect between drive alone and shared ride but not between shared ride 2 and shared ride 3⁺ (when this is done, the combined variable obtains a small negative parameter which is not statistically different from zero). This variable could be deleted or retained according to the judgment of the analyst as described in Section 6.2.1.

An alternative approach is to report the t-statistic to two or three decimal places and calculate the probability that a t-statistic value of that magnitude or higher would occur due to random variation in sampling as shown in Table 5-4. This is reported as the significance level, which is the complement of the level of confidence. The significance of each parameter can be read directly from the table. Parameters with significance greater than 0.05 (lower in magnitude but more significant), provide a stronger basis for rejecting the hypothesis that the true parameter is zero and that the corresponding variable can be excluded from the model. On the other hand, significance levels of 0.163 (for Income-SR2), 0.888 (for Income-SR3+) and 0.287 (for ASC-Walk) provide little evidence about whether the corresponding variable should or should not be included in the model²⁵.

²⁵ Generally, it is good policy to retain or eliminate full sets of constants and alternative specific variables unless there is a good reason to do otherwise until all other variables have been selected.

Table 5-4 Parameter Estimates, t-statistics and Significance for Base Model

Variables	Parameter Estimate	t-statistic	Significance Level
Travel Cost (1990 cents)	-0.0049	-20.60	0.0000 ²⁶
Total Travel Time (minutes)	-0.0513	-16.57	0.0000
Income (1,000's of 1990 dollars)			
Drive Alone (Base)	0.0	-----	-----
Shared Ride 2	-0.0022	-1.40	0.1628
Shared Ride 3+	0.0004	0.14	0.8879
Transit	-0.0053	-2.89	0.0039
Bike	-0.0128	-2.41	0.0164
Walk	-0.0097	-3.19	0.0015
Mode Constants			
Drive Alone (base)	0.0	-----	-----
Shared Ride 2	-2.178	-20.82	0.0000
Shared Ride 3+	-3.725	-21.96	0.0000
Transit	-0.6709	-5.106	0.0000
Bike	-2.376	-7.80	0.0000
Walk	-0.2068	-1.06	0.2867

It is important to recognize that a low t-statistic does not require removal of the corresponding variable from the model. If the analyst has a strong reason to believe that the variable is important, and the parameter sign is correct, it is reasonable to retain the variable in the model. A low t-statistic and corresponding low level of significance can best be interpreted as providing little or no information rather than as a basis for excluding a variable. Also, one should be cautious about prematurely deleting variables which are expected to be important as the same variable may be significant when other variables are added to or deleted from the model.

The lack of significance of the alternative specific walk constant is immaterial since the constants represent the average effect of all the variables not included in the model and should always be retained despite the fact that they do not have a well-understood behavioral interpretation.

²⁶ Significance levels reported as 0.0000 are equivalent to less than 0.00005

5.7.3.2 Test of Linear Relationship between Parameters

It is often interesting to determine if two parameters are statistically different from one another or if two parameters are related by a specific value ration. These tests are similarly based on the t-statistic; however, the formulation of the test is somewhat different from that described in Section 5.7.3.1. To test the hypothesis $H_0 : \beta_k = \beta_l$ vs. $H_A : \beta_k \neq \beta_l$; we use the asymptotic t-statistic, which takes the following form:

$$t - statistic = \frac{\hat{\beta}_k - \hat{\beta}_l}{\sqrt{S_k^2 + S_l^2 - 2S_{k,l}}} \quad 5.14$$

where $\hat{\beta}_k, \hat{\beta}_l$ are the estimates for the k^{th} and l^{th} parameters,
 S_k, S_l are the standard errors of the estimates for the k^{th} and l^{th} parameters and
 $S_{k,l}$ is the error covariance for the estimates for the k^{th} and l^{th} parameters.

That is the ratio is the differences between the two parameter estimates and the standard deviation of that difference. As before, sufficiently large absolute values of the *t-statistic* lead to the rejection of the null hypothesis that the parameters are equal. Again, rejection of this null hypothesis implies that the two corresponding variables have a significant different impact on the modal utilities and suggests that the variable should be retained in the model and low absolute values of the t-statistic imply that the variables do not have significantly different effects on the utility function or the explanatory power of the model and can be combined in the model.

This test can be readily extended to the test of a hypothesis that the two parameters are related by a predefined ratio; for example, the parameter for time and cost may be related by an *a priori* value of time. In that case, the null hypothesis becomes $H_0 : \beta_{cost} = (VOT) \times \beta_{time}$ and the alternative hypothesis is $H_A : \beta_{cost} \neq (VOT) \times \beta_{time}$. The corresponding t-statistic becomes

$$t - statistic = \frac{\hat{\beta}_{cost} - (VOT)\hat{\beta}_{time}}{\sqrt{S_{cost}^2 + (VOT)^2 S_{time}^2 - 2(VOT)S_{time, cost}}} \quad 5.15$$

The interpretation is the same as above except that the hypothesis refers to the equality of one parameter to the other parameter times an *a priori* fixed value. In this case, if it is concluded that the hypothesis is not rejected, the ratio constraint can be incorporated in the model by creating a new variable $X_{time, cost} = X_{cost} + (VOT)X_{time}$.

5.7.3.3 Tests of Entire Models

The t-statistic is used to test the hypothesis that a single parameter is equal to some pre-selected value or that there is a linear relationship between a pair of parameters. Sometimes, we wish to test multiple hypotheses simultaneously. This is done by formulating a test statistic which can be used to compare two models provided that one is a restricted version of the other; that is, the restricted model can be obtained by imposing restrictions (setting some parameters to zero, setting pairs of parameters equal to one another and so on) on parameters in the unrestricted model. This test statistic can then be used for any case when one or more restrictions are imposed on a model to obtain another model.

If all the restrictions that distinguish between the restricted and unrestricted models are valid, one would expect the difference in log-likelihood values (at convergence) of the restricted and unrestricted models to be small. If some or all the restrictions are invalid, the difference in log-likelihood values of the restricted and unrestricted models will be “sufficiently” large to reject the hypotheses. This underlying logic is the basis for the likelihood ratio test. The test statistic is:

$$-2 \times [LL_R - LL_U] \quad 5.16$$

where LL_R is the log-likelihood of the restricted model, and
 LL_U is the log-likelihood of the unrestricted model.

This test-statistic is chi-squared distributed. Example chi-squared distributions are shown in Figure 5.4. As with the test for individual parameters, the critical value for determining if the statistic is “sufficiently large” to reject the null hypothesis depends on the level of confidence desired by the model developer. It is also influenced by the number of restrictions²⁷ between the models. Table 5-5 shows the chi-squared values for selected confidence levels and for different numbers of restrictions.

Figure 5.5 illustrates the 90% and 95% confidence thresholds on the chi-squared distribution for five degrees of freedom. The critical chi-square values increase with the desired confidence level and the number of restrictions.

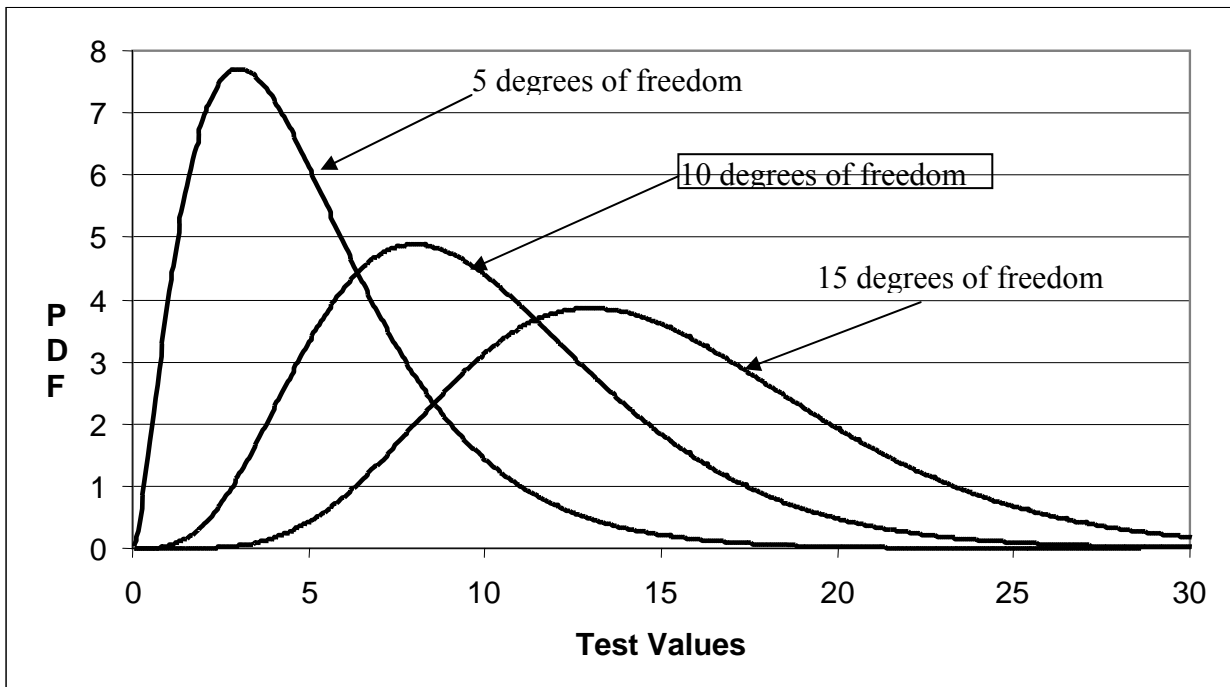


Figure 5.4 Chi-Squared Distributions for 5, 10, and 15 Degrees of Freedom

²⁷ The number of restrictions is the number of constraints imposed on the unrestricted model to obtain the restricted model. If three variables are deleted from the unrestricted mode; that is, their parameters are restricted to zero; the number of restrictions is three.

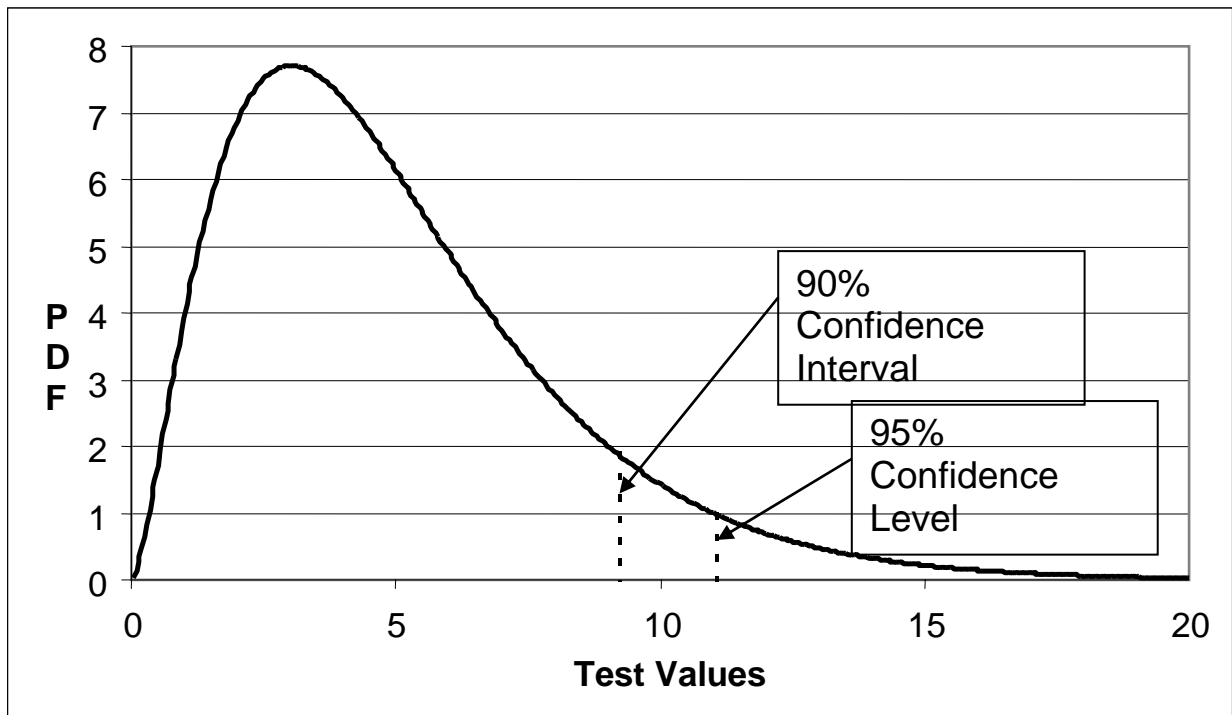


Figure 5.5 Chi-Squared Distribution for 5 Degrees of Freedom Showing 90% and 95% Confidence Thresholds

Table 5-5 Critical Chi-Squared (χ^2) Values for Selected Confidence Levels by Number of Restrictions

Level of Conf.	Number of Restrictions								
	1	2	3	4	5	7	10	12	15
90%	2.71	4.61	6.25	7.78	9.24	12.01	15.99	18.54	22.31
95%	3.84	5.99	7.81	9.49	11.07	14.06	18.31	21.02	25.00
99%	6.63	9.21	11.34	13.28	15.09	18.48	23.21	26.21	30.58
99.5%	7.88	10.60	12.84	14.86	16.75	20.28	25.19	28.30	32.80
99.9%	10.83	13.82	16.27	18.47	20.51	24.32	29.59	32.91	37.70

The likelihood ratio test can be applied to test null hypotheses involving the exclusion of groups of variables from the model. Table 5-6 illustrates the tests of two hypotheses describing restriction on some or all the parameters in the San Francisco Bay Area commuter mode choice model. The first hypothesis is that all the parameters are equal to zero. The formal statement of the null hypothesis in this case, is:

$$\begin{aligned}
 H_{0,a} : \beta_{Travel\ Time} &= \beta_{Travel\ Cost} = 0, \\
 \beta_{SR2} &= \beta_{SR3} = \beta_{TR} = \beta_{WK} = \beta_{BK} = 0, \text{ and} \\
 \beta_{Income-SR2} &= \beta_{Income-SR3+} = \beta_{Income-Transit} \\
 &= \beta_{Income-Bike} = \beta_{Income-Walk} = 0
 \end{aligned}
 \tag{5.17}$$

This test is not very useful because we almost always reject the null hypothesis that all coefficients are zero. A somewhat more useful null hypothesis is that the variables in the initial model specification provide no additional information in addition to the market share information represented by the alternative specific constants. The restrictions for this null hypothesis are:

$$\begin{aligned}
 H_{0,b} : \beta_{Travel\ Time} &= \beta_{Travel\ Cost} = 0, \text{ and} \\
 \beta_{Income-SR2} &= \beta_{Income-SR3+} = \beta_{Income-Transit}, \text{ and} \\
 &= \beta_{Income-Bike} = \beta_{Income-Walk} = 0
 \end{aligned}
 \tag{5.18}$$

The log-likelihood values needed to test each of these hypotheses are reported in Table 5-2. In each case, we include the log-likelihood of the restricted and unrestricted models, the calculated chi-square value and the number of restrictions or degrees of freedom as shown in Table 5-6. The confidence or significance of the rejection of the null hypothesis in each case can be obtained by referring to Table 5-5, more extensive published tables or software (most spreadsheet programs) that calculates the precise level of confidence/significance associated with each test result.

Table 5-6 Likelihood Ratio Test for Hypothesis $H_{0,a}$ and $H_{0,b}$

Variables	Test for Hypothesis $H_{0,a}$	Test for Hypothesis $H_{0,b}$
Log-Likelihood of Unrestricted Model (LL_U)	-3626.186	-3626.186
Log-Likelihood of Restricted Model (LL_R)	-7309.601	-4132.916
Test Statistics [$-2(LL_R-LL_U)$]	7366.830	1013.460
Number of Restrictions	12	7
Critical Chi-Squared Value at 99.9% Confidence	32.91	24.32
Rejection Confidence	99.9%	99.9%
Rejection Significance	0.001	0.001

These two applications of the likelihood ratio test correspond to situations where the null hypothesis leads to a highly restrictive model. These cases are not very interesting since the real value of the likelihood ratio test is in testing null hypotheses which are not so extreme. The log-likelihood ratio test can be applied to test null hypotheses involving the exclusion of selected groups of variables from the model. We consider two such hypotheses. The first is that the time and cost variables have no impact on the mode choice decision, that is,

$$H_{0,C} : \beta_{Travel\ Time} = \beta_{Travel\ Cost} = 0$$

The second is that income has no effect on the travel mode choice; that is

$$\begin{aligned} H_{0,D} : \beta_{Income-SR2} &= \beta_{Income-SR3+} = \beta_{Income-Transit} \\ &= \beta_{Income-Bike} = \beta_{Income-Walk} = 0 \end{aligned}$$

The restricted models that reflect each of these hypotheses and the corresponding unrestricted model are reported in Table 5-7 along with their log-likelihood values.

Table 5-7 Estimation Results for Base Models and its Restricted Versions

Variables	Base Model	Base Model without Time and Cost Variables	Base Model without Income Variables
Travel Cost (1990 cents)	-0.0049 (-20.6)		-0.0048 (-20.6)
Total Travel Time (minutes)	-0.0513 (-16.6)		-0.0514 (-16.6)
Income (1,000's of 1990 dollars)			
Drive Alone (Base)	0.0	0.0	
Shared Ride 2	-0.0022 (-1.4)	-0.0004 (-0.2)	
Shared Ride 3+	0.0004 (0.1)	0.0030 (1.3)	
Transit	-0.0053 (-2.9)	-0.0022 (-1.4)	
Bike	-0.0128 (-2.4)	-0.0122 (-2.3)	
Walk	-0.0097 (-3.2)	-0.0089 (-3.0)	
Mode Constants			
Drive Alone (base)	0.0	0.0	0.0
Shared Ride 2	-2.178 (-20.8)	-2.110 (-21.3)	-2.308 (-42.2)
Shared Ride 3+	-3.725 (-21.0)	-3.472 (-21.0)	-3.702 (-39.9)
Transit	-0.6709 (-5.1)	-1.820 (-17.8)	-0.974 (-11.0)
Bike	-2.376 (-7.8)	-2.672 (-8.8)	-3.071 (-19.9)
Walk	-0.2068 (-1.1)	-1.598 (-9.8)	-0.704 (-5.4)
Log-likelihood at Zero	-7309.601	-7309.601	-7309.601
Log-likelihood at Constant	-4132.916	-4132.916	-4132.916
Log-likelihood at Convergence	-3626.186	-4123.615	-3637.579
Rho-Squared w.r.t. Zero	0.5039	0.4359	0.5024
Rho-Squared w.r.t. Constants	0.1226	0.0023	0.1198
Adjusted Rho-Squared w.r.t. Zero	0.5023	0.4345	0.5014
Adjusted Rho-Squared w.r.t. Const.	0.1197	0.0010	0.1181

The statistical test of the hypothesis that time and cost have no effect has a chi-square value of

$$\chi^2 = -2(-4123.6 - (-3626.2)) = 994.8 \quad 5.19$$

with two degrees of freedom (two parameters constrained to zero). The critical χ^2 with two degrees of freedom at 99.9% confidence (or 0.001 level of significance) is 13.82. Similarly, the statistical test of the hypothesis that income has no effect on mode choice has a chi-square value of

$$\chi^2 = -2(-3637.6 - (-3626.2)) = 22.8, \quad 5.20$$

with five degrees of freedom (five income parameters are constrained to zero). The critical χ^2 with five degrees of freedom at 99.9% confidence level (or 0.001 level of significance) is 20.51. Thus, both null hypotheses can be rejected at very high levels; that is, neither time and cost nor the income variables should be excluded from the model. The log-likelihood ratio tests for both the above hypotheses are summarized in Table 5-8.

Table 5-8 Likelihood Ratio Test for Hypothesis $H_{0,c}$ and $H_{0,d}$

Variables	Test for Hypothesis $H_{0,c}$	Test for Hypothesis $H_{0,d}$
Log-Likelihood of Unrestricted Model (LL_U)	-3626.186	-3626.186
Log-Likelihood of Restricted Model (LL_R)	-4123.615	-3637.579
Test Statistics [$-2(LL_R - LL_U)$]	994.858	22.786
Number of Restrictions	2	5
Critical Chi-Squared Value at 99.9% Confidence	13.82	20.51
Rejection Confidence	>>99.9%	>>99.9%
Rejection Significance	0.000	0.000

5.7.3.4 Non-nested Hypothesis Tests

The likelihood ratio test can only be applied to compare models which differ due to the application of restrictions to one of the models. Such cases are referred to as nested hypothesis tests. However, there are important cases when the rival models do not have this type of restricted – unrestricted relationship. For example, we might like to compare the base model to an alternative specification in which the variable cost divided by income is used to replace cost. This reflects the expectation that the importance of cost diminishes with increasing income. This analysis can be performed by using the non-nested hypothesis test proposed by Horowitz (1982). The non-nested hypothesis test uses the adjusted likelihood ratio index, $\bar{\rho}^2$, to test the hypothesis that the model with the lower $\bar{\rho}^2$ value is the true model²⁸. In this test, the null hypothesis that the model with the lower value is the true model is rejected at the significance level determined by the following equation²⁹:

$$\text{Significance Level} = \Phi \left[- \left(-2 (\bar{\rho}_H^2 - \bar{\rho}_L^2) \times LL(0) + (K_H - K_L) \right)^{\frac{1}{2}} \right] \quad 5.21$$

where $\bar{\rho}_L^2$ is the adjusted likelihood ratio index for the model with the lower value,

$\bar{\rho}_H^2$ is the adjusted likelihood ratio index for the model with the higher value,

K_H, K_L are the numbers of parameters in models H and L, respectively,

and

Φ is the standard normal cumulative distribution function.

We illustrate the non-nested hypothesis test by applying it to compare the base model with alternative specifications that replace the cost variable with cost divided by income or cost

²⁸ The alternative test, that the model with the higher $\bar{\rho}^2$ value is the true model, cannot be undertaken as an inferior model can never be used to reject a superior model.

²⁹ Modified from Ben-Akiva and Lerman, Chapter 7, 1985

divided by $\ln(\text{income})$. The estimation results for all three models are presented in Table 5-9. Since the model using cost not adjusted for income has the best goodness of fit (highest $\bar{\rho}^2$), the null hypotheses for these tests is that the model with cost by income variable or the model with cost by $\ln(\text{income})$ is the true model. higher $\bar{\rho}^2$. Since all the models have the same number of parameters, the term $(K_H - K_L)$ drops out, and the equation for the test of the cost by income model being true is:

$$\begin{aligned} \Phi \left[\left(-2(\bar{\rho}_H^2 - \bar{\rho}_L^2) \times LL(0) \right)^{\frac{1}{2}} \right] &= \Phi \left[-(-2(0.5023 - 0.4897)(-7309.6))^{\frac{1}{2}} \right] & 5.22 \\ &= \Phi [-13.58] \ll 0.001 \end{aligned}$$

The corresponding test for the cost by $\ln(\text{income})$ being true is:

$$\begin{aligned} \Phi \left[\left(-2(\bar{\rho}_H^2 - \bar{\rho}_L^2) \times LL(0) \right)^{\frac{1}{2}} \right] &= \Phi \left[-(-2(0.5023 - 0.5015)(-7309.6))^{\frac{1}{2}} \right] & 5.23 \\ &= \Phi [-3.420] < 0.001 \end{aligned}$$

The above result implies that the null hypotheses that the models with cost by income variable or cost by $\ln(\text{income})$ are true are rejected at a significance level greater than 0.001. However, the significance of rejection is much lower for the cost by $\ln(\text{income})$ model and many analysts would adopt that specification on the grounds that it is conceptually more appropriate. This specification suggests that the value of money declines with income but the rate of decline diminishes at higher levels of income.

Table 5-9 Models with Cost vs. Cost/Income and Cost/Ln(Income)

Variables	Base Model (Cost Variable)	Model with Cost by Income Variable	Model with Cost by Ln(Income) Variable
Travel Cost (1990 cents)	-0.0049 (-20.6)		
Travel Cost by Income, Ln(Income) (1990 cents, 1,000 1990 dollars)		By Income -0.1692 (-17.4)	By Ln(Income) -0.0191 (-20.8)
Total Travel Time (minutes)	-0.0513 (-16.6)	-0.0512 (-16.7)	-0.0512 (-16.5)
Income (1,000's of 1990 dollars)			
Drive Alone (Base)	0.0	0.0	0.0
Shared Ride 2	-0.0022 (-1.4)	0.0030 (2.0)	-0.0004 (-0.3)
Shared Ride 3+	0.0004 (0.1)	0.0096 (4.1)	0.0035 (1.4)
Transit	-0.0053 (-2.9)	-0.0009 (-0.5)	-0.0039 (-2.2)
Bike	-0.0128 (-2.4)	-0.0042 (-0.8)	-0.0098 (-1.9)
Walk	-0.0097 (-3.2)	-0.0020 (-0.7)	-0.0072 (-2.4)
Mode Constants			
Drive Alone (base)	0.0	0.0	0.0
Shared Ride 2	-2.1780 (-20.8)	-2.3770 (-22.5)	-2.2817 (-21.7)
Shared Ride 3+	-3.7251 (-21.0)	-4.0797 (-23.3)	-3.9059 (-22.1)
Transit	-0.6708 (-5.1)	-0.7579 (-5.7)	-0.7312 (-5.5)
Bike	-2.3763 (-7.8)	-2.7714 (-9.4)	-2.5561 (-8.5)
Walk	-0.2068 (-1.1)	-0.5982 (-3.1)	-0.3686 (-1.9)
Log-likelihood at Zero	-7309.601	-7309.601	-7309.601
Log-likelihood at Constant	-4132.916	-4132.916	-4132.916
Log-likelihood at Convergence	-3626.19	-3718.39	-3629.00
Rho-Squared w.r.t. Zero	0.5039	0.4913	0.5035
Rho-Squared w.r.t. Constants	0.1226	0.1003	0.1219
Adjusted Rho-Squared w.r.t. Zero	0.5023	0.4897	0.5015
Adjusted Rho-Squared w.r.t. Const.	0.1197	0.0974	0.1197

5.8 Value of Time

5.8.1 Value of Time for Linear Utility Function

The value of time, as described in Section 5.7.1.3, is calculated as the ratio of the parameter for time over the parameter for cost. This ratio assumes that the utility function is linear in both time and cost and that neither value is interacted with any other variables. That is, when the time and cost portion of the utility function is

$$V_i = \dots + \beta_{TVT} TVT_i + \beta_{Cost} Cost_i + \dots \quad 5.24$$

the value of time is given by

$$VofT = \beta_{TVT} / \beta_{Cost} \quad 5.25$$

The units of time value are obtained from the units of the variables used to measure time and cost. In the Base Model in Table 5.9, the units are minutes and cents. Thus the value of time in cents per minute implied by this model is $-0.0513 / -0.0049 = 10.5$ cents per minute. This can be modified to \$ per hour by multiply by $0.6 = (1/100 \text{ \$ per cent}) / (1/60 \text{ hour per minute})$.

However, in general, the value of time is equal to the ratio between the derivative of utility with respect to time and the derivative of utility with respect to cost. That is

$$VofT = \frac{\frac{\partial V_i}{\partial Time_i}}{\frac{\partial V_i}{\partial Cost_i}} \quad 5.26$$

In the case described in Equation 5.24, this produces the ratio in Equation 5.25. However, this more general formulation allows us to infer the value of time for a variety of special cases in addition to the linear utility case described above.

5.8.2 Value of Time when Cost is Interacted with another Variable

This approach can be applied to any case including when either time or cost is interacted with another variable, usually a variable describing the decision maker or the decision context. For example, if cost is divided by income³⁰ as in the second model reported in Table 5-9 Models with Cost vs. Cost/Income and Cost/Ln(Income)Table 5-9, on the basis that a unit of cost is proportionally less important with increasing income, the utility expression becomes

$$V_{it} = \dots + \beta_{TVT}TVT_{it} + \beta_{CostInc} \frac{Cost_{it}}{Income_t} + \dots \tag{5.27}$$

and the value of time becomes

$$VofT = \frac{\frac{\partial V_i}{\partial Time_i}}{\frac{\partial V_i}{\partial Cost_i}} = \beta_{TVT} / \beta_{CostInc} / Income_t \tag{5.28}$$

which can be converted to

$$VofT = \frac{\beta_{IVT}}{\beta_{COST} \text{Income} (\$1000/\text{year})} \text{cents}/\text{minute} \tag{5.29}$$

or

$$VofT = \frac{0.6 \times \beta_{IVT}}{\beta_{COST} \text{Income} (\$1000/\text{year})} \$/\text{hour}$$

³⁰ The value of income commonly used is the total household income since that value is most commonly collected in surveys. This raises a variety of potential problems in interpretation as the hourly wage rate based on household income is only the wage rate of the worker in a single worker household and does not readily apply to any worker in multi-worker household to non-workers. This issue is not explicitly addressed in this manual but raises more general issues of model interpretation and use of the results.

Similarly, the value of time for the third specification in Table 5-9 Models with Cost vs. Cost/Income and Cost/Ln(Income) is

$$V_{ofT} = V_{ofT} = \frac{0.6 \times \frac{\beta_{IVT}}{\beta_{COST}}}{\ln \left[\text{Income} \left(\frac{\$1000}{\text{year}} \right) \right]} \$/\text{hour}$$

The value of time implied by each of these formulations is illustrated in Table 5-10 and by showing the values of time for different income levels. In each case, the values of travel time appear to be quite low. This will be addressed in model specification refinement (Section 6.2.6).

Table 5-10 Value of Time vs. Income

Annual Income	Hourly Wage	Value of Time		
		Linear	Cost/Income	Cost/Ln(Inc.)
\$25,000	\$12.50	\$6.26	\$4.53	\$5.18
\$50,000	\$25.00	\$6.26	\$9.07	\$6.29
\$75,000	\$37.50	\$6.26	\$13.60	\$6.95
\$100,000	\$50.00	\$6.26	\$18.13	\$7.41
\$125,000	\$62.50	\$6.26	\$21.94	\$7.72

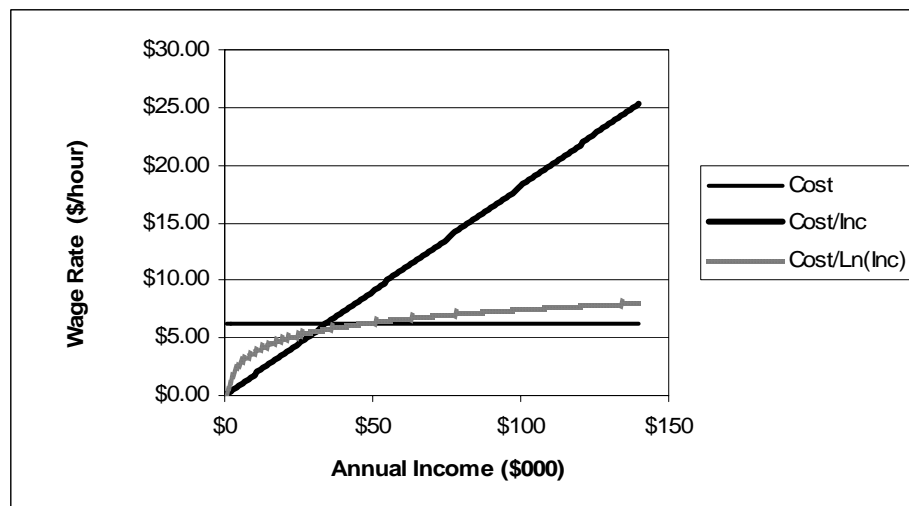


Figure 5.6 Value of Time vs. Income

Another approach that can be used in this case is to relate the value of travel directly to the wage rate by assuming that the working year consists of 2000 hours or 120,000 minutes and recognizing that \$1000 dollars is equivalent to 100,000 cents. This gives us

$$\begin{aligned}
 \text{Units}_{\text{VofT}} &= \frac{\text{cents/minute}}{\$1000/\text{year}} \\
 &= \frac{\text{cents/minute}}{100,000 \text{ cents}/120,000 \text{ minutes}} \\
 &= 1.2
 \end{aligned}
 \tag{5.30}$$

That is, there are no units but a simple factor of 1.2. This is interpreted as the value by which the ratio of the parameters should be multiplied to get the value of travel time as a fraction of the wage rate. In the cost by income model in Table 5-9 Models with Cost vs. Cost/Income and Cost/Ln(Income), this becomes

$$\text{VofT} = \frac{\beta_{TVT}}{\beta_{CostInc}} \times 1.2 = \frac{-0.0512}{-0.1692} \times 1.2 = 0.363 \text{ Wage Rate}
 \tag{5.31}$$

which, as before, is quite low. However, as shown in Figure 5.6 and discussed in Section 5.7.3.4, this specification and the related specification for cost by ln(income) have the advantage that the value of time is differentiated across households with different income.

5.8.3 Value of Time for Time or Cost Transformation

If time or cost is transformed, it becomes necessary to explicitly take the derivative of utility with respect to both time and cost. For example, if time enters the utility function using the natural log transformation, to suggest that the utility effect of increasing time decreases with time, the utility function becomes

$$V_{it} = \dots + \beta_{\ln(TVT)} \ln(TVT_{it}) + \beta_{Cost} Cost_{it} + \dots
 \tag{5.32}$$

and the value of time becomes

$$V_{ofT} = \frac{\frac{\partial V_{it}}{\partial Time_{it}}}{\frac{\partial V_{it}}{\partial Cost_{it}}} = \frac{\beta_{\ln(TVT)} / TVT_{it}}{\beta_{Cost}} = \frac{\beta_{\ln(TVT)}}{\beta_{Cost}} \times \frac{1}{TVT_{it}} \quad 5.33$$

Similarly, if cost is entered as the natural log of cost, the value of time becomes

$$V_{ofT} = \frac{\frac{\partial V_{it}}{\partial Time_{it}}}{\frac{\partial V_{it}}{\partial Cost_{it}}} = \frac{\beta_{TVT}}{\beta_{\ln(Cost)} / Cost_{it}} = \frac{\beta_{TVT}}{\beta_{\ln(Cost)}} \times Cost_{it} \quad 5.34$$

which can be reported in a table for selected values of time or plotted in a graph of Value of Time as a function of TVT or Cost, as appropriate (see below). Models using each of these formulations are estimated and reported, along with the Base Model in Table 5-11. The goodness of fit is substantially improved by using ln(time), which is generally expected, but is worse when using ln(cost), for which there is no conceptual basis.

Table 5-11 Base Model and Log Transformations

Variables	Base Model (Cost Variable)	Model with Log of Travel Time	Model with Log of Travel Cost
Travel Cost (1990 cents)	-0.0049 (-20.6)	-0.0034 (-15.1)	
Log of Travel Cost			-1.05 (-20.4)
Total Travel Time (minutes)	-0.0513 (-16.6)		-0.0598 (-19.0)
Log of Total Travel Time		-2.4 (-19.3)	
Income (1,000's of 1990 dollars)			
Drive Alone (Base)	0.0	0.0	0.0
Shared Ride 2	-0.0022 (-1.4)	-0.0026 (-1.6)	-0.0009 (-0.6)
Shared Ride 3+	0.0004 (0.1)	0.0001 (0.0)	0.0024 (1.0)
Transit	-0.0053 (-2.9)	-0.0066 (-3.5)	-0.0052 (-2.9)
Bike	-0.0128 (-2.4)	-0.0125 (-2.4)	-0.0132 (-2.4)
Walk	-0.0097 (-3.2)	-0.0090 (-2.9)	-0.0089 (-2.8)
Mode Constants			
Drive Alone (base)	0.0	0.0	0.0
Shared Ride 2	-2.1780 (-20.8)	-1.75 (-16.0)	-2.48 (-23.1)
Shared Ride 3+	-3.7250 (-21.0)	-3.14 (-17.4)	-4.31 (-24.2)
Transit	-0.6709 (-5.1)	0.0788 (0.5)	0.0062 (0.0)
Bike	-2.3760 (-7.8)	-1.64 (-5.3)	-5.95 (-16.2)
Walk	-0.2068 (-1.1)	1.19 (5.1)	-3.30 (-12.6)
Log-likelihood at Zero	-7309.601	-7309.601	-7309.601
Log-likelihood at Constant	-4132.916	-4132.916	-4132.916
Log-likelihood at Convergence	-3626.186	-3590.502	-3674.385
Rho-Squared w.r.t. Zero	0.5039	0.5088	0.4973
Rho-Squared w.r.t. Constants	0.1226	0.1312	0.1109
Adjusted Rho-Squared w.r.t. Zero	0.5023	0.5072	0.4957
Adjusted Rho-Squared w.r.t. Const.	0.1197	0.1294	0.1091

The implications for value of time of these different formulations are shown in Table 5-12 and Table 5-13 Value of Time for Log of Cost Model and also in Figure 5.7 and Figure 5.8.

Table 5-12 Value of Time for Log of Time Model

Trip Time (min)	Value of Time (¢/min)	Value of Time (\$/hr)
5	141.2	\$84.71
15	47.1	\$28.24
30	23.5	\$14.12
60	11.8	\$7.06
90	7.8	\$4.71
120	5.9	\$3.53

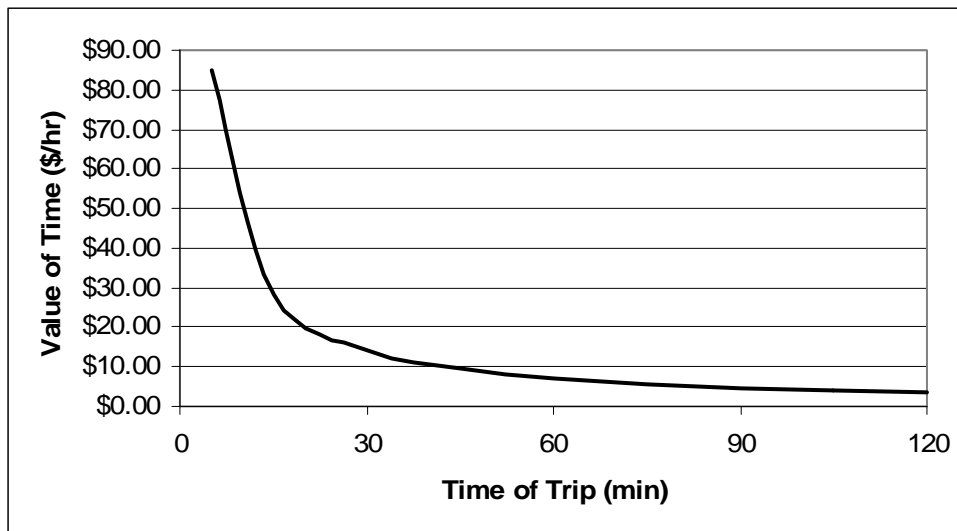


Figure 5.7 Value of Time for Log of Time Model

Table 5-13 Value of Time for Log of Cost Model

Trip Cost	Value of Time (\$/hr)
\$0.25	\$0.85
\$0.50	\$1.71
\$1.00	\$3.42
\$2.00	\$6.83
\$5.00	\$17.09

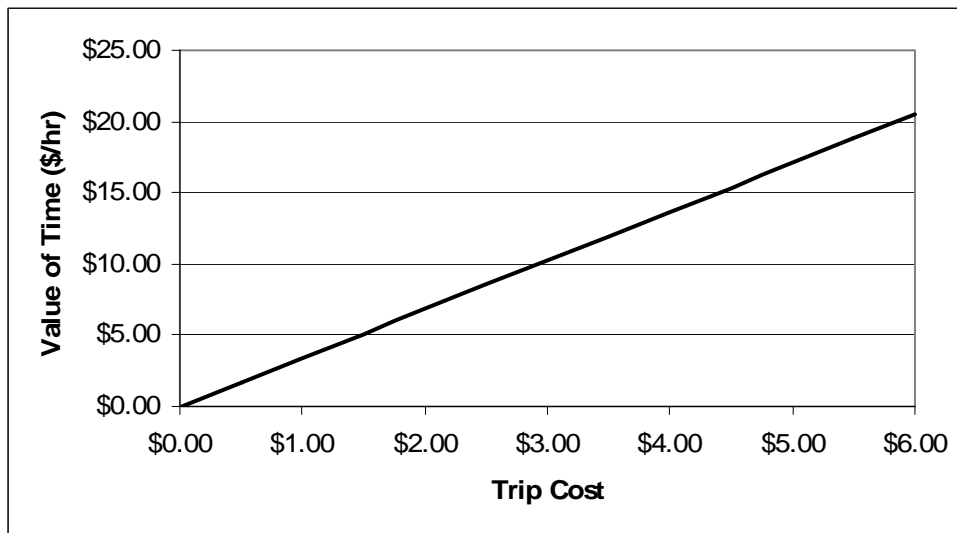


Figure 5.8 Value of Time for Log of Cost Model

CHAPTER 6: Model Specification Refinement: San Francisco Bay Area Work Mode Choice

6.1 Introduction

This chapter describes and demonstrates the refinement of the utility function specification for the multinomial logit (MNL) model for work mode choice in the San Francisco Bay Area. The process combines the use of intuition, statistical analysis and testing, and judgment. The intuition and judgment components of the model refinement process are based on theory, anecdotal evidence, logical analysis, and the accumulated empirical experience of the model developer. This empirical experience can be and often is enhanced through the advice of others or through review of reports and published papers documenting previous modeling studies for similar choice problems and contexts.

We explore a variety of different specifications of the utility functions to demonstrate some of the most common specifications and testing methods. These tests include both formal statistical tests and informal judgments about the signs, magnitudes, or relative magnitudes of parameters based on our knowledge about the underlying behavioral relationships that influence mode choice. The use of judgment and experience is an essential element of successful model development since it is almost impossible to determine the “best” model specification solely on the basis of statistical tests. A model that fits the data well may not necessarily describe the causal relationships and may not produce the most reasonable predictions. Also, it is not uncommon to find several model specifications that, for all practical purposes, fit the data equally well, but which have very different specifications and forecast implications. Therefore, practical model building involves considerable use of subjective judgment and is as much an art as it is a science.

Different modelers have different styles and approaches to the model development process. One of the most common approaches is to start with a minimal specification which includes those variables that are considered essential to any reasonable model. In the case of mode choice, such a specification might include travel time, travel cost and departure frequency

where appropriate for each alternative. Working from this minimal specification, incremental changes are proposed and tested in an effort to improve the model in terms of its behavioral realism and/or its empirical fit to the data while avoiding excessive complexity of the model. Another common approach is to start with a richer specification which represents the model developer's judgment about the set of variables that is likely to be included in the final model specification. For example, such a model might include travel time (separated into in-vehicle and out-of-vehicle time), out of vehicle travel time might be adjusted to take account of the total distance traveled, out of pocket travel cost (possibly adjusted by household income), frequency of departure for carrier modes, household automobile ownership or availability, household income, and size of the travel party.

We adopt the first of these methods in the following section for refinement of the specification of a model of work mode choice as it is the most appropriate approach for those who are new to discrete choice modeling. At each stage in the model development process, we introduce incremental changes to the modal utility functions and re-estimate the model with the objective of finding a more refined model specification that performs better statistically and is consistent with theory and our *a priori* expectations about mode choice behavior. We introduce small changes at each step as the estimation results for each stage provide useful insights which may be helpful in further refining the model. The appropriateness of each specification change is evaluated at each step using both judgmental and statistical tests.

In the rest of this chapter, we describe and demonstrate this process for work mode choice in the San Francisco Bay Area.

6.2 Alternative Specifications

The basic multinomial logit mode choice model for work commute in the San Francisco Bay Area was reported in Table 5-2 in CHAPTER 5. The refinements we consider include:

- Different specifications of the income effects,
- Different specifications of travel time,
- Additional decision maker related variables such as gender and automobiles owned,

- Additional variables that represent the interaction of decision maker related variables with mode related variables (*e.g.*, interaction of income with cost), and
- Additional trip context variables (*e.g.*, dummy variable indicating if the trip origin/destination is in a Central Business District).

6.2.1 Refinement of Specification for Alternative Specific Income Effects

The estimation results for the base model in CHAPTER 5 yielded time and cost parameter estimates that had the expected (negative) sign and were statistically significant. The parameters for the alternative specific income variables were significant and had the expected sign (negative relative to drive alone) except for the shared ride specific income variables (shared ride 2 and shared ride 3+) which were not significant and the sign on the shared ride 3+ income variable was counter-intuitive. All else being equal, we expect the preference for shared ride 2 to be negative relative to drive alone and for shared ride 3+ to be more negative than shared ride 2 because of the increasing inconvenience of coordinating with other travelers as the number of persons in the ride sharing group increases. However, the empirical results provide only limited support for the first expectation and are inconsistent with the second expectation. This suggests that the effect of income on choice is not necessarily different among the automobile modes.

We approach this inconsistency between expectation and empirical results by thinking of other plausible relationships for the effect of income on shared ride choice and developing alternative specifications which represent these relationships. Options for consideration include:

- The effect of income relative to drive alone is the same for the two shared ride modes (shared ride 2 and shared ride 3+) but is different from drive alone and different from the other modes. This relationship is represented by constraining the income coefficients in the two shared alternatives to be equal as follows:

$$H_0 : \beta_{IncomeSR2} = \beta_{IncomeSR3+} \quad 6.1$$

- The effect of income relative to drive alone is the same for both shared ride modes and transit but is different for the other modes. This is represented in the model by constraining the income coefficients in both shared ride modes and the transit mode to be equal as:

$$H_0 : \beta_{Income-SR2} = \beta_{Income-SR3+} = \beta_{Income-Transit} \quad 6.2$$

- The effect of income on all the automobile modes (drive alone, shared ride 2, and shared ride 3+) is the same, but the effect is different for the other modes. We include this constraint by setting the income coefficients in the utilities of the automobile modes to be equal. In this case, we set them to zero since drive alone is the reference mode.

$$H_0 : \beta_{IncomeSR2} = \beta_{IncomeSR3+} = 0 \quad 6.3$$

The estimation results for the base model (from CHAPTER 5) and for these three alternative models are reported in Table 6-1. The parameter estimates for all three models are consistent with expectations. That is, the effect of increasing income is neutral or negative for the shared ride modes relative to drive alone and equal to or more negative for transit, bike and walk than for shared ride. Further, all the parameters are significant except for the shared ride income parameters in Model 1W.

Selection of one of these four models to represent the effect of income should consider the statistical relationships among them and the reasonableness of the resultant models. Since Models 1W, 2W and 3W are constrained versions of the Base Model and Models 2W and 3W are constrained versions of Model 1W, we can use the likelihood ratio test to evaluate the hypotheses implied by each of these models (see section 5.7.3.2). We use this test to determine if the hypothesis that each of these models is the true model is or is not rejected by the less restricted model. The likelihood ratio statistics (equation 5.16), the degrees of freedom or number of restrictions and the level of significance for each test are reported relative to the Base Model and to Model 1W in the first and second rows of Table 6-2, respectively. The Base Model cannot reject any of the subsequent models at a reasonable level of significance. Further, the Base Model has a counter-intuitive relationship between the parameters for shared ride 2 and shared ride 3+. Thus, Model 1W or Model 3W can represent the effect of income on mode choice in this case. We choose Model 1W because it is most consistent with our prior hypotheses about the effect of income on preference between drive alone and shared ride and other modes. However, the differences among these models are small both statistically and

behaviorally so the decision should be subject to a review before adoption of the final specification.³¹

Table 6-1 Alternative Specifications of Income Variable

Variables	Base Model		Model 1W		Model 2W		Model 3W	
Travel Cost (1990 cents)	-0.0049	(-20.6)	-0.0049	(-0.6)	-0.0049	(-20.6)	-0.0049	(-20.6)
Total Travel Time (minutes)	-0.0513	(-16.6)	-0.0513	(-16.6)	-0.0514	(-16.6)	-0.0513	(-16.6)
Income (1000's of 1990 dollars)								
Drive Alone (Base)	0		0		0		0	
Shared Ride 2	-0.0022	(-1.4)	-0.0016	(-1.1)	-0.0029	(-2.3)		0
Shared Ride 3+	0.0004	-0.1	-0.0016	(-1.1)	-0.0029	(-2.3)		0
Transit	-0.0053	(-2.9)	-0.0053	(-2.9)	-0.0029	(-2.3)	-0.0049	(-2.7)
Bike	-0.0128	(-2.4)	-0.0128	(-2.4)	-0.0125	(-2.3)	-0.0125	(-2.3)
Walk	-0.0097	(-3.2)	-0.0098	(-3.2)	-0.0092	(-3.1)	-0.0093	(-3.1)
Mode Constants								
Drive Alone (base)	0		0		0		0	
Shared Ride 2	-2.178	(-20.8)	-2.212	(-22.7)	-2.137	(-24.2)	-2.304	(-42.1)
Shared Ride 3+	-3.725	(-21.0)	-3.612	(-29.3)	-3.532	(-30.6)	-3.704	(-39.8)
Transit	-0.6709	(-5.1)	-0.6698	(-5.1)	-0.7996	(-7.1)	-0.6976	(-5.4)
Bike	-2.376	(-7.8)	-2.377	(-7.8)	-2.39	(-7.9)	-2.398	(-7.9)
Walk	-0.2068	(-1.1)	-0.2075	(-1.1)	-0.2297	(-1.2)	-0.2292	(-1.2)
Log-likelihood at Zero	-7309.601		-7309.601		-7309.601		-7309.601	
Log-likelihood at Constant	-4132.916		-4132.916		-4132.916		-4132.916	
Log-likelihood at Convergence	-3626.186		-3626.590		-3628.286		-3627.234	
Rho-Squared w.r.t. Zero	0.5039		0.5039		0.5036		0.5038	
Rho-Squared w.r.t. Constants	0.1535		0.1534		0.153		0.1532	

Table 6-2 Likelihood Ratio Tests between Models in Table 6-1

(Likelihood Ratio Statistic, Degrees of Freedom, and Rejection Significance Level)

Reference Model→	Model 1W	Model 2W	Model 3W
Base Model	0.8, 1, 0.371	4.2, 2, 0.122	2.0, 2, 0.368
Model 1W	NA	3.4, 1, 0.065	1.2, 1, 0.273

31 As other variables are added to the model, the differences between these two specifications may change providing a stronger statistical basis for selecting Model 1 or 3.

6.2.2 Different Specifications of Travel Time

The specification for travel time in the above models implies that the utility value of time is equal for all the alternatives and between in-vehicle and out-of-vehicle time. However, we expect travelers in non-motorized modes to be more sensitive to travel time than travelers in motorized modes (since walking or biking is physically more demanding than traveling in a car) and we expect that travelers are more sensitive to out-of-vehicle travel time (OVT) than to in-vehicle travel time (IVT).

The estimation results for two specifications of travel time that relax these constraints are reported with those for Model 1W in Table 6-3. Model 5W relaxes the time constraints in Model 1W by specifying distinct time variables for the motorized and non-motorized modes based on our expectation that travelers are likely to be more sensitive to travel time by non-motorized modes. Model 6W relaxes the constraint further by disaggregating the travel time for motorized modes into distinct components for IVT and OVT. This specification allows the two components of travel time for motorized travel to have different effects on utility with the expectation that travelers are more sensitive to out-of-vehicle time than in-vehicle time.

The estimation results for Model 5W rejects the hypothesis of equal value of travel time across modes implied in Model 1W and Model 6W rejects the hypothesis of equal value of in and out of travel time for the motorized modes at a very high level of significance (0.001). The estimated parameters associated with travel time in Model 6W have the correct signs and the magnitude of the parameters for OVT for motorized modes and for time for non-motorized modes are larger in magnitude than the parameter for IVT, as expected; however, the parameter for IVT is very small and not statistically significant. Further, the ratio of OVT to IVT for motorized modes, 30 times, is far greater than expected. Nonetheless, since Model 6W rejects the constraints imposed by both Models 1W and 5W at a very high level of significance, we cannot discard this model without further exploration.

Another perspective on the suitability of these models can be obtained by calculating the relative importance of each component of travel time and cost which gives us the implied value of each component of time. The implied value of in-vehicle-time for motorized modes is

computed for each model using the estimated motorized in-vehicle-time and cost parameters and similarly for the other time components:

$$\text{Value of motorized IVTT (\$/hour)} = \frac{\beta_{\text{motorized ivtt (1/min.)}}}{\beta_{\text{cost (1/cents)}}} \times \frac{60 \text{ min./hour}}{100 \text{ cents/\$}} \quad 6.4$$

The implied values of in- and out-of-vehicle times for motorized modes in Models 1W, 5W, and 6W are reported in Table 6-4. The values of motorized in-vehicle time and non-motorized time are somewhat low but not unreasonable compared to the average wage rate of \$21.20 per hour in the region (1990 dollars); however, the value of in-vehicle time is unreasonably low. Nevertheless, the likelihood ratio tests reject both Model 5W and Model 1W at very high levels of significance. This raises doubt about the suitability of those models and suggests the need to consider other specifications to evaluate the influence of travel time components on the utilities of the different alternatives.

Two approaches are commonly taken to identify a specification which is not statistically rejected by other models and has good behavioral relationships among variables. The first is to examine a range of different specifications in an attempt to find one which is both behaviorally sound and statistically supported. The other is to constrain the relationships between or among parameter values to ratios which we are considered reasonable. The formulation of these constraints is based on the judgment and prior empirical experience of the analyst. Therefore, the use of such constraints imposes a responsibility on the analyst to provide a sound basis for his/her decision. The advice of other more experienced analysts is often enlisted to expand and/or support these judgments.

Table 6-3 Estimation Results for Alternative Specifications of Travel Time

Variables	Model 1W	Model 5W	Model 6W
Travel Cost (1990 cents)	-0.0049 (-20.6)	-0.005 (-20.7)	-0.0048 (-20.2)
Travel Time (minutes)			
All Modes	-0.0513 (-16.6)		
Motorized Modes Only		-0.0431 (-12.3)	
Non-Motorized Modes Only		-0.0687 (-12.9)	-0.0632 (-11.8)
In-vehicle Travel Time (Motorized Modes)			-0.0025 (-0.4)
Out-of-vehicle Travel Time (Motorized Modes)			-0.0759 (-13.0)
Income (1,000's of 1990 dollars)			
Drive Alone (Base)	0	0	0
Shared Ride 2	-0.0016 (-1.1)	-0.0015 (-1.1)	-0.0016 (-1.1)
Shared Ride and 3+ = Shared Ride 2	-0.0016 (-1.1)	-0.0015 (-1.1)	-0.0016 (-1.1)
Transit	-0.0053 (-2.9)	-0.0055 (-3.0)	-0.0057 (-3.1)
Bike	-0.0128 (-2.4)	-0.0125 (-2.4)	-0.0122 (-2.3)
Walk	-0.0098 (-3.2)	-0.0095 (-3.1)	-0.0093 (-3.0)
Constants Drive Alone (Base)	0	0	0
Shared Ride 2	-2.212 (-22.7)	-2.262 (-23.0)	-2.43 (-24.1)
Shared Ride 3+	-3.612 (-29.3)	-3.677 (-29.6)	-3.883 (-30.5)
Transit	-0.6698 (-5.1)	-0.852 (-6.1)	-0.49 (-3.3)
Bike	-2.377 (-7.8)	-1.844 (-5.7)	-1.719 (-5.3)
Walk	-0.2075 (-1.1)	0.477 (1.9)	0.409 (1.6)
Log-likelihood at Zero	-7309.601	-7309.601	-7309.601
Log-likelihood at Constant	-4132.916	-4132.916	-4132.916
Log-likelihood at Convergence	-3626.590	-3616.894	-3588.039
Rho-Squared w.r.t. Zero	0.5039	0.5052	0.5091
Rho-Squared w.r.t. Constants	0.1225	0.1249	0.1318
Likelihood Ratio Test vs Model 2W ³²	NA	19.4, 1, 0.001	123.6, 2, < 0.001
Likelihood Ratio Test vs Model 5W	NA	NA	57.8, 1, < 0.001

Table 6-4 Implied Value of Time in Models 1W, 5W, and 6W³³

	Model 1W	Model 5W	Model 6W
Value of Non-Motorized Time	\$6.28/hr	\$8.24/hr	\$7.90/hr
Value of Out-of-vehicle Time	\$6.28/hr	\$5.17/hr	\$9.49/hr
Value of In-vehicle Time	\$6.28/hr	\$5.17/hr	\$0.31/hr

32 Model in column used to test null hypothesis that the model identified in the row label is the true model. Values are log-likelihood test statistic, degrees of freedom, and significance of rejection of null hypothesis.

33 Values of time in this and subsequent tables are rounded to the nearest ten cents per hour.

The primary shortcoming of the specification in Model 6W is that the estimated value of IVT is unrealistically small. At least two alternatives can be considered for getting an improved estimate of the value of out-of-vehicle time. One is to use an approach that has been effective in other contexts; that is, to assume that the sensitivity of travelers to OVT diminishes with the trip distance. The idea behind this is that travelers are more willing to tolerate higher out-of-vehicle time for a long trip rather than for a short trip. We still expect that travelers will be more sensitive to OVT than IVT for any travel distance. A formulation which ensures this result is to include total travel time (the sum of in-vehicle and out-of-vehicle time) and out-of-vehicle time divided by distance in place of in- and out-of-vehicle travel time. This specification, as shown below, is consistent with our expectations provided that β_1 and β_2 are negative:

$$\begin{aligned}
 V_m &= \gamma_{0,m} + \beta_1 \times TTT_m + \beta_2 \times \left(\frac{OVT_m}{\text{Dist}} \right) + \dots \\
 &= \gamma_{0,m} + \beta_1 \times (IVT_m + OVT_m) + \frac{\beta_2}{\text{Dist}} \times OVT_m + \dots \\
 &= \gamma_{0,m} + \beta_1 \times IVT_m + \left(\beta_1 + \frac{\beta_2}{\text{Dist}} \right) \times OVT_m + \dots
 \end{aligned} \tag{6.5}$$

An alternative approach is to impose a constraint on the relative importance of OVT and IVT. This is achieved by replacing the travel time variables in the modal utility equations with a weighted travel time (*WTT*) variable defined as in-vehicle time plus the appropriate travel time importance ratio (*TIR*) times out-of-vehicle time ($IVT + TIR \times OVT$). The mechanics of how this constraint works is illustrated as follows:

$$\begin{aligned}
 V_m &= \gamma_{0,m} + \beta_1 \times IVT + (\beta_1 \times TIR) \times OVT + \dots \\
 &= \gamma_{0,m} + \beta_1 \times (IVT + TIR \times OVT) + \dots \\
 &= \gamma_{0,m} + \beta_1 \times WTT + \dots
 \end{aligned} \tag{6.6}$$

so that the parameter for out-of-vehicle time is equal to the parameter for in-vehicle time multiplied by the selected travel time ratio (*TTR*). In Models 8W and 9W, we use travel importance ratios of 2.5 and 4.0, respectively. The estimation results for these models compared to Model 6W are reported in

Table 6-5. The parameter estimates obtained for the travel time, cost, and income variables in all four models have the correct signs and are statistically significant. Model 7W has substantially better goodness-of-fit than Models 6W, 8W and 9W. Since none of the other models are constrained versions of Model 7W, we use the non-nested hypothesis test (see Section 5.7.3.2, Equation 5.21) to compare it with Models 6W, 8W, and 9W.

We illustrate the non-nested hypothesis test by applying it to the hypothesis that Model 6W is the true model given that Model 7W has a higher $\bar{\rho}^2$. Since both models have the same number of parameters, the term $(K_7 - K_6)$ drops out, and the equation becomes

$$\begin{aligned} (\text{Level of Rejection}) &= \Phi[-(-2(\bar{\rho}_7^2 - \bar{\rho}_6^2) \ell(0))^{1/2}] \\ &= \Phi[-(-2(0.5129 - 0.5074)(-7309.6))^{1/2}] && 6.7 \\ &= \Phi(-8.97) \ll 0.001 \end{aligned}$$

That is, the null hypothesis that Model 6W is the true model is rejected with significance much greater than 0.001. Models 8W and 9W are also rejected as the true model at an even higher level of significance.

Table 6-5 Estimation Results for Additional Travel Time Specification Testing

Variables	Model 6W	Model 7W	Model 8W	Model 9W
Travel Cost (1990 cents)	-0.0048 (-20.2)	-0.0041 (-17.2)	-0.0049 (-20.4)	-0.0048 (-20.3)
Travel Time (minutes)				
Motorized Modes		-0.0415 (-11.8)		
Non-Motorized Modes	-0.0632 (-11.8)	-0.0475 (-8.6)	-0.0663 (-12.6)	-0.0652 (-12.4)
IVT (Motorized Modes)	-0.0025 (-0.4)		-0.0254 (-13.4)	-0.0173 (-13.7)
OVT (Motorized Modes)	-0.0759 (-13.0)		-0.0635 (-13.4)	-0.0692 (-13.6)
OVT by Distance (mi.) (Motorized Modes)		-0.1812 (-10.1)		
Income (1,000's of 1990 dollars)				
Drive Alone (Base)	0	0	0	0
Shared Ride 2	-0.0016 (-1.1)	-0.0014 (-1.0)	-0.0016 (-1.1)	-0.0016 (-1.1)
Shared Ride 3+ = Shared Ride 2	-0.0016 (-1.1)	-0.0014 (-1.0)	-0.0016 (-1.1)	-0.0016 (-1.1)
Transit	-0.0057 (-3.1)	-0.0072 (-3.8)	-0.0055 (-3.0)	-0.0056 (-3.0)
Bike	-0.0122 (-2.3)	-0.0118 (-2.3)	-0.0123 (-2.4)	-0.0123 (-2.3)
Walk	-0.0093 (-3.0)	-0.0082 (-2.6)	-0.0095 (-3.1)	-0.0094 (-3.1)
Constants Drive Alone (Base)	0	0	0	0
Shared Ride 2	-2.43 (-24.1)	-2.188 (-22.4)	-2.33 (-24.0)	-2.364 (-24.4)
Shared Ride 3+	-3.883 (-30.5)	-3.518 (-28.6)	-3.756 (-30.6)	-3.799 (-31.1)
Transit	-0.49 (-3.3)	-0.042 (-0.3)	-0.582 (-4.0)	-0.527 (-3.6)
Bike	-1.719 (-5.3)	-2.687 (-8.1)	-1.802 (-5.6)	-1.775 (-5.5)
Walk	0.409 (1.6)	-1.023 (-3.5)	0.442 (1.7)	0.429 (1.7)
Log-likelihood at Zero	-7309.601	-7309.601	-7309.601	-7309.601
Log-likelihood at Constant	-4132.916	-4132.916	-4132.916	-4132.916
Log-likelihood at Convergence	-3588.039	-3547.344	-3595.317	-3590.929
Rho-Squared w.r.t. Zero	0.5091	0.5147	0.5081	0.5087
Rho-Squared w.r.t. Constants	0.1318	0.1417	0.1301	0.1311
Adjusted Rho-Squared w.r.t. Zero	0.5074	0.5129	0.5064	0.5070
Adjusted Rho-Squared w.r.t. Constants	0.1304	0.1402	0.1286	0.1297
Likelihood Ratio Test to reject Model 8W	14.4, 1, < 0.001	NA	NA	NA
Likelihood Ratio Test to reject Model 9W	5.8, 1, 0.016	NA	NA	NA
Non-nested Hyp. Test to reject Model 6W	NA	< 0.001	NA	NA

Before adopting Model 7W, it is a good idea to evaluate and interpret the relative importance of in-vehicle and out-of-vehicle time and between each component of time and cost. Despite the difference in the specification, this analysis is undertaken the same way as earlier; that is, the parameters for time is divided by the parameter for cost to obtain the values of time. The values of IVT and OVT in cents-per-minute (and dollars-per-hour) are shown in Table 6-6 as a function of distance. The time values are obtained as described earlier by dividing each of the time parameters (in utils-per-minute) by the cost parameter in utils per cent. For example, the values for Model 7W are:

$$\begin{aligned} \text{Value of IVTT} &= \frac{\beta_{\text{mot tvtt}}}{\beta_{\text{cost}}} = \frac{-0.0415}{-0.0041} \\ &= 10.1 \text{ cents/min} = \$6.07/\text{hr}. \\ \text{Value of OVT (5 Mile Trip)} &= \frac{\left(\beta_{\text{mot tvtt}} + \frac{\beta_{\text{OVT/Dist}}}{\text{Dist}} \right)}{\beta_{\text{cost}}} \\ &= \frac{-0.0415 + \frac{-0.1812}{5}}{-0.0041} \\ &= 19.0 \text{ cents/min} = \$11.38/\text{hr}. \end{aligned}$$

These values of time are fixed for IVT but vary with distance for OVT³⁴ as reported in Table 6-6 for Model 7W. The corresponding values of time for Models 6W, 8W and 9W are shown in Table 6-7.

³⁴ This formulation is similar to that of cost divided by income described Section 5.8.2.

Table 6-6 Model 7W Implied Values of Time as a Function of Trip Distance

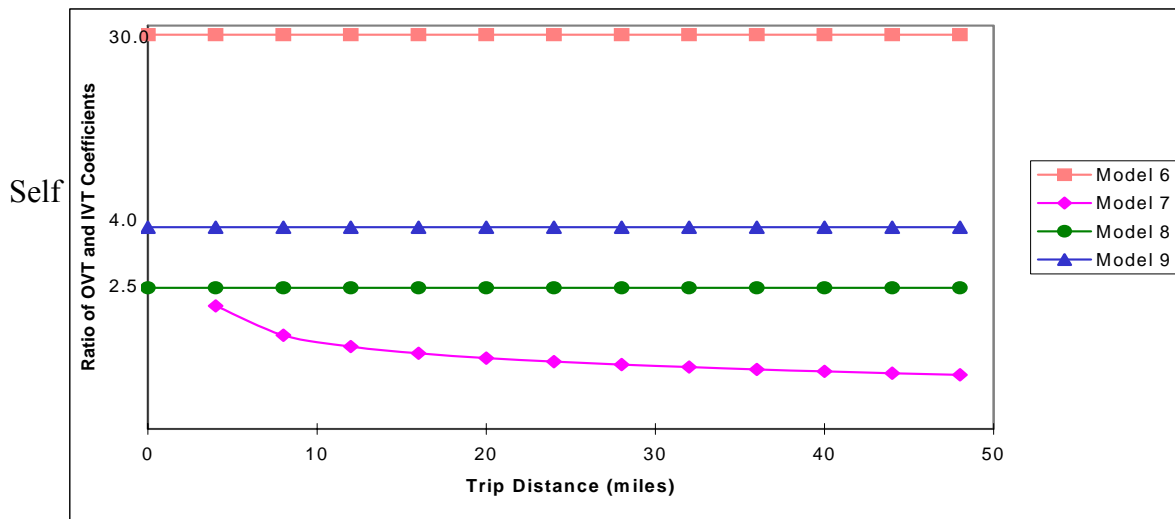
	Trip Distance		
	5 Miles	10 Miles	20 Miles
Value of Motorized Out-of-vehicle Time	19.0 cents/min (\$11.38/hr)	14.5 cents/min (\$8.72/hr)	12.3 cents/min (\$7.40/hr)
Value of Motorized In-vehicle Time	10.1 cents/min (\$6.07/hr)	10.1 cents/min (\$6.07/hr)	10.1 cents/min (\$6.07/hr)
Value of Non-Motorized Time	11.6 cents/min (\$6.95/hr)	11.6 cents/min (\$6.95/hr)	11.6 cents/min (\$6.95/hr)

Table 6-7 Implied Values of Time in Models 6W, 8W, 9W

	Model 6W	Model 8W	Model 9W
Value of Out-of-vehicle Time	\$9.50/hr	\$7.80/hr	\$8.70/hr
Value of In-vehicle Time	\$0.30/hr	\$3.10/hr	\$2.20/hr

The prevailing wage rate in the San Francisco Bay Area is \$21.20 per hour³⁵. In comparison, the values of in-vehicle time implied by Models 6W, 8W, and 9W are very low and the values of out of vehicle time are somewhat low. Model 7W produces higher, but still low, values of time. Finally, we can examine the ratio of time values of OVT relative to IVT for all four models as shown in Figure 6.1. The ratio for Model 6W is clearly unacceptable. Those for Models 7W, 8W and 9W are more reasonable.

³⁵ Refer to the “San Francisco Bay Area 1990 Travel Model Development Project”, Compilation of Technical Memoranda, Volume VI.



**Figure 6.1 Ratio of Out-of-Vehicle and In-Vehicle Time Coefficients
for Work Models 6, 7, 8, and 9**

The selection of a preferred travel time specification among the four alternative specifications tested is relatively straightforward in this case. Model 7W outperforms the other models in all the evaluations undertaken; it has the best goodness-of-fit, the most intuitive relationship between the IVT and OVT variables and the most acceptable values of time³⁶. Consequently, Model 7W is our preferred travel time specification. We can still consider imposing a constraint between the time and cost variables to force the value of time to more reasonable levels. However, we defer this until we explore other specification improvements.

6.2.3 Including Additional Decision Maker Related Variables

There are strong theoretical and empirical reasons to expect that a variety of decision maker related variables such as income, car availability, residential location, number of workers in the household and others, influence workers' choice of travel mode. The models reported to this point include income as the only decision maker related explanatory variable. To the extent that these variables influence the mode choice decision of travelers, their inclusion in the model will increase the explanatory power and predictive accuracy of the model.

³⁶ Based on these results, the model developer might impose constraints between the parameters for the time and cost to obtain higher values of time. The student can demonstrate this by modifying models 7 and 9 so that the value of IVT equals \$10/hour (retaining all other elements of the specifications).

There are two general approaches to including decision maker related variables in models. One is to include such variables as specific to each alternative (except for one base or reference alternative) to indicate the extent to which changes in the variable value will increase or decrease the utility of the mode to that traveler (relative to the reference alternative). The other is to include such variables as interactions with mode related characteristics. For example, dividing cost by income to reflect the decreasing importance of cost with increasing annual income. The inclusion of decision maker related variables as alternative specific variables is demonstrated in this section. Similar treatment of trip context variables is considered in section 6.2.4. Interactions with mode characteristics are demonstrated in section 6.2.5.

We consider number of automobiles in the household, the number of autos divided by the number of household workers and the number of autos divided by the number of persons of driving age in the household. Since these variables are constant across all alternatives, they must be included as distinct variables for each alternative (except for the reference alternative). This is considered a full set of alternative specific variables. The estimation results for these specifications and Model 7W are reported in Table 6-8.

These three new models have much better goodness-of-fit than Model 7W. Each model rejects Model 7W as the true model at a very high level of significance. The parameters for alternative specific automobile availability variables in all the three models have the expected signs, negative relative to drive alone, with the exception of the shared ride 3+ variable in Model 10W which is not significant. Further, the signs and magnitude of the parameters for time, cost, and income are stable across the models. Finally, Models 11W and 12W which include cars-per-worker and cars-per-number-of-adults, respectively, reject Model 10W as the true model.

Overall, Models 11W and 12W are superior to the other two models in terms of behavioral appeal, they provide an indication of automobile availability, and goodness of fit, they statistically reject Models 7W and 10W statistical fit. Model 11W has slightly better goodness-of-fit than Model 12W but the difference is so small that the non-nested hypothesis test is not able to distinguish between the two models. Therefore, selection of a preferred model is primarily a matter of judgment. We select Model 11W but selection of Model 12W would be equally appropriate.

Table 6-8 Estimation Results for Auto Availability Specification Testing

Variables	Model 7W	Model 10W	Model 11W	Model 12W
Travel Cost (1990 cents)	-0.004 (-17.2)	-0.004 (-16.9)	-0.004 (-17.4)	-0.004 (-17.2)
Travel Time (minutes)				
Motorized Modes Only	-0.042 (-11.8)	-0.038 (-10.5)	-0.038 (-10.7)	-0.038 (-10.6)
Non-Motorized Modes Only	-0.048 (-8.6)	-0.048 (-8.4)	-0.047 (-8.4)	-0.047 (-8.3)
OVT by Distance (mi.) Motorized Modes	-0.181 (-10.1)	-0.179 (-9.6)	-0.181 (-9.8)	-0.185 (-9.9)
Income (1,000's of 1990 dollars)				
Drive Alone (Base)	0	0	0	0
Shared Ride 2	-0.001 (-1.0)	-0.002 (-1.4)	-0.002 (-1.2)	-0.001 (-0.9)
Shared Ride 3+ = Shared Ride 2	-0.001 (-1.0)	-0.002 (-1.4)	-0.002 (-1.2)	-0.001 (-0.9)
Transit	-0.007 (-3.8)	-0.001 (-0.7)	-0.006 (-3.0)	-0.005 (-2.3)
Bike	-0.012 (-2.3)	-0.01 (-1.8)	-0.012 (-2.2)	-0.012 (-2.2)
Walk	-0.008 (-2.6)	-0.004 (-1.2)	-0.008 (-2.5)	-0.008 (-2.3)
Auto Ownership		(Autos per household)	(Autos per worker)	(Autos per adult)
Drive Alone (Base)		0	0	0
Shared Ride 2		-0.035 (-0.9)	-0.433 (-5.6)	-0.595 (-5.5)
Shared Ride 3+		0.0723 (1.3)	-0.267 (-2.4)	-0.448 (-2.7)
Transit		-0.555 (-8.0)	-0.99 (-8.6)	-1.409 (-9.1)
Bike		-0.229 (-1.7)	-0.673 (-2.7)	-0.642 (-2.1)
Walk		-0.366 (-3.7)	-0.628 (-3.9)	-0.794 (-3.8)
Constants				
Drive Alone (Base)	0	0	0	0
Shared Ride 2	-2.188 (-22.4)	-2.054 (-16.9)	-1.594 (-12.1)	-1.537 (-10.9)
Shared Ride 3+	-3.518 (-28.6)	-3.643 (-20.3)	-3.14 (-17.0)	-3.023 (-14.8)
Transit	-0.042 (-0.3)	0.574 (3.2)	0.963 (4.8)	1.05 (5.3)
Bike	-2.687 (-8.1)	-2.22 (-5.8)	-1.831 (-4.5)	-1.962 (-4.7)
Walk	-1.023 (-3.5)	-0.441 (-1.3)	-0.238 (-0.7)	-0.217 (-0.6)
Log-likelihood at Zero	-7309.601	-7309.601	-7309.601	-7309.601
Log-likelihood at Constant	-4132.916	-4132.916	-4132.916	-4132.916
Log-likelihood at Convergence	-3547.34	-3501.643	-3489.236	-3490.358
Rho-Squared w.r.t. Zero	0.5147	0.5210	0.5227	0.5225
Rho-Squared w.r.t. Constants	0.1417	0.1528	0.1558	0.1555
Adjusted Rho-Squared w.r.t. Zero	0.5122	0.5185	0.5202	0.5200
Adjusted Rho-Squared w.r.t. Constants	0.1411	0.1511	0.1541	0.1538
Likelihood Ratio Test vs. Model 7W	NA	91.4, 5, < 0.001	116.2, 5, < 0.001	113.8, 5, < 0.001
Adj. LRT vs. Model 10W	NA	NA	< 0.001	< 0.001

6.2.4 Including Trip Context Variables

The models considered to this point include variables that describe the attributes of alternatives, modes, and the characteristics of decision-makers (the work commuters). The mode choice

decision also is influenced by variables that describe the context in which the trip is made. For example, a work trip to the regional central business district (CBD) is more likely to be made by transit than an otherwise similar trip to a suburban work place because the CBD is generally well-served by transit, has more opportunities to make additional stops by walking and is less auto friendly due to congestion and limited and expensive parking. This suggests that the model specification can be enhanced by including variables related to the context of the trip, such as destination zone location.

We consider two distinct variables to describe the trip destination context. One is a dummy variable which indicates whether the destination zone (workplace) is located in the CBD; the other is the employment density of different workplace destinations. The CBD variable implies an abrupt increase in the likelihood of using public transit at the CBD boundary. The density variable implies a continuous increase in the likelihood of using public transit with increasing workplace density. A third option is to include both variables in the model. There is disagreement about whether to include such combinations of variables since they both represent the same underlying phenomenon: increasing transit use with increasing density of development. There is no firm rule about this point; each case must be evaluated on its merits based on statistical tests and reasonableness of the estimation results. As with the addition of characteristics of the traveler, we introduce each variable as a full set of alternative specific variables, each of which represents the effect of a change in that variable on the utility of the alternative relative to the reference alternative (drive alone). Model 13W adds the alternative specific CBD dummy variables to the variables in Model 11W. Model 14W adds the alternative specific employment density variables and Model 15W adds both. Estimation results for these specifications and Model 11W are reported in Table 6-9.

Table 6-9 Estimation Results for Models with Trip Context Variables

Variables	Model 11W		Model 13W		Model 14W		Model 15W	
Travel Cost (1990 cents)	-0.0042	(-17.4)	-0.0033	(-13.0)	-0.0029	(-9.4)	-0.0024	(-7.7)
Travel Time (minutes)								
Motorized Modes Only	-0.0384	(-10.7)	-0.0286	(-10.6)	-0.0299	(-8.0)	-0.0231	(-5.9)
Non-Motorized Modes Only	-0.0470	(-8.4)	-0.0464	(-8.1)	-0.0459	(-8.1)	-0.0467	(-8.1)
OVT by Distance (mi.) Motorized Modes	-0.1814	(-9.8)	-0.1501	(-7.6)	-0.1575	(-8.3)	-0.1324	(-7.9)

Variables	Model 11W		Model 13W		Model 14W		Model 15W	
Income (1,000's of 1990 dollars)								
Drive Alone (Base)	0		0		0		0	
Shared Ride 2 and Shared Ride 3+	-0.002	(-1.2)	-0.002	(-1.2)	-0.002	(-1.3)	-0.002	(-1.3)
Transit	-0.006	(-3.0)	-0.006	(-3.1)	-0.007	(-3.5)	-0.007	(-3.4)
Bike	-0.012	(-2.2)	-0.011	(-2.1)	-0.011	(-2.1)	-0.011	(-2.1)
Walk	-0.008	(-2.5)	-0.008	(-2.5)	-0.008	(-2.4)	-0.008	(-2.5)
Autos per Worker								
Drive Alone (Base)	0		0		0		0	
Shared Ride 2	-0.433	(-5.6)	-0.415	(-5.4)	-0.407	(-5.3)	-0.401	(-5.2)
Shared Ride 3+	-0.267	(-2.4)	-0.212	(-1.9)	-0.237	(-2.1)	-0.183	(-1.7)
Transit	-0.99	(-8.6)	-0.911	(-7.9)	-0.995	(-8.4)	-0.93	(-7.9)
Bike	-0.673	(-2.7)	-0.698	(-2.7)	-0.714	(-2.8)	-0.715	(-2.8)
Walk	-0.628	(-3.9)	-0.719	(-4.3)	-0.681	(-4.1)	-0.727	(-4.3)
CBD Dummy (1 = in CBD; 0 = not in CBD)								
Drive Alone (Base)			0				0	
Shared Ride 2			0.256				0.204	
Shared Ride 3+			1.057				1.018	
Transit			1.356				1.204	
Bike			0.376				0.462	
Walk			0.175				0.109	
Emp. Density - Work Zone (Emp/sq. mi.)								
Drive Alone (Base)					0		0	
Shared Ride 2					0.0011		(3.1) 0.001	
Shared Ride 3+					0.0022		(5.1) 0.0013	
Transit					0.0027		(7.0) 0.0021	
Bike					0.0011		(1.0) 0.0008	
Walk					0.0015		(2.2) 0.0018	
Constants								
Drive Alone (Base)	0		0		0		0	
Shared Ride 2	-1.594	(-12.1)	-1.634	-12	-1.605	(-12.1)	-1.64	(-12.0)
Shared Ride 3+	-3.14	(-17.0)	-3.537	(-17.3)	-3.212	(-17.0)	-3.55	(-17.4)
Transit	0.963	-4.8	-0.202	(-0.8)	0.419	(2.0)	-0.471	(-1.9)
Bike	-1.831	(-4.5)	-1.651	(-3.9)	-1.597	(-3.8)	-1.515	(-3.5)
Walk	-0.238	(-0.7)	0.083	-0.2	-0.041	(-0.1)	0.21	(0.6)
Log-likelihood at Zero	-7309.601		-7309.601		-7309.601		-7309.601	
Log-likelihood at Constant	-4132.916		-4132.916		-4132.916		-4132.916	
Log-likelihood at Convergence	-3489.236		-3440.644		-3460.678		-3424.550	
Rho-Squared w.r.t. Zero	0.5227		0.5293		0.5266		0.5315	
Rho-Squared w.r.t. Constants	0.1558		0.1675		0.1627		0.1714	
Adjusted Rho-Squared w.r.t. Zero	0.5202		0.5262		0.5234		0.5277	
Adjusted Rho-Squared w.r.t. Constants	0.1524		0.1630		0.1581		0.1656	
Likelihood Ratio Test versus Model 11W	NA		97.2, 5, < 0.001		57.0, 5, < 0.001		129.2, 10, < 0.001	
Likelihood Ratio Test Model 15W vs. 13W	NA		NA		NA		32.0, 5, < 0.001	
Likelihood Ratio Test Model 15W vs. 14W	NA		NA		NA		72.2, 5, < 0.001	

Table 6-10 Implied Values of Time in Models 13W, 14W, and 15W

	Model 13W	Model 14W	Model 15W
Value of Motorized IVT	\$5.22/hr	\$6.22/hr	\$5.88/hr
Value of Motorized OVT (10 mile trip)	\$7.96/hr	\$9.50/hr	\$9.25/hr
(20 mile trip)	\$6.59/hr	\$7.86/hr	\$7.57/hr
Value of Non-Motorized Time	\$8.48/hr	\$9.54/hr	\$11.88/hr

Each of the new Models (13W, 14W and 15W) significantly reject Model 11W as the true model at a very high level of significance. Further, the parameters for all of the alternative specific CBD dummy and employment density variables have a positive sign, implying that all else being equal, an individual is less likely to choose drive alone mode for trips destined to a CBD and/or high employment density zones, as expected.

Since Models 13W and 14W are restricted versions of Model 15W, we can use the log-likelihood test which rejects the hypothesis that each of these models is the true model. Therefore, purely on statistical grounds, Model 15W is preferred over Models 13W and 14W. However, this improvement in statistical fit comes at the cost of increased model complexity, and it may be appropriate to adopt Model 13W or 14W, sacrificing statistical fit in favor of parsimony³⁷. For now, we choose Model 15W as the preferred model for its stronger statistical results, but we will return to the issue of model complexity.

6.2.5 Interactions between Trip Maker and/or Context Characteristics and Mode Attributes

Another approach to the inclusion of trip maker or context characteristics is through interactions with mode attributes. The most common example of this approach is to take account of the expectation that low-income travelers will be more sensitive to travel cost than high-income travelers by using cost divided by income in place of cost as an explanatory variable. Such a specification implies that the importance of cost in mode choice diminishes with increasing

³⁷ Parsimony emphasizes the use of less extensive specifications to reduce the burden of forecasting predictive variables and to provide simpler model interpretation.

household income. Table 6-11 portrays the estimation results for two models that differ only in how they represent cost; Model 15W includes travel cost while Model 16W includes travel cost divided by income.

Table 6-11 Comparison of Models with and without Income as Interaction Term

Variables	Model 15W		Model 16W	
Travel Cost (1990 cents)	-0.0024	(-7.7)		
Travel Cost by Income (1990 cents per 1000 1990 dollars)			-0.0518	(-4.8)
Travel Time (minutes)				
Motorized Modes Only	-0.0231	(-5.9)	-0.0202	(-5.3)
Non-Motorized Modes Only	-0.0467	(-8.1)	-0.0455	(-7.9)
Out-of-vehicle Travel Time by Distance (miles) Motorized Modes	-0.1324	(-7.9)	-0.1328	(-6.8)
Income (1,000's of 1990 dollars)				
Drive Alone (Base)	0		0	
Shared Ride 2	-0.002	(-1.3)	-1E-04	(-0.1)
Shared Ride 3+	-0.002	(-1.3)	-1E-04	(-0.1)
Transit	-0.007	(-3.4)	-0.005	(-2.6)
Bike	-0.011	(-2.1)	-0.009	(-1.7)
Walk	-0.008	(-2.5)	-0.006	(-1.9)
Autos per Worker				
Drive Alone	0		0	
Shared Ride 2	-0.401	(-5.2)	-0.382	(-5.0)
Shared Ride 3+	-0.183	(-1.7)	-0.139	(-1.3)
Transit	-0.93	(-7.9)	-0.938	(-7.9)
Bike	-0.715	(-2.8)	-0.704	(-2.7)
Walk	-0.727	(-4.3)	-0.724	(-4.3)
CBD Dummy (1 = in CBD; 0 = not in CBD)				
Drive Alone (Base)	0		0	
Shared Ride 2	0.204	(1.6)	0.247	(2.0)
Shared Ride 3+	1.018	(5.3)	1.094	(5.7)
Transit	1.204	(7.2)	1.306	(7.9)
Bike	0.462	(1.3)	0.486	(1.3)
Walk	0.109	(0.4)	0.098	(0.4)
Employ. Density at Work Zone (employees per square mile)				
Drive Alone (Base)	0		0	
Shared Ride 2	0.001	(2.5)	0.0016	(4.1)
Shared Ride 3+	0.0013	(2.7)	0.0022	(4.8)
Transit	0.0021	(5.5)	0.0031	(8.6)
Bike	0.0008	(0.7)	0.0019	(1.6)
Walk	0.0018	(2.4)	0.0029	(3.9)

Variables	Model 15W	Model 16W
Constants		
Drive Alone (Base)	0	0
Shared Ride 2	-1.64 (-12.0)	-1.73 (-12.5)
Shared Ride 3+	-3.55 (-17.4)	-3.656 (-17.7)
Transit	-0.471 (-1.9)	-0.692 (-2.8)
Bike	-1.515 (-3.5)	-1.622 (-3.8)
Walk	0.21 (0.6)	0.075 (0.2)
Log-likelihood at Zero	-7309.601	-7309.601
Log-likelihood at Constant	-4132.916	-4132.916
Log-likelihood at Convergence	-3424.550	-3442.334
Rho-Squared w.r.t. Zero	0.5315	0.5291
Rho-Squared w.r.t. Constants	0.1714	0.1671

The cost by income variable has the expected sign and is statistically significant, but the overall goodness-of-fit for the cost divided by income model is lower than that for model 15 that uses cost without interaction with income. However, because theory and common sense suggest that the importance of cost should decrease with income, we may choose Model 16W despite the differences in the goodness-of-fit statistics. Since the estimation results contradict our understanding of the decision making behavior, it is useful to consider other aspects of model results. In the case of mode choice, we are particularly interested in the relative value of the time and cost parameters because it measures the implied value of time used by travelers in choosing their travel mode. Values of time evaluated with earlier models were somewhat lower than expected when compared to the average wage rate. Using the cost by income formulation in Model 16W, we can calculate the implied value of time using the relationship developed in Section 5.8.2.

The implied values of IVT and OVT from Model 16W are substantially higher than those from Model 15W (Table 6-12) and more in line with our *a priori* expectations. This improvement in the estimate of values of time more than offsets the difference in goodness-of-fit so we adopt Model 16W as our preferred specification. Thus, our strong belief in both valuing time relative to wage rate and higher estimates of the value of time provide evidence which is strong enough to override the statistical test results. Nonetheless, we may still decide to impose parameter constraints to obtain higher values of time.

Table 6-12 Implied Value of Time in Models 15W and 16W

	Model 15W	Model 16W Wage Rate = \$21.20
Value of In-Vehicle Time	\$5.88/hr	$0.47 \times \text{Wage Rate}$ (\$9.90/hr)
Value of Out-of-Vehicle Time (10 mile trip)	\$9.25/hr	$0.77 \times \text{Wage Rate}$ (\$16.42/hr)
(20 mile trip)	\$7.57/hr	$0.62 \times \text{Wage Rate}$ (\$13.16/hr)

6.2.6 Additional Model Refinement

Generally, it is appropriate to test the preferred model specification against a variety of other specifications; particularly reviewing decisions made earlier in the model development process. Such testing would include reducing model complexity by the elimination of selected variables (*e.g.*, dropping either the CBD Dummy or Employment Density variables or combining some of the alternative specific parameters), changing the form used for inclusion of different variables (*e.g.*, replacing income by log of income) or adding new variables which substantially improve the explanatory power and behavioral realism of the model.

In this section, we consider simplifying the model specification by dropping variables that are not statistically significant or by collapsing alternative specific variables that do not differ across alternatives. The cost and time parameters are all significant and should be included because they represent the impact of policy changes in mode service attributes. Among the traveler and context variables, those for income have the lowest t-statistics so might be considered for elimination; however, we prefer to keep these in the model since income differences are important in mode selection, particularly for transit. However, the extremely low

values and lack of significance for the shared ride alternatives suggest that income has no differential impact on the choice of drive alone versus any of the shared ride alternatives and these variables should be dropped from the model (or constrained to zero). In addition, the parameter for the number of automobiles by number of workers variable for shared ride 3+ alternative is smaller in magnitude than the parameter for the shared ride 2 alternative. This is counter-intuitive as we expect shared ride 3+ travelers to be more sensitive to automobile availability. This can be addressed by constraining the alternative specific variables for the shared ride modes to be equal (we accomplish by summing the two variables). The estimation results for the simplified specification (constraining income for the shared ride alternatives to zero, and constraining the automobile ownership by number of workers variable for the two shared ride alternatives to be equal) and Model 16W are reported in Table 6-13.

The goodness-of-fit for the two models are very close, suggesting that the constraints imposed to simplify the model do not significantly impact the explanatory power of the model. The results of the likelihood ratio test confirm that the restrictions imposed in Model 17W cannot be statistically rejected. The parameter estimates for all the variables have the right sign and are all statistically significant (except CBD dummy for bike and walk). We therefore select Model 17W as our preferred model.

As discussed in the next section, the other major approach to searching for improved models is market segmentation and segmenting the population into groups which are expected to use different criteria in making their mode choice decisions.

Table 6-13 Estimation Results for Model 16W and its Constrained Version

Variables	Model 16W		Model 17W	
Travel Cost by Income (1990 cents per 1000 1990 \$)	-0.052	(-4.8)	-0.052	(-5.0)
Travel Time (minutes)				
Motorized Modes Only	-0.02	(-5.3)	-0.02	(-5.3)
Non-Motorized Modes Only	-0.045	(-7.9)	-0.045	(-7.9)
Out-of-vehicle Travel Time by Distance in Miles (Motorized Modes)	-0.133	(-6.8)	-0.133	(-6.8)

Variables	Model 16W		Model 17W	
Income (1990 dollars)				
Drive Alone (Base)	0		0	
Shared Ride 2	0.0001	(-0.1)	0	
Shared Ride 3+	0.0001	(-0.1)	0	
Transit	-0.005	(-2.6)	-0.005	(-2.7)
Bike	-0.009	(-1.7)	-0.009	(-1.7)
Walk	-0.006	(-1.9)	-0.006	(-1.9)
Autos per Worker				
Drive Alone (Base)	0		0	
Shared Ride 2	-0.382	(-5.0)	-0.317	(-4.8)
Shared Ride 3+	-0.139	(-1.3)	-0.317	(-4.8)
Transit	-0.938	(-7.9)	-0.946	(-8.0)
Bike	-0.704	(-2.7)	-0.702	(-2.7)
Walk	-0.724	(-4.3)	-0.722	(-4.3)
CBD Dummy (1 = in CBD; 0 = not in CBD)				
Drive Alone (Base)	0		0	
Shared Ride 2	0.247	(2.0)	0.26	(2.1)
Shared Ride 3+	1.094	(5.7)	1.069	(5.6)
Transit	1.306	(7.9)	1.309	(7.9)
Bike	0.486	(1.3)	0.489	(1.4)
Walk	0.098	(0.4)	0.102	(0.4)
Empl. Density - Work Zone (employees / square mile)				
Drive Alone (Base)	0		0	
Shared Ride 2	0.0016	(4.1)	0.0016	(4.0)
Shared Ride 3+	0.0022	(4.8)	0.0023	(5.0)
Transit	0.0031	(8.6)	0.0031	(8.7)
Bike	0.0019	(1.6)	0.0019	(1.6)
Walk	0.0029	(3.9)	0.0029	(3.9)
Constants				
Drive Alone (Base)	0		0	
Shared Ride 2	-1.73	(-12.5)	-1.808	(-17.0)
Shared Ride 3+	-3.656	(-17.7)	-3.434	(-22.6)
Transit	-0.692	(-2.8)	-0.685	(-2.8)
Bike	-1.622	(-3.8)	-1.629	(-3.8)
Walk	0.075	(0.2)	0.068	(0.2)
Log-likelihood at Zero	-7309.601		-7309.601	
Log-likelihood at Constant	-4132.916		-4132.916	
Log-likelihood at Convergence	-3442.334		-3444.185	
Rho-Squared w.r.t. Zero	0.5291		0.5288	
Rho-Squared w.r.t. Constants	0.1671		0.1666	
Likelihood Ratio Test versus Model 16W	NA		3.8, 3, 0.28	

6.3 Market Segmentation

The models considered to this point implicitly assume that the entire population, represented by the sample, uses the same model decision structure, variable and importance weights (parameters) to select their commute to work mode. That is, we assume that the population is

homogeneous with respect to the importance it places on different aspects of service except as differentiated by decision-maker characteristics included in the model specification. If this assumption is incorrect, the estimated model will not adequately represent the underlying decision processes of the entire population or of distinct behavioral groups within the population. For example, mode preference may differ between low and high-income travelers as low-income travelers are expected to be more sensitive to cost and less sensitive to time than high-income travelers. This phenomenon is incorporated in the preceding models to a limited extent through the use of alternative specific income variables and cost divided by income in the utility specification. Market segmentation can be used to determine whether the impact of other variables is different among population groups. The most common approach to market segmentation is for the analyst to consider sample segments which are mutually exclusive and collectively exhaustive (that is, each case is included in one and only one segment). Models are estimated for the sample associated with each segment and compared to the pooled model (all segments represented by a single model) to determine if there are statistically significant and important differences among the market segments.

Market segmentation is usually based on socio-economic and trip related variables such as income, auto ownership and trip purpose which may be used separately or jointly. Trip purpose has already been used in our analysis by considering work commute trips exclusively. Once segmentation variables are selected (income, auto ownership, etc.), different numbers of segments may be considered for each dimension (*e.g.*, we could use high, medium and low income segments or only high and low income segments). All members of each segment are assumed to have identical preferences and identical sensitivities to all the variables in the utility equation.

Analysts will often have some *a priori* ideas about the best segmentation variables and the appropriate groupings of the population with respect to these variables. In the case of continuous variables, such as income, the analyst may consider different boundaries between segments. In cases where the analyst does not have a strong basis for selecting model segments, he/she can test different combinations of socio-economic and trip-related variables in the data for segmentation. This approach is limited by the fact that the number of segments grows very fast

with the number of segmentation variables (*e.g.*, three income segments, two gender segments and three home location segments results in 18 distinct groups). The multiplicity of segments creates interpretational problems due to the complexity of comparing results among segments and estimation problems due to the small number of observations in some of the segments (with as many as 2,000 cases, eighteen segments would be likely to produce many segments with fewer than 100 cases and some with fewer than 50 cases, well below the threshold for reliable estimation results). The alternative of pre-defining market segments along one dimension at a time is practical and easy to implement but it has the disadvantage that this approach does not account for interactions among the segmentation variables.

6.3.1 Market Segmentation Tests

The determination of whether to segment the data is based on a comparison of the pooled model for the entire sample/population and a set of segment specific models for each segment of the sample/population. This comparison includes: (1) a statistical test, referred to as the market segmentation or taste variation test, to determine if the segments are statistically different from one another, (2) statistical significance and reasonableness of the parameters in each of the segments, and (3) reasonableness of the relationships among parameters in each segment and between parameters in the different market segments.

The statistical test for market segmentation consists of three steps. First, the sample is divided into a number of market segments which are mutually exclusive and collectively exhaustive. A preferred model specification is used to estimate a pooled model for the entire data set and to estimate models for each market segment. Finally, the goodness-of-fit differences between the segmented models (taken as a group) and the pooled model are evaluated to determine if they are statistically different. This test is an extension of the likelihood ratio test described earlier to test the difference between two models. In this case, the unrestricted model is the set of all the segmented models and the restricted model is the pooled model which imposes the restriction that the parameters for each segment are identical.

Thus, the null hypothesis is that $\underline{\beta}_1 = \underline{\beta}_2 = \dots = \underline{\beta}_s = \dots = \underline{\beta}_S$, where $\underline{\beta}_s$, is the vector of coefficients for the s^{th} market segment. Following the approach described in CHAPTER 5, we reject the null hypothesis that the restricted model is the correct model at significance level p if the calculated value of the statistic is greater than the test or critical value. That is, if:

$$-2 \times [\ell_R - \ell_U] \geq \chi_{n,(p)}^2 \tag{6.8}$$

Substituting the log-likelihood for the pooled model for ℓ_R and the sum of market segment model log-likelihoods for ℓ_U in equation 5.16, the null hypothesis, that all segments have the same choice function, is rejected at level p if:

$$-2 \times \left[\ell(\beta) - \sum_{s=1}^S \ell(\beta_s) \right] \geq \chi_{n,(p)}^2 \tag{6.9}$$

- where $\ell(\beta)$ is the log-likelihood for the pooled model,
- $\ell(\beta_s)$ is the log-likelihood of the model estimated with s^{th} market segment,
- χ_n^2 is the chi-square distribution with n degrees of freedom,
- n is equal to the number of restrictions, $\sum_{s=1}^S K_s - K$
- K is the number of coefficients in the pooled model, and
- K_s is the number of coefficients in the s^{th} market segment model.

K_s is generally equal to K in which case n is given by $K \times (S-1)$ ³⁸.

38 If one or more segments is defined so that one or more of the variables is fixed for all members of the segment, the parameters for that segment, K_s , will be fewer than K . For example, if none of the members of the low income group owned cars in income segmentation, it would not be possible to estimate parameters for the effect of auto ownership in that segment.

6.3.2 Market Segmentation Example

We illustrate the market segmentation test for two cases, automobile ownership (zero/one car households and households with more than one car), and gender (male and female). In the case of segmentation by automobile ownership, it is appealing to include a distinct segment for households with no cars since the mode choice behavior of this segment is very different from the rest of the population due to their dependence on non-automobile modes. However, the small size of this segment in the data set, only 160 of the 5029 work trip reports from households with no cars, precludes use of a no car segment; this group is combined with the one car ownership households for estimation. Using the same utility specification as in Model 17W, the estimation results for the pooled and segmented models for auto ownership and for gender are reported in Table 6-14 and Table 6-15.

We can make the following observations from the estimation results of the automobile ownership segmentation models (Table 6-14):

- The segmented model rejects the pooled model at a very high level of statistical significance.

$$-2 \times \left[\ell(\beta) - \sum_{s=1}^S \ell(\beta) \right] = -2 \times [-3444.2 - (-1049.3 - 2296.7)] = 196.4$$

- The alternative specific constants for all other modes relative to drive alone are much more negative for the higher auto ownership group than for the lower auto ownership group. These differences indicate the increased preference for drive alone among persons from multi-car households. This makes intuitive sense, as travelers in households with fewer automobiles are more likely to choose non-automobile modes, all else being equal.
- The alternative specific income coefficients are insignificant or marginally significant for both segments suggesting that the effect of income differences is adequately explained by the segment difference.

Table 6-14 Estimation Results for Market Segmentation by Automobile Ownership

Variables	Pooled Model		0-1 Car HH's		2+ Car HH's	
Travel Cost by Income (1990 cents per 1000 1990 dollar)	-0.052	(-5.0)	-0.023	(-1.6)	-0.098	(-6.1)
Travel Time (minutes)						
Motorized Modes	-0.02	(-5.3)	-0.021	(-3.5)	-0.019	(-3.6)
Non-Motorized Modes	-0.045	(-7.9)	-0.044	(-5.4)	-0.045	(-5.2)
OVT by Distance Motorized Modes	-0.133	(-6.8)	-0.113	(-4.4)	-0.194	(-5.9)

Variables	Pooled Model	0-1 Car HH's	2+ Car HH's
Income (1,000's of 1990 dollars)			
Drive Alone (Base), Shared Ride 2 and Shared Ride 3+	0	0	0
Transit	-0.005 (-2.7)	-0.007 (-1.8)	0.0004 (0.1)
Bike	-0.009 (-1.7)	-0.001 (-1.2)	-0.002 (-0.3)
Walk	-0.006 (-1.9)	-0.001 (-2.0)	0.0007 (0.2)
Autos per Worker			
Drive Alone (Base)	0	0	0
Shared Ride 2 and Shared Ride 3+	-0.317 (-4.8)	-3.015 (-8.6)	-0.241 (-3.3)
Transit	-0.946 (-8.0)	-3.963 (-10.5)	-0.24 (-1.8)
Bike	-0.702 (-2.7)	-2.664 (-4.0)	-0.191 (-0.6)
Walk	-0.722 (-4.3)	-3.32 (-7.5)	-0.098 (-0.5)
CBD Dummy (1 = in CBD; 0 = not in CBD)			
Drive Alone (Base)	0	0	0
Shared Ride 2	0.26 (2.1)	0.372 (1.5)	0.163 (1.1)
Shared Ride 3+	1.069 (5.6)	0.229 (0.6)	1.33 (6.0)
Transit	1.309 (7.9)	1.106 (4.3)	1.279 (5.2)
Bike	0.489 (1.4)	0.395 (0.7)	0.487 (1.0)
Walk	0.102 (0.4)	0.03 (0.1)	0.111 (0.3)
Empl. Density - Work Zone (employees per square mile)			
Drive Alone (Base)	0	0	0
Shared Ride 2	0.0016 (4.0)	0.002 (2.8)	0.0011 (2.2)
Shared Ride 3+	0.0023 (5.0)	0.0035 (3.9)	0.0013 (2.5)
Transit	0.0031 (8.7)	0.0032 (4.7)	0.0029 (6.4)
Bike	0.0019 (1.6)	0.0015 (0.8)	0.0016 (1.0)
Walk	0.0026 (3.9)	0.0038 (3.9)	0.0001 (-0.1)
Constants			
Drive Alone (Base)	0	0	0
Shared Ride 2	-1.808 (-17.0)	0.593 (2.0)	-1.979 (-15.4)
Shared Ride 3+	-3.434 (-22.6)	-0.765 (-2.2)	-3.717 (-20.0)
Transit	-0.685 (-2.8)	2.258 (5.1)	-2.163 (-5.6)
Bike	-1.629 (-3.8)	0.977 (1.4)	-3.218 (-4.4)
Walk	0.061 (0.2)	2.907 (5.2)	-1.535 (-2.7)
Log-likelihood at Zero	-7309.601	-1775.420	-5534.280
Log-likelihood at Constant	-4132.916	-1309.145	-2716.215
Log-likelihood at Convergence	-3444.185	-1049.280	-2296.667
Rho-Squared w.r.t. Zero	0.5288	0.4090	0.5850
Rho-Squared w.r.t. Constants	0.1666	0.1985	0.1544
Sample Size	5029	1221	3808
Likelihood Ratio Test versus Pooled model		196.4, 26, < 0.001	

- The sensitivity to automobile availability is much higher among low auto ownership households where an increase in availability (from 0) will be relatively important, than among higher auto ownership households where the number of cars is likely to closely

approximate the number of drivers and an increase in availability will be relatively unimportant.

- The differences in the alternative specific CBD dummy variables and the Employment Density variables are very small and not significant suggesting that these variables could be constrained to be equal across auto ownership segments.
- The differences in the time parameters also are very small and not significant suggesting that these variables could be constrained to be equal across auto ownership segments.
- The magnitude of the cost by income parameter is much smaller in the lower automobile ownership segment than in the higher automobile ownership segment indicating that cost may be of little importance in households with low car availability.

We can make the following observations from the estimation results of the gender segmentation models (Table 6-15):

- The segmented model rejects the pooled model at a very high level of statistical significance.
- The alternative specific constants relative to the drive alone mode are less negative (more positive) in the female segment suggesting the preference for drive alone mode is less pronounced among females. This is especially true for the non-motorized modes (bike and walk) where the difference in the modal constants between the two groups is large and highly significant.

Table 6-15 Estimation Results for Market Segmentation by Gender

Variables	Pooled Model		Males Only		Females Only	
Travel Cost by Income (1990 cents per 1000 1990 dollar)	-0.0524	(-5.0)	-0.064	(-4.4)	-0.0437	(-2.9)
Travel Time (minutes)						
Motorized Modes Only	-0.0202	(-5.3)	-0.0195	(-3.7)	-0.0191	(-3.4)
Non-Motorized Modes Only	-0.0454	(-7.9)	-0.0245	(-3.3)	-0.0703	(-7.5)
OVT by Distance (miles) for Motorized Modes	-0.1329	(-6.8)	-0.1865	(-6.0)	-0.09	(-3.6)
Income (1,000's of 1990 dollars)						
Drive Alone (Base), SR2 and SR3+	0		0		0	
Transit	-0.0053	(-2.7)	-0.0021	(-0.8)	-0.0089	(-3.0)
Bike	-0.0086	(-1.7)	-0.0014	(-0.2)	-0.0378	(-2.7)
Walk	-0.006	(-1.9)	-0.005	(-1.1)	-0.0049	(-1.1)

Variables	Pooled Model	Males Only	Females Only
Autos per Worker			
Drive Alone (Base)	0	0	0
Shared Ride 2 and Shared Ride 3+	-0.3166 (-4.8)	-0.21 (-2.8)	-0.607 (-4.5)
Transit	-0.9462 (-8.0)	-0.833 (-5.3)	-1.173 (-6.2)
Bike	-0.7021 (-2.7)	-0.992 (-3.1)	-0.056 (-0.1)
Walk	-0.7218 (-4.3)	-0.611 (-2.7)	-0.904 (-3.3)
CBD Dummy (1 = in CBD; 0 = not in CBD)			
Drive Alone (Base)	0	0	0
Shared Ride 2	0.26 (2.1)	0.028 (0.2)	0.454 (2.5)
Shared Ride 3+	1.069 (5.6)	1.424 (6.0)	0.377 (1.1)
Transit	1.309 (7.9)	1.196 (5.0)	1.381 (6.0)
Bike	0.489 (1.4)	0.311 (0.7)	1.039 (1.7)
Walk	0.102 (0.4)	0.224 (0.6)	0.081 (0.1)
Empl. Density - Work Zone (employees per square mile)			
Drive Alone (Base)	0	0	0
Shared Ride 2	0.0016 (4.0)	0.0009 (1.7)	0.003 (4.5)
Shared Ride 3+	0.0023 (5.0)	0.0006 (1.0)	0.0051 (6.6)
Transit	0.0031 (8.7)	0.0025 (5.5)	0.0046 (7.0)
Bike	0.0019 (1.6)	0.0005 (0.3)	0.0041 (1.9)
Walk	0.0026 (3.9)	0.0013 (1.2)	0.0055 (4.7)
Constants			
Drive Alone (Base)	0	0	0
Shared Ride 2	-1.808 (-17.0)	-1.912 (-14.2)	-1.564 (-8.6)
Shared Ride 3+	-3.434 (-22.6)	-3.551 (-17.9)	-3.199 (-13.0)
Transit	-0.685 (-2.8)	-0.865 (-2.4)	-0.477 (-1.3)
Bike	-1.629 (-3.8)	-1.928 (-3.6)	-1.153 (-1.5)
Walk	0.061 (0.2)	-1.221 (-2.4)	1.305 (2.6)
Log-likelihood at Zero	-7309.601	-4068.809	-3240.792
Log-likelihood at Constant	-4132.916	-2239.877	-1884.644
Log-likelihood at Convergence	-3444.185	-1889.784	-1511.319
Rho-Squared w.r.t. Zero	0.5288	0.5355	0.5337
Rho-Squared w.r.t. Constants	0.1666	0.1563	0.1981
Sample Size	5029	2842	2187
Likelihood Ratio Test versus Pooled Model		86.2, 26, < 0.001	

- The female segment parameters for alternative specific variables; Income, Autos per Worker, CBD Dummy and Employment Density are generally more favorable to non-auto modes and especially bike and walk, but the differences are small and marginally or not significant.
- Both groups show almost identical sensitivity to motorized in-vehicle travel time. However, the female group is more sensitive to non-motorized travel time while the male group is more sensitive to out-of-vehicle time.
- The female segment exhibits a much lower sensitivity to cost than males.

The above observations demonstrate that taste variations exist between the auto ownership segments and between the gender segments. However, in each case, the differences appear to be associated with a subset of parameters. One approach to simplifying the segmentation is to adopt a pooled model which includes segment related parameters where the differences are important³⁹. For example, such a model would at a minimum include different parameters for each of the segment for the following variables:

- Travel cost by income,
- Total travel time for non-motorized modes, and
- Out-of-vehicle time by distance.

6.4 Summary

This chapter demonstrates the development of an MNL model specification for work mode to choice using data from the San Francisco Bay Area for a realistic context. We start with relatively simple model specifications and develop more complex models which provide additional insight into the behavioral choices being made. We begin with the variables: travel cost, total travel time and household income. We then develop a more comprehensive model which includes: 1) cost divided by income to account for travelers different sensitivity to cost depending on household income, 2) two variables for time by motorized vehicle (which capture the constraint that OVTT is valued less for longer trips than shorter trips but is valued more highly than IVTT for all trip distances) and an additional variable for non-motorized personal transport (walk and bike), 3) alternative specific income variables, 4) number of autos per worker in the household, 5) location of the work zone (CBD or not), and 6) employment density of the work location.

The specification search was not necessarily exhaustive and improvements to the final preferred model specification are possible. The example describes the basis for the decisions made at each point in the model specification search process. Clearly, different decisions could

³⁹ For a more extensive discussion see Chapter 7, Section 7.5, in Ben-Akiva and Lerman [1985].

be made at some of these points. Thus, the final model result is based on a complex mix of empirical results, statistical analysis and judgment. The challenge to the analyst is to make good judgment, describe the basis for the judgments made, and be prepared to demonstrate the implications of making different judgments.

In the next chapter, we extend our work to consideration of home-based shop/other trips and we consider adoption of the more sophisticated nested logit model.

CHAPTER 7: San Francisco Bay Area Shop/Other Mode Choice

7.1 Introduction

This chapter extends the work of the preceding chapter by application of mode choice model development to the choice of mode for shop/other purpose trips in the San Francisco Bay Area. The modal alternatives in this case are more complex because a substantial share of the sample uses different modes to and from the shopping destination. The specific alternatives considered include:

- Drive Alone: Car driver both outbound to and return from shopping,
- Shared Ride 2: Shared ride with one other person both outbound to and return from shopping,
- Shared Ride 3+: Shared ride with two or more other people both outbound to and return from shopping,
- Shared Ride 2+ & Drive Alone: Shared ride with one or more other people outbound to or return from shopping and drive alone on the other trip,
- Shared Ride 2/3+: Shared ride with one or other person outbound to or return from shopping and shared ride 3+ on the other trip,
- Bike
- Walk

The frequency of mode availability and choice for these alternatives in the sample of 3157 cases is shown in Table 7-1.

Table 7-1 Sample Statistics for Bay Area Home-Based Shop/Other Trip Modal Data

Mode	Fraction of Sample with Mode Available	Market Share	Average IVTT (minutes)	Average OVTT (minutes)	Average Cost (1990 cents)
1. Transit	94.9%	1.2%	13.7	82.2	92.0
2. Shared Ride 2	100.0%	22.2%	7.6	1.8	31.0
3. Shared Ride 3+	100.0%	10.9%	7.6	1.8	17.6
4. Shared Ride 2+ & Drive Alone	100.0%	17.5%	7.6	1.8	43.4
5. Shared Ride 2/3+	100.0%	3.7%	7.6	1.8	24.1
6. Bike	67.7%	0.9%	12.6	0.0	0.0
7. Walk	60.4%	5.1%	34.2	0.0	0.0
8. Drive Alone	97.4%	38.6%	7.8	1.7	62.3

We will build on what has been learned in the preceding chapter to simplify the variety of specifications testing while recognizing that the factors that determine mode choice may be different or have different impacts when applied to travel for a different purpose. In addition to building on intuition, statistical analysis and testing, and judgment; we will be building on the experience of the estimation undertaken in the preceding chapter.

In this case, we adopt the approach of starting with a moderately advanced model specification and explore variations through addition, deletion or substitution of variables to identifying preferred specification.

7.2 Specification for Shop/Other Mode Choice Model

The choice of mode for shop/other trips is determined differently than the choice of mode for work trips. Nonetheless, some of the things we learned in the preceding estimation are likely to be relevant to this choice. Thus, we begin with a specification that reflects some of the results of the previous section. We have five groups of variables to consider for specification of the Shop/Other Mode Choice model. These are alternative specific constants, measures of household size, measures of vehicle availability, measures of income, various specifications of time components and different specifications of travel cost. Our initial specification includes the alternative specific constants and one selection from each other group of variables, as well as a few additional alternative specific dummy variables we found helpful. Subsequently, we estimate alternative forms from each major group of variables. After assessing the results of each of these estimations, we select a combined specification including the best specification results from each of the groups of variables.

7.3 Initial Model Specification

The first model includes alternative specific constants, the number of persons in the household (alternative specific), the number of vehicles in the household (alternative specific), household income (alternative specific), three measures of travel time (non-motorized time, motorized time and motorized out of vehicle time divided by trip distance) and travel cost. The results of this estimation are reported in Table 7-2.

Some of the most interesting observations are:

- The effect of persons per household increases the utility of every mode relative to drive alone. All of these parameters are significantly different from zero.
- The opposite effect is observed for number of vehicles with the greatest negative impact for the transit alternative. Again, all of these parameters are highly significant.
- The effect of household income (in 1990 \$000) is small and insignificant suggesting that any wealth effect is adequately captured by the automobile ownership variable.

- The travel time and cost variables are all negative, as expected; however, the value of in vehicle time for motorized modes is low for this time period.
- The zero vehicle ownership variables, constrained to be equal for transit, bike and walk is positive and significant as expected.
- The core destination variable is positive and significant for transit but small and not significant for all other alternatives relative to drive alone.
- Overall, the goodness of fit is reasonable but not as high as for the work mode choice. However, we can reasonably expect that improvements in the model will substantially improve the goodness of fit.

Table 7-2 Base Shopping/Other Mode Choice Model

Variables	Model 1 S/O	
Constants		
Transit	0.373	(0.5)
Shared Ride 2	-1.20	(-8.4)
Shared Ride 3+	-2.88	(-13.6)
Shared Ride 2+ & Drive Alone	-2.00	(-12.3)
Shared Ride 2/3+	-3.86	(-12.3)
Bike	-4.69	(-7.7)
Walk	0.304	(0.7)
Drive Alone (base)	0.00	----
Persons per Household		
Transit	0.733	(5.4)
Shared Ride 2	0.364	(8.6)
Shared Ride 3+	0.915	(18.3)
Shared Ride 2+ & Drive Alone	0.639	(14.4)
Shared Ride 2/3+	0.815	(11.8)
Bike	0.798	(5.9)
Walk	0.535	(6.8)
Drive Alone (base)	0.00	----
Number of Vehicles		
Transit	-1.84	(-5.8)
Shared Ride 2	-0.182	(-3.4)
Shared Ride 3+	-0.768	(-9.4)
Shared Ride 2+ & Drive Alone	-0.422	(-6.8)
Shared Ride 2/3+	-0.651	(-5.4)
Bike	-0.376	(-1.8)
Walk	-0.679	(-5.0)
Drive Alone (base)	0.00	----

Variables	Model 1 S/O	
Household Income (1,000's of 1990 dollars)		
Transit	-0.0062	(-0.8)
Shared Ride 2	-0.0023	(-1.4)
Shared Ride 3+	0.0016	(0.7)
Shared Ride 2+ & Drive Alone	0.0017	(0.9)
Shared Ride 2/3+	0.0034	(1.1)
Bike	-0.0010	(0.0)
Walk	-0.0049	(-1.4)
Drive Alone (base)	0.00	----
Travel Time (minutes)		
Non-Motorized Modes Only	-0.0853	(-8.1)
IVTT, Motorized Modes	-0.0343	(-3.1)
OVT by Distance (mi.), Motorized Modes	-0.215	(-4.1)
Travel Cost (1990 cents)	-0.00345	(-3.7)
Zero Vehicle Household Dummy Variable		
Transit, Bike and Walk	1.63	(4.4)
Dummy Variable for Destination in Core		
Transit	1.94	(3.0)
Shared Ride 2	0.492	(1.1)
Shared Ride 3+	-0.0481	(-0.1)
Shared Ride 2+ & Drive Alone	1.42	(3.4)
Shared Ride 2/3+	-25.0 ⁴⁰	(0.0)
Bike	1.25	(1.1)
Walk	0.147	(0.2)
Drive Alone (base)	0.0	----
Log-likelihood at Zero	-6201.516	
Log-likelihood at Constants	-4962.194	
Log-likelihood at Convergence	-4486.445	
Rho Squared w.r.t Zero	0.2766	
Rho Squared w.r.t. Constants	0.0959	

⁴⁰ The large negative parameter effectively drives the probability of choosing Shared Ride 2/3+ to zero for trips to the core of the region. This is consistent with the data as no trips to the core are made by Shared Ride 2/3+.

Table 7-3 Implied Value of Time in Base S/O Model

Value of Time Components	Base S/O Model
Non-Motorized Time	\$14.83/hr
In-Vehicle Time	\$5.97/hr
Out-of-Vehicle Time (10 mile trip)	\$9.70/hr
(20 mile trip)	\$7.83/hr

Although several of the model parameters are not statistically significant, and could be removed or combined, they are left for the time being while alternative specifications are explored in part because their significance may be affected by other variables.

7.4 Exploring Alternative Specifications

The base model specification for shop/other trips will be enhanced by considering alternative specification for each group of variables as stated earlier. The first examination is to consider alternative specifications for persons per household. Table 7-4 shows the results for the base model, a second model that separates the effect of increasing from one to two persons to increases beyond two persons (partially recognizing that the third and additional persons are likely to be minors or dependent adults and a third model in which persons is replaced for $\ln(\text{persons})$). It is obvious that both new specifications give substantially better goodness of fit indicating that the effect of increasing household size should be treated non-linearly. In this case either of these representations, the natural log transformation or a spline (larger effect for the second person than for additional persons) produce almost identical goodness of fit. This suggests that one of these two variations should be considered for inclusion in the final model

specification. Neither of the changes has any substantial impact on other model parameters except for the constants which are expected to change with any change in model specification.

Table 7-4 Alternative Specifications for Household Size

Variables	Model 1 S/O		Model 2 S/O		Model 3 S/O	
Constants						
Transit	0.373	(0.5)	0.763	(0.9)	0.696	(0.9)
Shared Ride 2	-1.20	(-8.4)	-1.84	(-8.7)	-1.10	(-7.9)
Shared Ride 3+	-2.88	(-13.6)	-2.88	(-7.2)	-3.15	(-13.0)
Shared Ride 2+ & Drive Alone	-2.00	(-12.3)	-1.65	(-7.9)	-1.83	(-11.1)
Shared Ride 2/3+	-3.86	(-12.3)	-4.93	(-4.9)	-4.11	(-11.2)
Bike	-4.69	(-7.7)	-4.00	(-5.1)	-4.44	(-6.9)
Walk	0.304	(0.7)	0.837	(2.0)	0.581	(1.4)
Drive Alone (base)	0.00		0.00		0.00	
Household Size						
	Persons		2+ Persons DV		Log of Persons	
Transit	0.733	(5.4)	1.65	(2.8)	2.11	(5.4)
Shared Ride 2	0.364	(8.6)	1.84	(8.6)	1.16	(10.0)
Shared Ride 3+	0.915	(18.3)	2.18	(5.4)	3.19	(18.2)
Shared Ride 2+ & Drive Alone	0.639	(14.4)	1.16	(5.4)	1.90	(14.3)
Shared Ride 2/3+	0.815	(11.8)	3.14	(3.1)	2.89	(11.2)
Bike	0.798	(5.9)	1.12	(1.3)	2.29	(4.8)
Walk	0.535	(6.8)	0.581	(2.0)	1.37	(6.3)
Drive Alone (base)	0.00		0.00		0.00	
Additional Person beyond 2						
Transit			0.585	(3.6)		
Shared Ride 2			0.206	(4.5)		
Shared Ride 3+			0.797	(15.2)		
Shared Ride 2+ & Drive Alone			0.547	(11.5)		
Shared Ride 2/3+			0.680	(9.3)		
Bike			0.721	(5.0)		
Walk			0.527	(6.0)		
Drive Alone (base)			0.00			
Number of Vehicles						
Transit	-1.84	(-5.8)	-2.01	(-6.0)	-1.91	(-5.9)
Shared Ride 2	-0.182	(-3.4)	-0.234	(-4.4)	-0.233	(-4.3)
Shared Ride 3+	-0.768	(-9.4)	-0.801	(-9.8)	-0.835	(-10.1)
Shared Ride 2+ & Drive Alone	-0.422	(-6.8)	-0.444	(-7.1)	-0.456	(-7.3)
Shared Ride 2/3+	-0.651	(-5.4)	-0.696	(-5.7)	-0.740	(-6.0)
Bike	-0.376	(-1.8)	-0.405	(-1.9)	-0.388	(-1.8)
Walk	-0.679	(-5.0)	-0.691	(-5.1)	-0.669	(-5.0)
Drive Alone (base)	0.00		0.00		0.00	
Household Income (1,000's of 1990 dollars)						
Transit	-0.0062	(-0.8)	-0.0068	(-0.8)	-0.0072	(-0.9)
Shared Ride 2	-0.0023	(-1.4)	-0.0035	(-2.1)	-0.0033	(-1.9)
Shared Ride 3+	0.0016	(0.7)	0.0008	(0.3)	-0.0006	(-0.3)
Shared Ride 2+ & Drive Alone	0.0017	(0.9)	0.0011	(0.6)	0.0004	(0.2)
Shared Ride 2/3+	0.0034	(1.1)	0.0023	(0.7)	0.0015	(0.5)
Bike	0.0001	(0.0)	-0.0006	(-0.1)	-0.0015	(-0.2)
Walk	-0.0049	(-1.4)	-0.0053	(-1.5)	-0.0055	(-1.5)
Drive Alone (base)	0.00		0.00		0.00	

Variables	Model 1 S/O		Model 2 S/O		Model 3 S/O	
Travel Time (minutes)						
Non-Motorized Modes Only	-0.0853	(-8.1)	-0.0850	(-8.0)	-0.0850	(-8.0)
Motorized Modes Only	-0.0343	(-3.1)	-0.0340	(-3.1)	-0.0343	(-3.1)
OVT by Distance (mi.) Motorized Modes	-0.215	(-4.1)	-0.215	(-4.1)	-0.214	(-4.1)
Travel Cost (1990 cents)	-0.0035	(-3.7)	-0.0036	(-3.7)	-0.0037	(-3.9)
Zero Vehicle Household Dummy Variable						
Transit, Bike and Walk	1.63	(4.4)	1.36	(3.5)	1.58	(4.2)
Dummy Variable for Destination in Core						
Transit	1.94	(3.0)	1.86	(2.9)	1.88	(2.9)
Shared Ride 2	0.492	(1.1)	0.397	(0.9)	0.443	(1.0)
Shared Ride 3+	-0.0481	(-0.1)	-0.124	(-0.2)	-0.113	(-0.2)
Shared Ride 2+ & Drive Alone	1.42	(3.4)	1.36	(3.2)	1.38	(3.3)
Shared Ride 2/3+	-25.0	(0.0)	-12.7	(0.0)	-26.7	(0.0)
Bike	1.25	(1.1)	1.19	(1.1)	1.22	(1.1)
Walk	0.147	(0.2)	0.0223	(0.0)	0.0358	(0.0)
Drive Alone (base)	0.00		0.00		0.00	
Log-likelihood at Zero	-6201.516		-6201.516		-6201.516	
Log-likelihood at Constants	-4962.194		-4962.194		-4962.194	
Log-likelihood at Convergence	-4486.445		-4445.800		-4445.078	
Rho Squared w.r.t Zero	0.2766		0.2831		0.2832	
Rho Squared w.r.t. Constants	0.0959		0.1041		0.1042	

The next variation to be considered is the replacement of the number of cars by the number of cars per person or the number of cars per worker. As can be seen in Table 7-5, the use of number of cars per person substantially improves model goodness of fit while the effect of number of cars per worker results in poorer goodness of fit. We would prefer cars per driver or per adults or at least cars per person greater than five years of age but these are not available in the current data. Nonetheless there is a clear indication that some normalization by household size (which relates to travel needs) should be included in the model.

Table 7-5 Alternative Specifications for Vehicle Availability

Variables	Model 1 S/O		Model 4 S/O		Model 5 S/O	
Constants						
Transit	0.373	(0.5)	1.76	(2.0)	0.713	(0.9)
Shared Ride 2	-1.20	(-8.4)	-0.437	(-2.0)	-1.22	(-7.9)
Shared Ride 3+	-2.88	(-13.6)	-0.84	(-2.8)	-3.25	(-14.1)
Shared Ride 2+ & Drive Alone	-2.00	(-12.3)	-0.866	(-3.5)	-1.95	(-11.0)
Shared Ride 2/3+	-3.86	(-12.3)	-1.77	(-4.0)	-4.08	(-11.7)
Bike	-4.69	(-7.7)	-4.55	(-5.8)	-4.79	(-7.4)
Walk	0.304	(0.7)	0.802	(1.6)	0.331	(0.8)
Drive Alone (base)	0.00		0.00		0.00	

Variables	Model 1 S/O	Model 4 S/O	Model 5 S/O
Persons per Household			
Transit	0.733 (5.4)	0.129 (0.9)	0.452 (3.6)
Shared Ride 2	0.364 (8.6)	0.164 (3.8)	0.305 (7.9)
Shared Ride 3+	0.915 (18.3)	0.398 (7.8)	0.738 (16.3)
Shared Ride 2+ & Drive Alone	0.639 (14.4)	0.303 (6.6)	0.522 (13.0)
Shared Ride 2/3+	0.815 (11.8)	0.316 (4.2)	0.654 (10.2)
Bike	0.798 (5.9)	0.573 (4.3)	0.688 (5.5)
Walk	0.535 (6.8)	0.185 (2.3)	0.366 (5.1)
Drive Alone (base)	0.00	0.00	0.00
Vehicle Availability			
	Autos per HH	Autos per Person	Autos per Worker
Transit	-1.84 (-5.8)	-3.44 (-4.6)	-1.65 (-4.2)
Shared Ride 2	-0.182 (-3.4)	-0.715 (-5.0)	-0.105 (-1.8)
Shared Ride 3+	-0.768 (-9.4)	-2.84 (-10.4)	-0.269 (-3.2)
Shared Ride 2+ & Drive Alone	-0.422 (-6.8)	-1.28 (-7.1)	-0.332 (-4.7)
Shared Ride 2/3+	-0.651 (-5.4)	-2.69 (-6.6)	-0.296 (-2.2)
Bike	-0.376 (-1.8)	-0.278 (-0.5)	-0.197 (-0.8)
Walk	-0.679 (-5.0)	-0.897 (-2.9)	-0.503 (-3.4)
Drive Alone (base)	0.00	0.00	0.00
Household Income (1,000's of 1990 dollars)			
Transit	-0.0062 (-0.8)	-0.0114 (-1.4)	-0.0178 (-2.2)
Shared Ride 2	-0.0023 (-1.4)	-0.0019 (-1.2)	-0.0033 (-2.0)
Shared Ride 3+	0.0016 (0.7)	0.0011 (0.5)	-0.0038 (-1.8)
Shared Ride 2+ & Drive Alone	0.0017 (0.9)	0.0014 (0.8)	-0.0007 (-0.4)
Shared Ride 2/3+	0.0034 (1.1)	0.0037 (1.2)	-0.0008 (-0.3)
Bike	-0.0001 (0.0)	-0.001 (-0.2)	-0.002 (-0.3)
Walk	-0.0049 (-1.4)	-0.0075 (-2.1)	-0.009 (-2.6)
Drive Alone (base)	0.00	0.00	0.00
Travel Time (minutes)			
Non-Motorized Modes Only	-0.0853 (-8.1)	-0.0856 (-8.1)	-0.0856 (-8.1)
Motorized Modes Only	-0.0343 (-3.1)	-0.0340 (-3.0)	-0.0357 (-3.2)
OVT by Distance (mi.) Motorized Modes	-0.215 (-4.1)	-0.239 (-4.4)	-0.218 (-4.2)
Travel Cost (1990 cents)			
	-0.0035 (-3.7)	-0.0038 (-4.0)	-0.0034 (-3.7)
Zero Vehicle Household Dummy Variable			
Transit, Bike and Walk	1.63 (4.4)	2.48 (6.2)	1.78 (4.8)
All Private Vehicle Modes	0.00	0.00	0.00
Dummy Variable for Destination in Core			
Transit	1.94 (3.0)	1.93 (3.0)	1.92 (3.1)
Shared Ride 2	0.492 (1.1)	0.409 (0.9)	0.487 (1.1)
Shared Ride 3+	-0.0481 (-0.1)	-0.125 (-0.2)	0.0913 (0.1)
Shared Ride 2+ & Drive Alone	1.42 (3.4)	1.37 (3.3)	1.41 (3.4)
Shared Ride 2/3+	-25.0 (0.0)	-25.1 (0.0)	-24.9 (0.0)
Bike	1.25 (1.1)	1.23 (1.1)	1.24 (1.1)
Walk	0.147 (0.2)	0.136 (0.2)	0.107 (0.1)
Drive Alone (base)	0.00	0.00	0.00
Log-likelihood at Zero	-6201.516	-6201.516	-6201.516
Log-likelihood at Constants	-4962.194	-4962.194	-4962.194
Log-likelihood at Convergence	-4486.445	-4469.880	-4543.708
Rho Squared w.r.t Zero	0.2766	0.2792	0.2673
Rho Squared w.r.t. Constants	0.0959	0.0992	0.0843

The third variation, shown in Table 7-6, is an alternative specification for including household income. We consider income as a linear variable and as a log transformation. The

goodness of fit for both models is almost identical. Thus, the analyst can choose to use either specification. It is also reasonable for the analyst to reconsider this issue when the final model specification is being formulated.

Table 7-6 Alternative Specifications for Income

Variables	Model 1 S/O		Model 6 S/O	
Constants				
Transit	0.373	(0.5)	0.829	(0.8)
Shared Ride 2	-1.20	(-8.4)	-0.958	(-3.7)
Shared Ride 3+	-2.88	(-13.6)	-2.81	(-7.3)
Shared Ride 2+ & Drive Alone	-2.00	(-12.3)	-2.31	(-7.4)
Shared Ride 2/3+	-3.86	(-12.3)	-4.36	(-7.1)
Bike	-4.69	(-7.7)	-4.84	(-4.3)
Walk	0.304	(0.7)	0.605	(1.1)
Drive Alone (base)	0.00		0.00	
Persons per Household				
Transit	0.733	(5.4)	0.732	(5.4)
Shared Ride 2	0.364	(8.6)	0.366	(8.6)
Shared Ride 3+	0.915	(18.3)	0.913	(18.2)
Shared Ride 2+ & Drive Alone	0.639	(14.4)	0.636	(14.2)
Shared Ride 2/3+	0.815	(11.8)	0.813	(11.7)
Bike	0.798	(5.9)	0.797	(5.8)
Walk	0.535	(6.8)	0.532	(6.8)
Drive Alone (base)	0.00		0.00	
Number of Vehicles				
Transit	-1.84	(-5.8)	-1.84	(-5.7)
Shared Ride 2	-0.182	(-3.4)	-0.181	(-3.4)
Shared Ride 3+	-0.768	(-9.4)	-0.748	(-9.1)
Shared Ride 2+ & Drive Alone	-0.422	(-6.8)	-0.431	(-6.9)
Shared Ride 2/3+	-0.651	(-5.4)	-0.660	(-5.4)
Bike	-0.376	(-1.8)	-0.394	(-1.8)
Walk	-0.679	(-5.0)	-0.689	(-5.1)
Drive Alone (base)	0.00		0.00	
Household Income (1,000's of 1990 dollars)				
		Income		Log of Income
Transit	-0.0062	(-0.8)	-0.195	(-0.8)
Shared Ride 2	-0.0023	(-1.4)	-0.0969	(-1.3)
Shared Ride 3+	0.0016	(0.7)	-0.0045	(0.0)
Shared Ride 2+ & Drive Alone	0.0017	(0.9)	0.112	(1.3)
Shared Ride 2/3+	0.0034	(1.1)	0.187	(1.1)
Bike	-0.0001	(0.0)	0.049	(0.2)
Walk	-0.0049	(-1.4)	-0.144	(-1.1)
Drive Alone (base)	0.00		0.00	
Travel Time (minutes)				
Non-Motorized Modes Only	-0.0853	(-8.1)	-0.0847	(-8.0)
Motorized Modes Only	-0.0343	(-3.1)	-0.0343	(-3.1)
OVT by Distance (mi.) Motorized Modes	-0.215	(-4.1)	-0.218	(-4.1)
Travel Cost (1990 cents)				
	-0.0035	(-3.7)	-0.0034	(-3.7)

Variables	Model 1 S/O		Model 6 S/O	
Zero Vehicle Household Dummy Variable				
Transit, Bike, and Walk	1.63	(4.4)	1.59	(4.3)
All Private Vehicle Modes (base)	0.00		0.00	
Dummy Variable for Destination in Core				
Transit	1.94	(3.0)	1.91	(3.0)
Shared Ride 2	0.492	(1.1)	0.492	(1.1)
Shared Ride 3+	-0.0481	(-0.1)	-0.0499	(-0.1)
Shared Ride 2+ & Drive Alone	1.42	(3.4)	1.42	(3.4)
Shared Ride 2/3+	-25.0	(0.0)	-25.0	(0.0)
Bike	1.25	(1.1)	1.27	(1.1)
Walk	0.147	(0.2)	0.130	(0.2)
Drive Alone (base)	0.00		0.00	
Log-likelihood at Zero	-6201.516		-6201.516	
Log-likelihood at Constants	-4962.194		-4962.194	
Log-likelihood at Convergence	-4486.445		-4486.560	
Rho Squared w.r.t Zero	0.2766		0.2765	
Rho Squared w.r.t. Constants	0.0959		0.0959	

The fifth set of variables concerns travel time. In this case, we consider two alternatives for out-of-vehicle time: time divided by distance and time divided by the natural log of distance. In terms of goodness of fit, the results in Table 7-7 are almost identical. There is no strong reason to choose one formulation over the other. The analyst can select a preferred specification at this point of when the final specification is being developed.

Table 7-7 Alternative Specifications for Travel Time

Variables	Model 1 S/O		Model 7 S/O	
Constants				
Transit	0.373	(0.5)	0.525	(0.6)
Shared Ride 2	-1.20	(-8.4)	-1.20	(-8.4)
Shared Ride 3+	-2.88	(-13.6)	-2.88	(-13.7)
Shared Ride 2+ & Drive Alone	-2.00	(-12.3)	-2.00	(-12.3)
Shared Ride 2/3+	-3.86	(-12.3)	-3.87	(-12.3)
Bike	-4.69	(-7.7)	-4.64	(-7.6)
Walk	0.304	(0.7)	0.335	(0.8)
Drive Alone (base)	0.00		0.00	
Persons per Household				
Transit	0.733	(5.4)	0.729	(5.4)
Shared Ride 2	0.364	(8.6)	0.364	(8.6)
Shared Ride 3+	0.915	(18.3)	0.916	(18.3)
Shared Ride 2+ & Drive Alone	0.639	(14.4)	0.639	(14.4)
Shared Ride 2/3+	0.815	(11.8)	0.816	(11.8)
Bike	0.798	(5.9)	0.800	(5.9)
Walk	0.535	(6.8)	0.535	(6.8)
Drive Alone (base)	0.00		0.00	

Variables	Model 1 S/O		Model 7 S/O	
Number of Vehicles				
Transit	-1.84	(-5.8)	-1.77	(-5.6)
Shared Ride 2	-0.182	(-3.4)	-0.182	(-3.4)
Shared Ride 3+	-0.768	(-9.4)	-0.769	(-9.4)
Shared Ride 2+ & Drive Alone	-0.422	(-6.8)	-0.423	(-6.8)
Shared Ride 2/3+	-0.651	(-5.4)	-0.652	(-5.4)
Bike	-0.376	(-1.8)	-0.380	(-1.8)
Walk	-0.679	(-5.0)	-0.687	(-5.1)
Drive Alone (base)	0.00		0.00	
Household Income (1,000's of 1990 dollars)				
Transit	-0.0062	(-0.8)	-0.0065	(-0.8)
Shared Ride 2	-0.0023	(-1.4)	-0.0023	(-1.4)
Shared Ride 3+	0.0016	(0.7)	0.0016	(0.7)
Shared Ride 2+ & Drive Alone	0.0017	(0.9)	0.0017	(0.9)
Shared Ride 2/3+	0.0034	(1.1)	0.0034	(1.1)
Bike	-0.0001	(0.0)	-0.0001	(0.0)
Walk	-0.0049	(-1.4)	-0.0048	(-1.4)
Drive Alone (base)	0.00		0.00	
Travel Time (minutes)				
Non-Motorized Modes Only	-0.0853	(-8.1)	-0.0839	(-7.8)
Motorized Modes Only	-0.0343	(-3.1)	-0.0035	(-0.3)
OVT by Distance (mi.) Motorized Modes	-0.215	(-4.1)		
OVT by Log of Distance (mi.) Motorized Modes			-0.166	(-4.2)
Travel Cost (1990 cents)	-0.0035	(-3.7)	-0.0036	(-3.9)
Zero Vehicle Household Dummy Variable				
Transit, Bike, and Walk	1.63	(4.4)	1.66	(4.4)
All Private Vehicle Modes	0.00		0.00	
Dummy Variable for Destination in Core				
Transit	1.94	(3.0)	1.89	(3.0)
Shared Ride 2	0.492	(1.1)	0.470	(1.0)
Shared Ride 3+	-0.0481	(-0.1)	-0.0548	(-0.1)
Shared Ride 2+ & Drive Alone	1.42	(3.4)	1.41	(3.3)
Shared Ride 2/3+	-25.0	(0.0)	-26.7	(0.0)
Bike	1.25	(1.1)	1.31	(1.2)
Walk	0.147	(0.2)	0.226	(0.3)
Drive Alone (base)	0.00		0.00	
Log-likelihood at Zero	-6201.516		-6201.516	
Log-likelihood at Constants	-4962.194		-4962.194	
Log-likelihood at Convergence	-4486.445		-4486.394	
Rho Squared w.r.t Zero	0.2766		0.2766	
Rho Squared w.r.t. Constants	0.0959		0.0959	

The last specification variation to be considered is the adjustment of cost by income either as cost divided by income or cost divided by the natural log of income. There are strong theoretical reasons to adjust the effect of cost by household income. The non-nested hypothesis test rejects the cost divided by income specification at a very high level but is not able to reject the cost divided by $\ln(\text{income})$ at any reasonable level of significance. This and the strong

theoretical reasons for adjusting cost by income support the adoption of the cost by $\ln(\text{income})$ specification.

Table 7-8 Alternative Specifications for Cost

Variables	Model 1 S/O		Model 8 S/O		Model 9 S/O	
Constants						
Transit	0.373	(0.5)	0.536	(0.7)	0.482	(0.6)
Shared Ride 2	-1.20	(-8.4)	-1.16	(-8.1)	-1.22	(-8.5)
Shared Ride 3+	-2.88	(-13.6)	-2.83	(-13.3)	-2.92	(-13.6)
Shared Ride 2+ & Drive Alone	-2.00	(-12.3)	-1.98	(-12.1)	-2.01	(-12.4)
Shared Ride 2/3+	-3.86	(-12.3)	-3.81	(-12.1)	-3.89	(-12.3)
Bike	-4.69	(-7.7)	-4.69	(-7.7)	-4.73	(-7.7)
Walk	0.304	(0.7)	0.304	(0.7)	0.277	(0.7)
Drive Alone (base)	0.00		0.00		0.00	
Persons per Household						
Transit	0.733	(5.4)	0.738	(5.5)	0.739	(5.5)
Shared Ride 2	0.364	(8.6)	0.361	(8.6)	0.365	(8.6)
Shared Ride 3+	0.915	(18.3)	0.915	(18.3)	0.916	(18.3)
Shared Ride 2+ & Drive Alone	0.639	(14.4)	0.636	(14.3)	0.638	(14.4)
Shared Ride 2/3+	0.815	(11.8)	0.813	(11.8)	0.816	(11.8)
Bike	0.798	(5.9)	0.801	(5.9)	0.799	(5.9)
Walk	0.535	(6.8)	0.536	(6.8)	0.534	(6.8)
Drive Alone (base)	0.00		0.00		0.00	
Number of Vehicles						
Transit	-1.84	(-5.8)	-1.85	(-5.8)	-1.86	(-5.8)
Shared Ride 2	-0.182	(-3.4)	-0.178	(-3.4)	-0.180	(-3.4)
Shared Ride 3+	-0.768	(-9.4)	-0.762	(-9.3)	-0.764	(-9.3)
Shared Ride 2+ & Drive Alone	-0.422	(-6.8)	-0.420	(-6.8)	-0.422	(-6.8)
Shared Ride 2/3+	-0.651	(-5.4)	-0.646	(-5.4)	-0.648	(-5.4)
Bike	-0.376	(-1.8)	-0.371	(-1.8)	-0.371	(-1.8)
Walk	-0.679	(-5.0)	-0.681	(-5.0)	-0.677	(-5.0)
Drive Alone (base)	0.00		0.00		0.00	
Household Income (1,000's of 1990 dollars)						
Transit	-0.0062	(-0.8)	-0.0095	(-1.1)	-0.0084	(-1.0)
Shared Ride 2	-0.0023	(-1.4)	-0.0016	(-0.9)	-0.0017	(-1.0)
Shared Ride 3+	0.0016	(0.7)	0.0025	(1.1)	0.0024	(1.1)
Shared Ride 2+ & Drive Alone	0.0017	(0.9)	0.0021	(1.2)	0.0020	(1.1)
Shared Ride 2/3+	0.0034	(1.1)	0.0041	(1.3)	0.0041	(1.3)
Bike	-0.0001	(0.0)	0.0008	(0.1)	0.0007	(0.1)
Walk	-0.0049	(-1.4)	-0.0038	(-1.1)	-0.004	(-1.1)
Drive Alone (base)	0.00		0.00		0.00	
Travel Time (minutes)						
Non-Motorized Modes Only	-0.0853	(-8.1)	-0.0849	(-8.0)	-0.0855	(-8.1)
Motorized Modes Only	-0.0343	(-3.1)	-0.0370	(-3.3)	-0.0350	(-3.2)
OVT by Distance (mi.) Motorized Modes	-0.215	(-4.1)	-0.216	(-4.1)	-0.213	(-4.0)
Travel Cost (1990 cents)	-0.0035	(-3.7)				
Travel Cost by Income (1990 cents per 1000 1990 dollars)			-0.0349	(-2.2)		
Travel Cost by Log of Income (1990 cents per log of 1000 1990 dollars)					-0.0111	(-3.6)
Zero Vehicle Household Dummy Variable						
Transit, Bike, and Walk	1.63	(4.4)	1.63	(4.4)	1.61	(4.3)
All Private Vehicle Modes (base)	0.00		0.00		0.00	

Variables	Model 1 S/O		Model 8 S/O		Model 9 S/O	
Dummy Variable for Destination in Core						
Transit	1.94	(3.0)	2.20	(3.5)	1.96	(3.0)
Shared Ride 2	0.492	(1.1)	0.653	(1.4)	0.494	(1.1)
Shared Ride 3+	-0.0481	(-0.1)	0.150	(0.2)	-0.0578	(-0.1)
Shared Ride 2+ & Drive Alone	1.42	(3.4)	1.53	(3.6)	1.43	(3.4)
Shared Ride 2/3+	-25.0	(0.0)	-26.5	(0.0)	-26.7	(0.0)
Bike	1.25	(1.1)	1.41	(1.2)	1.2	(1.1)
Walk	0.147	(0.2)	0.277	(0.3)	0.0709	(0.1)
Drive Alone (base)	0.00		0.00		000	
Log-likelihood at Zero	-6201.516		-6201.516		-6201.516	
Log-likelihood at Constants	-4962.194		-4962.194		-4962.194	
Log-likelihood at Convergence	-4486.445		-4490.817		-4486.695	
Rho Squared w.r.t Zero	0.2766		0.2759		0.2765	
Rho Squared w.r.t. Constants	0.0959		0.0950		0.0958	
Adjusted Rho Squared w.r.t Zero	0.27011		0.2694		0.27007	
Non-nest hypothesis of cost by income and cost by ln(income)	-----		-8.2539		-0.7131	
	-----		<<0.001		0.237	

The next step in developing a preferred model specification is to combine the preferred alternative from each of the specifications considered to this point. The log of household size is selected over the spline variables, as it achieves the same goodness-of-fit with fewer variables. Autos per person is used for vehicle availability, and income and travel time are retained as modeled in the base model. Given the very similar goodness-of-fit of the unadjusted cost variable and the cost divided by the log of income, both are tested. The resulting models, Models 10 S/O and 11 S/O, are displayed in Table 7-9.

Table 7-9 also includes Model 3 S/O and a variation on it, dividing cost by the log of income, showing that Model 3 S/O has superior goodness-of-fit to both Models 10 S/O and 11 S/O. While this result may seem surprising, it illustrates the possible interactions which can occur between variables in a model and highlights the importance of reviewing earlier decisions in the development of the utility specification as the final specification is approached. Unfortunately, there is no approach, other than testing every possible combination of variables, to ensure that the best statistical model is obtained. However, the time and effort required to test every possible specification may not be justified when a simpler approach, with a reasonable, but limited, amount of checking alternative specifications is likely to produce nearly as good a model as the full factorial search. This is particularly true if one has a judgmental basis to guide the specification selection process.

Table 7-9 Composite Specifications from Earlier Results Compared with other Possible Preferred Specifications

Variables	Model 10 S/O		Model 11 S/O		Model 3 S/O		Model 12 S/O	
Constants								
Transit	1.79	(2.1)	1.91	(2.3)	0.696	(0.9)	0.803	(1.0)
Shared Ride 2	-0.659	(-3.2)	-0.700	(-3.4)	-1.10	(-7.9)	-1.13	(-8.1)
Shared Ride 3+	-1.30	(-4.0)	-1.36	(-4.1)	-3.15	(-13.0)	-3.20	(-13.1)
Shared Ride 2+ & Drive Alone	-0.914	(-3.7)	-0.933	(-3.8)	-1.83	(-11.1)	-1.84	(-11.2)
Shared Ride 2/3+	-2.38	(-4.8)	-2.42	(-4.9)	-4.11	(-11.2)	-4.15	(-11.3)
Bike	-4.33	(-5.2)	-4.38	(-5.3)	-4.44	(-6.9)	-4.48	(-7.0)
Walk	0.967	(2.0)	0.941	(1.9)	0.581	(1.4)	0.551	(1.4)
Drive Alone (base)	0.00		0.00		0.00		0.00	
Log of Persons per Household								
Transit	0.426	(1.1)	0.441	(1.2)	2.11	(5.4)	2.14	(5.5)
Shared Ride 2	0.668	(5.5)	0.677	(5.5)	1.16	(10.0)	1.17	(10.1)
Shared Ride 3+	1.55	(8.5)	1.56	(8.6)	3.19	(18.2)	3.19	(18.2)
Shared Ride 2+ & Drive Alone	0.965	(6.9)	0.968	(6.9)	1.90	(14.3)	1.90	(14.3)
Shared Ride 2/3+	1.37	(5.0)	1.39	(5.1)	2.89	(11.2)	2.90	(11.2)
Bike	1.68	(3.5)	1.69	(3.5)	2.29	(4.8)	2.29	(4.8)
Walk	0.399	(1.8)	0.397	(1.8)	1.37	(6.3)	1.36	(6.3)
Drive Alone (base)	0.00		0.00		0.00		0.00	
Number of Vehicles								
Transit					-1.91	(-5.9)	-1.93	(-6.0)
Shared Ride 2					-0.233	(-4.3)	-0.231	(-4.3)
Shared Ride 3+					-0.835	(-10.1)	-0.831	(-10.0)
Shared Ride 2+ & Drive Alone					-0.456	(-7.3)	-0.456	(-7.3)
Shared Ride 2/3+					-0.740	(-6.0)	-0.737	(-6.0)
Bike					-0.388	(-1.8)	-0.382	(-1.8)
Walk					-0.669	(-5.0)	-0.667	(-4.9)
Drive Alone (base)					0.00		0.00	
Vehicles per Person								
Transit	-3.57	(-4.7)	-3.62	(-4.7)				
Shared Ride 2	-0.567	(-4.0)	-0.559	(-3.9)				
Shared Ride 3+	-2.63	(-9.6)	-2.61	(-9.5)				
Shared Ride 2+ & Drive Alone	-1.22	(-6.7)	-1.21	(-6.6)				
Shared Ride 2/3+	-2.38	(-5.7)	-2.37	(-5.7)				
Bike	-0.386	(-0.7)	-0.371	(-0.6)				
Walk	-0.948	(-3.0)	-0.949	(-3.0)				
Drive Alone (base)	0.00		0.00					
Household Income (1,000's of 1990 dollars)								
Transit	-0.0123	(-1.5)	-0.0150	(-1.8)	-0.0072	(-0.9)	-0.0096	(-1.2)
Shared Ride 2	-0.0033	(-2.0)	-0.0026	(-1.6)	-0.0033	(-1.9)	-0.0026	(-1.5)
Shared Ride 3+	-0.0007	(-0.3)	0.0001	(0.0)	-0.0006	(-0.3)	0.0003	(0.1)
Shared Ride 2+ & Drive Alone	0.0003	(0.2)	0.0007	(0.4)	0.0004	(0.2)	0.0008	(0.4)
Shared Ride 2/3+	0.0017	(0.5)	0.0024	(0.7)	0.0015	(0.5)	0.0022	(0.7)
Bike	-0.0022	(-0.3)	-0.0013	(-0.2)	-0.0015	(-0.2)	-0.0006	(-0.1)
Walk	-0.0073	(-2.1)	-0.0064	(-1.8)	-0.0055	(-1.5)	-0.0046	(-1.3)
Drive Alone (base)	0.00		0.00		0.00		0.00	

Variables	Model 10 S/O		Model 11 S/O		Model 3 S/O		Model 12 S/O	
Travel Time (minutes)								
Non-Motorized Modes Only	-0.0846	(-8.0)	-0.0849	(-8.0)	-0.0850	(-8.0)	-0.0853	(-8.1)
Motorized Modes Only	-0.0331	(-2.9)	-0.0334	(-2.9)	-0.0343	(-3.1)	-0.0348	(-3.1)
OVT by Distance (mi.) Motorized Modes	-0.236	(-4.4)	-0.232	(-4.3)	-0.214	(-4.1)	-0.210	(-4.0)
Travel Cost (1990 cents)	-0.0039	(-4.0)			-0.0037	(-3.9)		
Travel Cost by Log of Income (1990 cents per log of 1000 1990 dollars)			-0.0128	(-4.0)			-0.0124	(-3.9)
Zero Vehicle Household Dummy Variable								
Transit, Bike, Walk	2.2	(5.4)	2.17	(5.3)	1.58	(4.2)	1.56	(4.2)
All Private Vehicle Modes	0.00		0.00		0.00		0.00	
Dummy Variable for Destination in Core								
Transit	1.94	(3.1)	1.95	(3.1)	1.88	(2.9)	1.87	(2.9)
Shared Ride 2	0.426	(0.9)	0.415	(0.9)	0.443	(1.0)	0.422	(0.9)
Shared Ride 3+	-0.107	(-0.2)	-0.107	(-0.2)	-0.113	(-0.2)	-0.127	(-0.2)
Shared Ride 2+ & Drive Alone	1.38	(3.3)	1.38	(3.3)	1.38	(3.3)	1.37	(3.3)
Shared Ride 2/3+	-25.0	(0.0)	-25.0	(0.0)	-26.7	(0.0)	-26.7	(0.0)
Bike	1.24	(1.1)	1.16	(1.0)	1.22	(1.1)	1.15	(1.0)
Walk	0.112	(0.1)	0.0105	(0.0)	0.0358	(0.0)	-0.0649	(-0.1)
Drive Alone (base)	0.00		0.00		0.00		0.00	
Log-likelihood at Zero	-6201.516		-6201.516		-6201.516		-6201.516	
Log-likelihood at Constants	-4962.194		-4962.194		-4962.194		-4962.194	
Log-likelihood at Convergence	-4455.014		-4455.143		-4445.078		-4444.851	
Rho Squared w.r.t Zero	0.2816		0.2816		0.2832		0.2833	
Rho Squared w.r.t. Constants	0.1022		0.1022		0.1042		0.1043	

Among the models in Table 7-9, Model 12 S/O is chosen as the preferred specification, because it's goodness of fit is slightly better than any of the other models, it obtains slightly better values of time than any other models and theoretical considerations support the use of cost divided by a function of income. Before accepting it as the final MNL model, however, it is important to return to the issue of the significance of the variables included. Variables that are not statistically significant are usually excluded to simplify the model unless there are important theoretical or policy reasons for their retention. However, case descriptive variables are generally included or excluded as a full set of alternative specific variables although this is not a firm rule and special cases are included in this chapter.

Therefore, Table 7-10 presents two final additional models which simplify Model 12 S/O by eliminating or combining insignificant variables. Model 13 S/O eliminates household income as an alternative specific variable, since none of these variables are significant and income is included in the model through its interaction with cost. Model 14 S/O combines the dummy

variables for destination in the core to improve the significance of these variables. The chi-squared log-likelihood test rejects Model 13 S/O but does not reject Model 14 S/O in favor of Model 12 S/O.

Table 7-10 Refinement of Final Specification Eliminating Insignificant Variables

Variables	Model 12 S/O		Model 13 S/O		Model 14 S/O	
Constants						
Transit	0.803	(1.0)	0.546	(0.7)	0.53	(0.7)
Shared Ride 2	-1.13	(-8.1)	-1.20	(-9.0)	-1.19	(-9.0)
Shared Ride 3+	-3.20	(-13.1)	-3.19	(-13.6)	-3.20	(-13.7)
Shared Ride 2+ & Drive Alone	-1.84	(-11.2)	-1.82	(-11.6)	-1.78	(-11.4)
Shared Ride 2/3+	-4.15	(-11.3)	-4.08	(-11.7)	-4.11	(-11.7)
Bike	-4.48	(-7.0)	-4.49	(-7.3)	-4.41	(-7.3)
Walk	0.551	(1.4)	0.462	(1.2)	0.461	(1.2)
Drive Alone (base)	0.00		0.00		0.00	
Log of Persons per Household						
Transit	2.14	(5.5)	2.10	(5.3)	2.10	(5.4)
Shared Ride 2	1.17	(10.1)	1.14	(9.9)	1.14	(9.9)
Shared Ride 3+	3.19	(18.2)	3.19	(18.3)	3.19	(18.3)
Shared Ride 2+ & Drive Alone	1.90	(14.3)	1.90	(14.5)	1.90	(14.5)
Shared Ride 2/3+	2.90	(11.2)	2.90	(11.3)	2.89	(11.3)
Bike	2.29	(4.8)	2.28	(4.8)	2.27	(4.8)
Walk	1.36	(6.3)	1.32	(6.2)	1.32	(6.2)
Drive Alone (base)	0.00		0.00		0.00	
Number of Vehicles						
Transit	-1.93	(-6.0)	-2.05	(-6.6)	-2.06	(-6.6)
Shared Ride 2	-0.231	(-4.3)	-0.250	(-4.7)	-0.248	(-4.7)
Shared Ride 3+	-0.831	(-10.0)	-0.826	(-10.3)	-0.819	(-10.3)
Shared Ride 2+ & Drive Alone	-0.456	(-7.3)	-0.448	(-7.4)	-0.456	(-7.5)
Shared Ride 2/3+	-0.737	(-6.0)	-0.712	(-6.0)	-0.700	(-6.0)
Bike	-0.382	(-1.8)	-0.383	(-1.9)	-0.401	(-2.0)
Walk	-0.667	(-4.9)	-0.715	(-5.5)	-0.714	(-5.5)
Drive Alone (base)	0.00		0.00		0.00	
Household Income (1,000's of 1990 dollars)						
Transit	-0.0096	(-1.2)				
Shared Ride 2	-0.0026	(-1.5)				
Shared Ride 3+	0.0003	(0.1)				
Shared Ride 2+ & Drive Alone	0.0008	(0.4)				
Shared Ride 2/3+	0.0022	(0.7)				
Bike	-0.0006	(-0.1)				
Walk	-0.0046	(-1.3)				
Drive Alone (base)	0.00					
Travel Time (minutes)						
Non-Motorized Modes Only	-0.0853	(-8.1)	-0.0850	(-8.0)	-0.0848	(-8.0)
Motorized Modes Only	-0.0348	(-3.1)	-0.0341	(-3.1)	-0.0344	(-3.1)
OVT by Distance (mi.) Motorized Modes	-0.210	(-4.0)	-0.208	(-3.9)	-0.208	(-3.9)
Travel Cost by Log of Income (1990 cents per log of 1000 1990 dollars)						
	-0.0124	(-3.9)	-0.0126	(-4.0)	-0.0111	(-3.6)
Zero Vehicle Household Dummy Variable						
Transit, Bike, Walk	1.56	(4.2)	1.56	(4.2)	1.57	(4.2)

Variables	Model 12 S/O		Model 13 S/O		Model 14 S/O	
All Private Vehicle Modes (base)	0.00		0.00		0.00	
Dummy Variable for Destination in Core						
Transit	1.87	(2.9)	1.87	(2.9)	2.01	(3.2)
Shared Ride 2	0.422	(0.9)	0.404	(0.9)	0.716	(1.8)
Shared Ride 3+	-0.127	(-0.2)	-0.147	(-0.2)	0.716	(1.8)
Shared Ride 2+ & Drive Alone	1.37	(3.3)	1.36	(3.2)	0.716	(1.8)
Shared Ride 2/3+	-26.7	(0.0)	-26.7	(0.0)	0.716	(1.8)
Bike	1.15	(1.0)	1.11	(1.0)	0.301	(0.2)
Walk	-0.0649	(-0.1)	-0.0736	(-0.1)	0.301	(0.2)
Drive Alone (base)	0.00		0.00		0.00	
Log-likelihood at Zero	-6201.516		-6201.516		-6201.516	
Log-likelihood at Constants	-4962.194		-4962.194		-4962.194	
Log-likelihood at Convergence	-4444.851		-4448.481		-4457.235	
Rho Squared w.r.t Zero	0.2833		0.2827		0.2813	
Rho Squared w.r.t. Constants	0.1043		0.1035		0.1018	
Chi-Squared vs. Model 12 S/O			7.2600		24.7680	
Confidence			59.8%		99.0%	

CHAPTER 8: Nested Logit Model

8.1 Motivation

The Multinomial Logit Model (MNL) structure has been widely used for both urban and intercity mode choice models primarily due to its simple mathematical form, ease of estimation and interpretation, and the ability to add or remove choice alternatives. However, the MNL model has been widely criticized for its Independence of Irrelevant Alternatives (IIA) property (see Section 4.2). The IIA property of the MNL restricts the ratio of the choice probabilities for any pair of alternatives to be independent of the existence and characteristics of other alternatives in the choice set. This restriction implies that introduction of a new mode or improvements to any existing mode will reduce the probability of existing modes in proportion to their probabilities before the change.

The IIA property is a major limitation of the MNL model as it implies equal competition between all pairs of alternatives, an inappropriate assumption in many choice situations. For example, in the case of urban mode choice among drive alone, shared ride, bus and light rail; the bus and light rail alternatives are likely to be more similar to each other than they are to either of the other alternatives due to shared attributes which are not included in the measured portion of the utility function; for example, bus and light rail may have the same fare structure and operating policies, the same lack of privacy, control of the environment, and so on. Such similarities, if not included in the measured portion of the utility function, lead to correlation between the errors associated with these alternatives, a violation of the assumptions which underlie the derivation of the MNL.

The way in which this undesirable characteristic of the IIA property manifests itself can be illustrated using this example. Assume that the choice probabilities (for an individual or a homogeneous group of individuals) are 65%, 15%, 10% and 10% for drive alone, shared ride, bus and light rail, respectively. If the light rail service were to be improved in such a way as to increase its choice probability to 19%, the MNL model would predict that the shares of the other alternatives would decrease proportionately as shown in Table 8-1, decreasing the probability for

the drive alone, shared ride and bus alternatives by a factor of 0.90. As a result, the MNL model predicts that most of the increased light rail ridership comes from drive alone (6.5%) while only 1.5% comes from car pool and 1% from bus. This is inconsistent with expectations and empirical evidence that most of the new light rail riders will be diverted from bus and carpool. This inconsistency is a direct result of the IIA property of the MNL model. Thus, in these types of choice situations, the MNL model will yield incorrect predictions of diversions from existing modes.

Table 8-1 Illustration of IIA Property on Predicted Choice Probabilities

Alternative	Choice Probability Before Improvements to LRT	Choice Probability After Improvements to LRT	Proportional Change in Choice Probabilities	Algebraic Change in Choice Probabilities
Drive Alone	0.650	0.585	0.900	-0.065
Carpool	0.150	0.135	0.900	-0.015
Bus	0.100	0.009	0.900	-0.010
Light Rail	0.100	0.190	1.900	+0.090

More extreme examples of this phenomenon have been described in the modeling literature. Most widely known among them is the red bus/blue bus problem described in section 4.2.1.

These and numerous other examples illustrate that the IIA property is difficult to justify in situations where some alternatives compete more closely with each other than they do with other alternatives. This limitation of the MNL model results from the assumption of the independence of error terms in the utility of the alternatives (section 3.5), which is used to derive the model. Different models can be derived through the use of different assumptions concerning the structure of the error distributions of alternative utilities. Among them, the Nested Logit (NL) model (Williams, 1977; McFadden, 1978; Daly and Zachary, 1978), is the simplest and most widely used. The NL model represents important deviations from the IIA property but retains

most of the computational advantages of the MNL model (Borsch-Supan, 1987). The NL model is characterized by grouping (or nesting) subsets of alternatives that are more similar to each other with respect to excluded characteristics than they are to other alternatives.

Alternatives in a common nest exhibit a higher degree of similarity and competitiveness than alternatives in different nests. This level of competitiveness, represented by cross-elasticities between pairs of alternatives (the impact of a change in one mode on the probability of another mode) is identical for all pairs of alternatives in the nest. Complex tree structures can be developed which offer substantial flexibility in representing differential competitiveness between pairs of alternatives; however, the nesting structure imposes a system of restrictions concerning relationships between pairs of alternatives as will be discussed later in this chapter.

8.2 Formulation of Nested Logit Model

The derivation of the nested logit model is based on the assumption that some of the alternatives share common components in their random error terms. That is, the random term of the nested alternatives can be decomposed into a portion associated with each alternative and a portion associated with groups of alternatives. For example, consider an urban mode choice where a traveler has four modes (drive alone, shared ride, commuter rail and bus) available for making an intercity trip. The utility equations for these alternatives are:

$$\begin{aligned}
 U_{DA} &= V_{DA} + \varepsilon_{DA} \\
 U_{SR} &= V_{SR} + \varepsilon_{SR} \\
 U_{Bus} &= V_{PT} + V_{Bus} + \varepsilon_{PT} + \varepsilon_{Bus} \\
 U_{LTR} &= V_{PT} + V_{LTR} + \varepsilon_{PT} + \varepsilon_{LTR}
 \end{aligned}
 \tag{8.1}$$

The utility terms for bus and light rail each include a distinct observed component, V_{Bus} and V_{LTR} , and a common observed component, V_{PT} , for public transit (PT); they also include distinct random components, ε_{Bus} and ε_{LTR} , and a common random component, ε_{PT} . The common error component creates a covariance between the total errors for bus, $\varepsilon_{PT} + \varepsilon_{Bus}$, and Light Rail, $\varepsilon_{PT} + \varepsilon_{LTR}$. This covariance violates the assumption underlying the MNL

model representing an increased similarity between pairs of nested alternatives (bus and Light Rail, in this case) and leads to greater cross-elasticity between these alternatives.

The total error for each of the four alternatives is assumed to be distributed Gumbel with scale parameter equal to one, as in the MNL model. The variance of these distributions is:

$$Var(\varepsilon_{DA}) = Var(\varepsilon_{SR}) = Var(\varepsilon_{PT} + \varepsilon_{Bus}) = Var(\varepsilon_{PT} + \varepsilon_{LTR}) = \frac{\pi^2}{6} \quad 8.2$$

The distinct error components, ε_{Bus} and ε_{LTR} , also are assumed to be distributed Gumbel, but

with scale parameter, μ_{PT} . However, in practice we estimate $\theta_{PT} = \frac{1}{\mu_{PT}}$, the inverse of the

Gumbel scale parameter. θ_{PT} , commonly referred to as the logsum parameter, is bounded by zero and one. That is, the variance of these distributions is:

$$Var(\varepsilon_{Bus}) = Var(\varepsilon_{LTR}) = \frac{\pi^2}{6\mu_{PT}^2} = \frac{\pi^2\theta_{PT}^2}{6} \quad 8.3$$

where μ_{PT} is bounded by one and positive infinity (θ_{PT} is bounded by zero and one) to ensure that the conditional variance for bus, $Var(\varepsilon_{Bus})$, is less than the total variance for bus, $Var(\varepsilon_{PT} + \varepsilon_{Bus})$, and similarly for Light Rail. This is required to ensure that the variance for the common public transit error component, ε_{PT} , is non-negative. These assumptions are adequate to derive a nested logit model using utility maximization principles. The choice structure implied by these equations is depicted by the nesting structure in Figure 8.3, in which bus and light rail are more similar to each other than they are to drive alone and shared ride.

It is convenient to interpret this structure as if there are two levels of choice even though the derivation of the model makes no assumptions about the structure of the choice process⁴¹. The figure depicts an upper level (marginal) choice among drive alone, shared ride, and public transit and a lower level (conditional) choice between bus and light rail, given that public transit is chosen.

41 The hierarchy of choice and the notion of a “decision tree” are purely analytical devices - they do not necessarily imply that an individual makes choices in a certain order (Borsch-Supan, 1987).

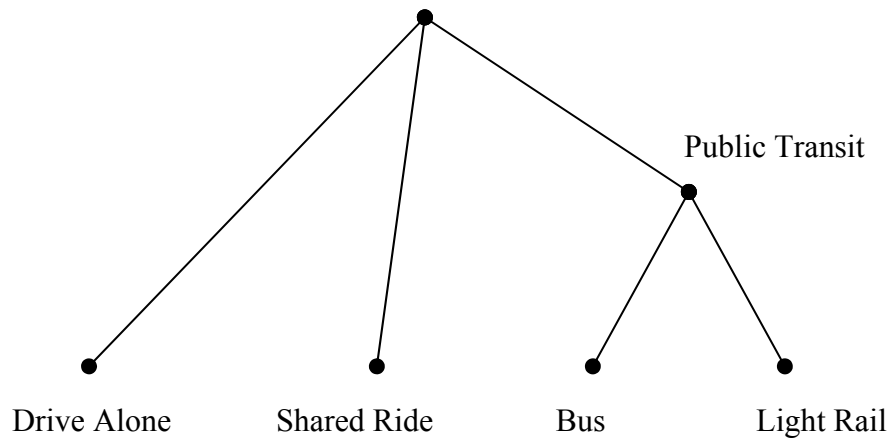


Figure 8.1 Two-Level Nest Structure with Two Alternatives in Lower Nest

The choice probabilities for the lower level nested alternatives (commuter rail or bus), conditional on choice of these alternatives are given by:

$$\Pr(Bus / PT) = \frac{\exp\left(\frac{V_{Bus}}{\theta_{PT}}\right)}{\exp\left(\frac{V_{Bus}}{\theta_{PT}}\right) + \exp\left(\frac{V_{LTR}}{\theta_{PT}}\right)} \quad 8.4$$

$$\Pr(LTR / PT) = \frac{\exp\left(\frac{V_{LTR}}{\theta_{PT}}\right)}{\exp\left(\frac{V_{Bus}}{\theta_{PT}}\right) + \exp\left(\frac{V_{LTR}}{\theta_{PT}}\right)} \quad 8.5$$

This is the standard logit form except for the inclusion of the logsum parameter in the denominator of each utility function. The marginal choice probabilities for the drive alone, shared ride, and public transit alternatives are:

$$\Pr(DA) = \frac{\exp(V_{DA})}{\exp(V_{DA}) + \exp(V_{SR}) + \exp(V_{PT} + \theta_{PT}\Gamma_{PT})} \quad 8.6$$

$$\Pr(SR) = \frac{\exp(V_{SR})}{\exp(V_{DA}) + \exp(V_{SR}) + \exp(V_{PT} + \theta_{PT}\Gamma_{PT})} \quad 8.7$$

$$\Pr(PT) = \frac{\exp(V_{PT} + \theta_{PT}\Gamma_{PT})}{\exp(V_{DA}) + \exp(V_{SR}) + \exp(V_{PT} + \theta_{PT}\Gamma_{PT})} \quad 8.8$$

where Γ_{PT} represents the expected value of the maximum of the bus and light rail utility.

The expected utility of the public transit alternatives equals this value, Γ_{PT} , times the logsum parameter, θ_{PT} , plus other attributes common to the pair of alternatives, V_{PT} . Γ_{PT} is computed from the log of the sum of the exponents of the nested utilities (equations 8.4 and 8.5), commonly referred to as the “logsum” variable.

$$\Gamma_{PT} = \log \left[\exp \left(\frac{V_{Bus}}{\theta_{PT}} \right) + \exp \left(\frac{V_{LTR}}{\theta_{PT}} \right) \right] \quad 8.9$$

An important feature of these equations is that the logsum parameter, θ_{PT} , appears in the denominator of the conditional utility for all the nested alternatives. The implication of this is that all of the utility function parameters, β , are scaled by a common value. Since θ_{PT} is bounded by zero and one, the magnitudes of all the resultant parameters, β / θ_{PT} , are increased implying that the choice between the nested alternatives is more sensitive to changes to any of the variables in these functions than are alternatives not in the nest. Further, if the number of alternatives in any nest is reduced to one, the utility in the marginal model becomes identically equal to the utility for the remaining alternative. That is, if the bus alternative is not available to some travelers, the utility of the nest becomes:

$$\begin{aligned}
 V_{PT}^* &= V_{PT} + \theta_{PT} \log \left[\exp \left(\frac{V_{LTR}}{\theta_{PT}} \right) \right] \\
 &= V_{PT} + V_{LTR}
 \end{aligned}
 \tag{8.10}$$

The probability of choosing the nested alternatives can be obtained by multiplying the conditional probability of the nested alternative by the marginal probability as follows:

$$\Pr(Bus) = \Pr(Bus | PT) \times \Pr(PT) \tag{8.11}$$

$$\Pr(LTR) = \Pr(LTR | PT) \times \Pr(PT) \tag{8.12}$$

8.2.1 Interpretation of the Logsum Parameter

The logsum parameter, θ , (sometimes called the “dissimilarity parameter” or the “nesting coefficient”), is a function of the underlying correlation between the unobserved components for pairs of alternatives in that nest, and it characterizes the degree of substitutability between those alternatives. The value of the logsum parameter is bounded by zero and one to ensure consistency with random utility maximization principles. Different values of the parameter indicate the degree of dissimilarity between pairs of alternatives in the nest. The interpretation of different values of the logsum parameter is as follows:

- $\theta > 1$ Not consistent with the theoretical derivation. Reject NL model.
- $\theta = 1$ Implies zero correlation among mode pairs in the nest so the NL model collapses to the MNL model.
- $0 < \theta < 1$ Implies non-zero correlation among pairs. This range of values is appropriate for the nested logit model. Decreasing values of θ indicate increased substitution between/among alternatives in the nest.
- $\theta = 0$ Implies perfect correlation between pairs of alternatives in the nest. That is, the choice between the nested alternatives, conditional on the nest, is deterministic.
- $\theta < 0$ Not consistent with the theoretical derivation. Therefore, we reject the nested logit model.

8.2.2 Disaggregate Direct and Cross-Elasticities

The differences between the nested logit model and the multinomial logit model can be illustrated by comparison of the elasticities of each alternative to changes in the value of a

variable associated with it (direct elasticity) or with another alternative (cross elasticity) as reported in Table 8-2. The MNL direct- and cross-elasticity equations are the same for all alternatives. This is a manifestation of the IIA property of the MNL model.

However, the elasticity expressions for the NL model are differentiated between cases in which the alternative being considered is or is not in the same nest as the alternative which is changed. Both models produce identical direct- and cross-elasticities when the attribute that is changed is for one of the non-nested alternatives. However, the elasticity equations for changes in the attributes of nested alternatives are different. These differences are attributable to the value of θ in the elasticity equations. When θ is equal to one, its maximum value, the expression, $\frac{1 - \theta}{\theta}$, in the elasticity formulae in Table 8-2 becomes zero and the direct and cross-elasticity expressions for the nested alternatives collapse to the corresponding equations for the alternatives not in the nest. As the scale parameter, θ , decreases from one to zero, this expression increases and the direct- and cross-elasticities within the nest become larger (possibly, much larger) than the direct- and cross-elasticities between the nests. That is, the sensitivity of a nested alternative to changes in its attributes or to changes in the attributes of other nested alternatives becomes much greater than the corresponding changes for non-nested alternatives. Although the same comparison appears to exist between nested alternatives and similar alternatives in the MNL model, this can only be evaluated by taking account of differences in the β parameters between the MNL and NL models.

Table 8-2 Elasticity Comparison of Nested Logit vs. MNL Models

Elasticity of Probability of Choosing Mode	Changes in Non-Nested Alternative, 'j'	Changes in Nested Alternative, 'k'
I. Direct Elasticity Multinomial Logit Nested Logit	$(1 - P_j)\beta_{LOS}LOS_j$ $(1 - P_j)\beta_{LOS}LOS_j$	Not Applicable $\left[(1 - P_k) + \left(\frac{1 - \theta_N}{\theta_N} \right) (1 - P_{k N}) \right]$ $\times \beta_{LOS}LOS_k$
II. Cross Elasticity <i>a.) Effect on Non-nested Alts.</i> Multinomial Logit Nested Logit <i>b.) Effect on Nested Alts.</i> Multinomial Logit Nested Logit	$-P_j\beta_{LOS}LOS_j$ $-P_j\beta_{LOS}LOS_j$ Not Applicable $-P_j\beta_{LOS}LOS_j$	Not Applicable $-P_k\beta_{LOS}LOS_k$ Not Applicable $-\left[P_k + \left(\frac{1 - \theta_N}{\theta_N} \right) (P_{k N}) \right]$ $\times \beta_{LOS}LOS_k$

Source: Modified from Forinash and Koppelman, 1993

8.3 Nesting Structures

The assumptions underlying the model described in the preceding section (correlation of error terms for the bus and light rail alternatives) results in a two-level model with a single public transit nest for the four alternative (drive alone, shared ride, commuter rail, and bus) mode

choice problem. This nesting structure is one of many possible two-level nested logit models which can be constructed for a choice set with four alternatives. The possible set of two-level nested logit models includes six combinations of two alternatives in a nest with the remaining alternatives at the upper level, four combinations of three alternatives in a nest and one at the upper level and three combinations of two alternatives in a nest and two alternatives in a parallel nest for a total of thirteen two-level nest structures (Figure 8.2). The number of two level nest structures increases rapidly with the number of alternatives as shown in Table 8-3 below.

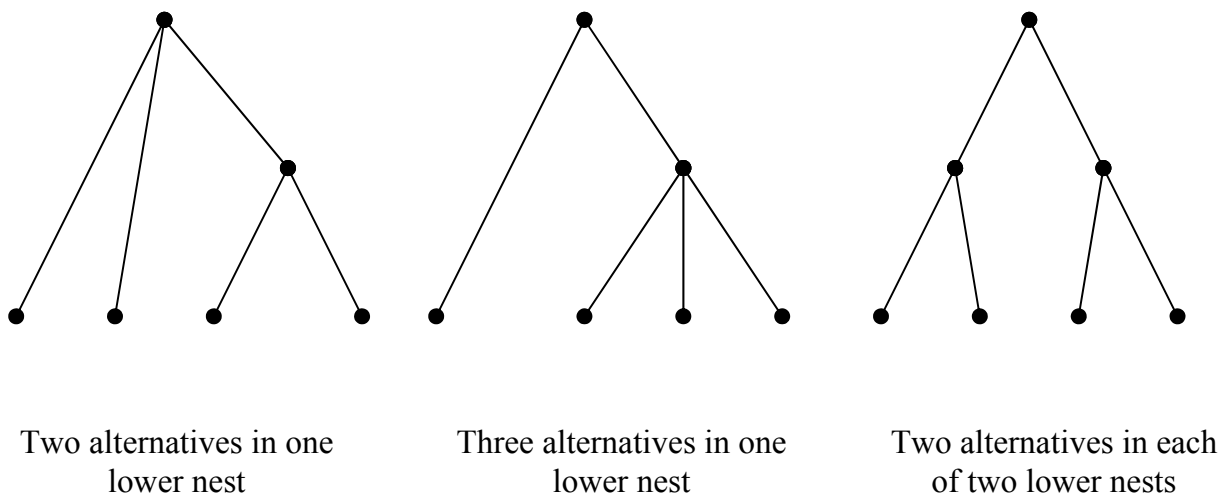


Figure 8.2 Three Types of Two Level Nests

Some of these nesting structures are likely to be behaviorally unreasonable; for example, it doesn't seem reasonable to include bus and car in the same nest. The selection of a preferred nesting structure requires a combination of judgment (about reasonable nesting structures) and statistical hypothesis testing; specifically, testing the hypothesis that the MNL or a simpler NL model is the true model.

Further, sub-groups of alternatives within any group may themselves be more similar to each other than to other alternatives in the larger group. Representation of these differences in similarity can result in multiple levels of nesting, hierarchically identifying increasingly similar alternatives at each level (Borsch-Supan, 1987). For example, in the four alternative case, the

case of three alternatives in a lower nest might represent alternatives that include group riding (shared ride, bus and light rail). If we believe that the bus and light rail alternatives are more similar than either alternative is to shared ride, utility equations with the following common error terms will show an intermediate level of error correlation among all group travel modes, and an additional level of error correlation between commuter rail and bus.

$$\begin{aligned}
 U_{DA} &= V_{DA} && + \varepsilon_{DA} \\
 U_{SR} &= V_{GRP} + V_{SR} && + \varepsilon_{GRP} + \varepsilon_{SR} \\
 U_{Bus} &= V_{GRP} + V_{PT} + V_{Bus} && + \varepsilon_{GRP} + \varepsilon_{PT} + \varepsilon_{Bus} \\
 U_{LTR} &= V_{GRP} + V_{PT} + V_{LTR} && + \varepsilon_{GRP} + \varepsilon_{PT} + \varepsilon_{LTR}
 \end{aligned}
 \tag{8.13}$$

In this case, the group travel modes (shared ride, commuter rail and bus) will be nested at the second level and commuter rail and bus will be nested at the third or lowest level as shown in Figure 8.3. In this model structure, the lower level nest is a binary choice between commuter rail and bus, conditional on choice of public transit; the second level nest represents a choice between shared ride and public transit conditional on group travel; and the highest level represents a choice between drive alone and group travel.

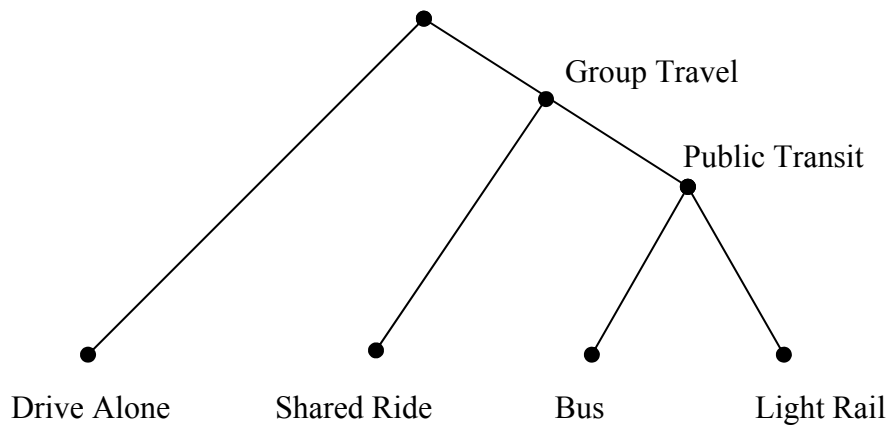


Figure 8.3 Three-Level Nest Structure for Four Alternatives

Twelve three-level nested structures are possible with four alternatives; these are combinations of two alternatives at the lowest level, one alternative at the intermediate level and one alternative at the upper level.

The probability equations for the two-level nested logit model can be extended readily to the three-level case as illustrated below. The probabilities for each nested alternative in the lowest level, commuter rail or bus, conditional on choice of public transit (PT) are given by:

$$\Pr(Bus | PT) = \frac{\exp\left(\frac{V_{Bus}}{\theta_{PT}}\right)}{\exp\left(\frac{V_{Bus}}{\theta_{PT}}\right) + \exp\left(\frac{V_{LTR}}{\theta_{PT}}\right)} \tag{8.14}$$

$$\Pr(LTR | PT) = \frac{\exp\left(\frac{V_{LTR}}{\theta_{PT}}\right)}{\exp\left(\frac{V_{Bus}}{\theta_{PT}}\right) + \exp\left(\frac{V_{LTR}}{\theta_{PT}}\right)} \tag{8.15}$$

where θ_{PT} is the logsum parameter at the lowest (*i.e.*, public transit) nest level.

The probabilities for each alternative in the second level nest, shared ride and public transit, conditional on the choice of group travel (GRP) modes are:

$$\Pr(SR / GRP) = \frac{\exp\left(\frac{V_{SR}}{\theta_{GRP}}\right)}{\exp\left(\frac{V_{SR}}{\theta_{GRP}}\right) + \exp\left(\frac{V_{PT} + \theta_{PT}\Gamma_{PT}}{\theta_{GRP}}\right)} \tag{8.16}$$

$$\Pr(PT / GRP) = \frac{\exp\left(\frac{V_{PT} + \theta_{PT}\Gamma_{PT}}{\theta_{GRP}}\right)}{\exp\left(\frac{V_{SR}}{\theta_{GRP}}\right) + \exp\left(\frac{V_{PT} + \theta_{PT}\Gamma_{PT}}{\theta_{GRP}}\right)} \tag{8.17}$$

where θ_{GRP} is the logsum parameter for the intermediate level (*i.e.*, for the group travel modes) and ,

Γ_{PT} is the “logsum” of the exponents of the nested utilities for the lower nest level:

$$\Gamma_{PT} = \log \left[\exp \left(\frac{V_{Bus}}{\theta_{PT}} \right) + \exp \left(\frac{V_{LTR}}{\theta_{PT}} \right) \right] \quad 8.18$$

Finally, the probabilities for automobile and the common carrier nest are:

$$\begin{aligned} \Pr(DA) &= \frac{\exp(V_{DA})}{\exp(V_{DA}) + \exp(V_{GRP} + \theta_{GRP}\Gamma_{GRP})} \\ \Pr(GRP) &= \frac{\exp(V_{GRP} + \theta_{GRP}\Gamma_{GRP})}{\exp(V_{DA}) + \exp(V_{GRP} + \theta_{GRP}\Gamma_{GRP})} \end{aligned} \quad 8.19$$

where θ_{GRP} is the logsum parameter for the intermediate level (*i.e.*, for the common carrier modes), and

Γ_{GRP} is the logsum of the exponents of the nested utilities for the intermediate nest:

$$\Gamma_{GRP} = \log \left\{ \exp \left(\frac{V_{SR}}{\theta_{GRP}} \right) + \exp \left(\frac{V_{PT} + \theta_{PT}\Gamma_{PT}}{\theta_{GRP}} \right) \right\} \quad 8.20$$

The marginal probabilities of shared ride, commuter rail and bus are the product of the probabilities of each branch from the root (top of the tree) to the alternative:

$$\begin{aligned} \Pr(SR) &= \Pr(SR | GRP) \times \Pr(GRP) \\ \Pr(Bus) &= \Pr(Bus | PT) \times \Pr(PT | GRP) \times \Pr(GRP) \\ \Pr(LTR) &= \Pr(LTR | PT) \times \Pr(PT | GRP) \times \Pr(GRP) \end{aligned} \quad 8.21$$

The value of the logsum parameters decrease as we go down the tree. This follows from the requirement that the error variance at each level of the tree must be lower than at the next higher level since the total variance for each alternative is fixed and the variance at each level of the tree must be positive (non-negative). The total variance for each alternative, $Var(\varepsilon_{Mode}^*)$, is given by:

$$Var(\varepsilon_{DA}^*) = \frac{\pi^2}{6} \tag{8.22}$$

$$\begin{aligned} Var(\varepsilon_{SR}^*) &= Var(\varepsilon_{SR} + \varepsilon_{GRP}) \\ &= Var(\varepsilon_{SR}) + Var(\varepsilon_{GRP}) = \frac{\pi^2}{6} \end{aligned} \tag{8.23}$$

$$\begin{aligned} Var(\varepsilon_{Bus}^*) &= Var(\varepsilon_{GRP} + \varepsilon_{PT} + \varepsilon_{Bus}) \\ &= Var(\varepsilon_{GRP}) + Var(\varepsilon_{PT}) + Var(\varepsilon_{Bus}) = \frac{\pi^2}{6} \end{aligned} \tag{8.24}$$

$$\begin{aligned} Var(\varepsilon_{LTR}^*) &= Var(\varepsilon_{GRP} + \varepsilon_{PT} + \varepsilon_{LTR}) \\ &= Var(\varepsilon_{GRP}) + Var(\varepsilon_{PT}) + Var(\varepsilon_{LTR}) = \frac{\pi^2}{6} \end{aligned} \tag{8.25}$$

The variance components for alternatives, group elements and public transit elements of each alternative are shown in the following figure.

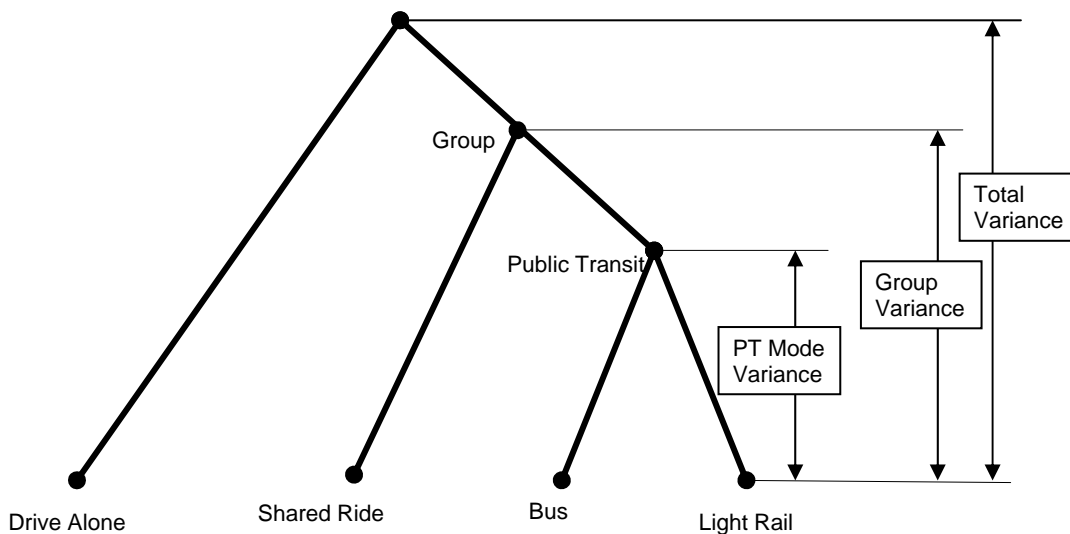


Figure 8.4 Decomposition of Error Variance

This figure helps to explain the hierarchical restrictions on logsum parameters. Section 8.2.1 described the initial restriction on the logsum for a single nest level or a two level tree as follows: $0 < \theta < 1$. However, in multi-level tree structures, as shown in Section 8.3, the logsum parameter at each level is restricted to be between zero and the logsum parameter at the

next higher level of the nesting structure. Thus, in the example in Figure 8.4 above, $0 < \theta_{Grp} < 1$ and $0 < \theta_{PT} < \theta_{Grp}$. This hierarchical restriction ensures that the variance of each error term in equations 8.22 through 8.25 is positive, as required. In particular, the variances are defined as

$$Var(\varepsilon_{Total}) = \frac{\pi^2}{6} \quad 8.26$$

$$Var(\varepsilon_{GRP}) = \frac{\pi^2 \theta_{Grp}^2}{6} \quad 8.27$$

$$Var(\varepsilon_{PT}) = \frac{\pi^2 \theta_{PT}^2}{6} \quad 8.28$$

$$Var(\varepsilon_{Total} - \varepsilon_{GRP}) = \frac{\pi^2}{6} (1 - \theta_{Grp}^2) \quad 8.29$$

$$Var(\varepsilon_{Grp} - \varepsilon_{PT}) = \frac{\pi^2}{6} (\theta_{Grp}^2 - \theta_{PT}^2) \quad 8.30$$

To ensure that all the variance components are positive, as required, the nesting parameters must be constrained as shown $\theta_{PT} < \theta_{Grp} < 1$. These hierarchical constraints apply to all levels of nested logit models. The examples of two-level and three-level nesting structures shown in Figure 8.2 and Figure 8.3 represent four of the twenty-eight different nested models (13 two-level and 12 three-level) that are feasible for a four-alternative case. The number of distinct nests increases rapidly with an increasing alternatives (see Table 8-3).

The large number of feasible nesting structures poses a substantial problem of determining which one best reflects the choice behavior of the population. The analyst's judgment can be used to substantially reduce this to a smaller number of realistic structures (based on our understanding of the competitive relationships); however, the analyst needs to be cautious in excluding structures since some apparently non-intuitive structures may have good fit statistics and their interpretation may provide useful insight into the choice behavior under study.

Number of Alternatives	Possible 2-Level Nesting Structures	Possible 3-Level Nesting Structures	Total Possible Nesting Structures All Levels ⁴²
3	3	0	3
4	13	12	25
5	50	125	235
6	201	1040	2711

Table 8-3 Number of Possible Nesting Structures

8.4 Statistical Testing of Nested Logit Structures

Adopting a nested logit model implies rejection of the MNL⁴³. We can use standard statistical tests of the hypothesis that the MNL model is the true model since the nested logit model is a generalization of the MNL model. In the case of multiple nests, the hypothesis that the MNL is the true model is equivalent to the hypothesis that all the logsum parameters are equal to one. We can use the likelihood ratio statistic with degrees of freedom equal to the number of logsum parameters (the number of restrictions between the NL and the MNL) to test this hypothesis. We reject the null hypothesis that the MNL model is the correct model if the calculated value is greater than the test or critical value for the distribution as:

$$-2 \times [\ell^{\text{MNL}} - \ell^{\text{NL}}] \geq \chi_n^2 \quad 8.31$$

where n is the number of restrictions (nests) between the MNL and NL models.

⁴² The maximum number of levels is one fewer than the number of alternatives

⁴³ Test can also be made between any NL and a simpler NL that is a reduced form of the initial model.

In the case of a simple NL model with a single nest, we can use the t-statistic to test the hypothesis that the logsum parameter is equal to one. Even in a more complex NL model, we can test each logsum term to determine if a portion of the nesting structure can be eliminated. For each case in which the hypothesis that the logsum parameter is equal to one is not rejected, the corresponding branch of the tree can be eliminated and the alternatives can go directly to the next level. The t-statistics must be evaluated for the appropriate null hypothesis. For nests directly under the root of the tree, the null hypothesis is $H_0 : \mu_k = 1$; for other nests, the null hypothesis is that the parameter for that nest is equal to the parameter for the next higher nest in the tree. That is, for a top level nest,

$$\text{t-statistic} = \frac{\hat{\theta}_k - 1}{S_k} \quad 8.32$$

where $\hat{\theta}_k$ is the estimate of the logsum parameter for nest k ,

1 is the hypothesized value against which the logsum parameter is being tested;

S_k is the standard error of the parameter estimate.

For other nests,

$$\text{t-statistic} = \frac{\hat{\theta}_{k \subset j} - \hat{\theta}_j}{\sqrt{S_{k \subset j}^2 + S_j^2 - 2S_{k \subset j, j}}} \quad 8.33$$

where $\hat{\theta}_{k \subset j}$ is the estimate of the logsum parameter for nest k that is included under nest j ,

$\hat{\theta}_j$ is the estimate of the logsum parameter for nest j ,

$S_{k \subset j}^2$ is the error variance of the logsum parameter for nest k that is included under nest j ,

S_j^2 is the error variance of the logsum parameter for nest j , and

$S_{k \subset j, j}$ is the error covariance of the two logsum parameters.

It is important to note that these are not necessarily the test values that will be reported by all computer programs, many of which apply the test for the null hypothesis that the logsum is equal to zero or one only. If the test against zero is reported⁴⁴, the user must obtain or calculate the standard errors of estimate for the relevant parameters and calculate the t-statistic(s) as defined above.

The likelihood ratio test can also be used to test the significance of more complex nested models or between models with different nesting structures as long as the nesting structure of one model can be obtained as a restriction of the other. To choose between two nested logit models where neither model is a restricted version of the other, we use the non-nested hypothesis test discussed in CHAPTER 5 (section 5.7.3.2).

⁴⁴ Also, the statistic must be calculated if the test against one is reported for logsums not directly below the root of the tree.

CHAPTER 9: Selecting a Preferred Nesting Structure

9.1 Introduction

This chapter re-examines the San Francisco mode choice models estimated in CHAPTER 6 and CHAPTER 7 and evaluates whether the MNL models should be replaced by nested logit models. The final, un-segmented, specifications, Model 17W for work trips and Model 14 S/O for shop/other trips, are taken as the base specifications for estimating nested logit models⁴⁵. Nested models for work trips are examined first, followed by shop/other trips. The chapter concludes with consideration of the policy implications of adopting alternative nesting structures.

Although the number of possible nests for six alternatives is large, the nature of the alternatives allows certain nests to be rejected as implausible. For example, it is unreasonable to suppose that Bike and Transit have substantial unobserved characteristics that lead to correlation in the error terms of their utility functions. For this reason, this chapter considers four plausible nests which are combined in different ways. For Work trips⁴⁶, these are:

- Motorized (M) alternatives, comprised of Drive Alone, Shared Ride 2, Shared Ride 3+, and Transit;
- Private automobile (P) alternatives, including Drive Alone, Shared Ride 2 and Shared Ride 3+ alternatives;
- Shared ride (S) alternatives; and
- Non-Motorized (NM) alternatives comprised of Bike and Walk.

Given these four nests or groupings of alternatives and recognizing that Motorized includes both Automobile and Transit alternatives and that Automobile includes Drive Alone

⁴⁵ Estimated parameters from the MNL model are typically used as initial values for utility function parameters and one is typically used for the initial values of logsum parameters of the nested logit model. However, as noted in this chapter and discussed more thoroughly in the next chapter it is sometimes desirable to start the estimation with other values for the logsum parameters.

⁴⁶ For Shop/Other trips, the shared ride alternatives include Shared Ride 2, Shared Ride 3+, Shared Ride 2+ & Drive Alone, and Shared Ride 2/3+ as defined in Section 7.1. All other nests are as described for Work trips.

and the Shared Ride alternatives, fifteen nest structures as described below. We begin with four single nest models (and, based on the results of these estimations, we explore a selection of more complex structures).

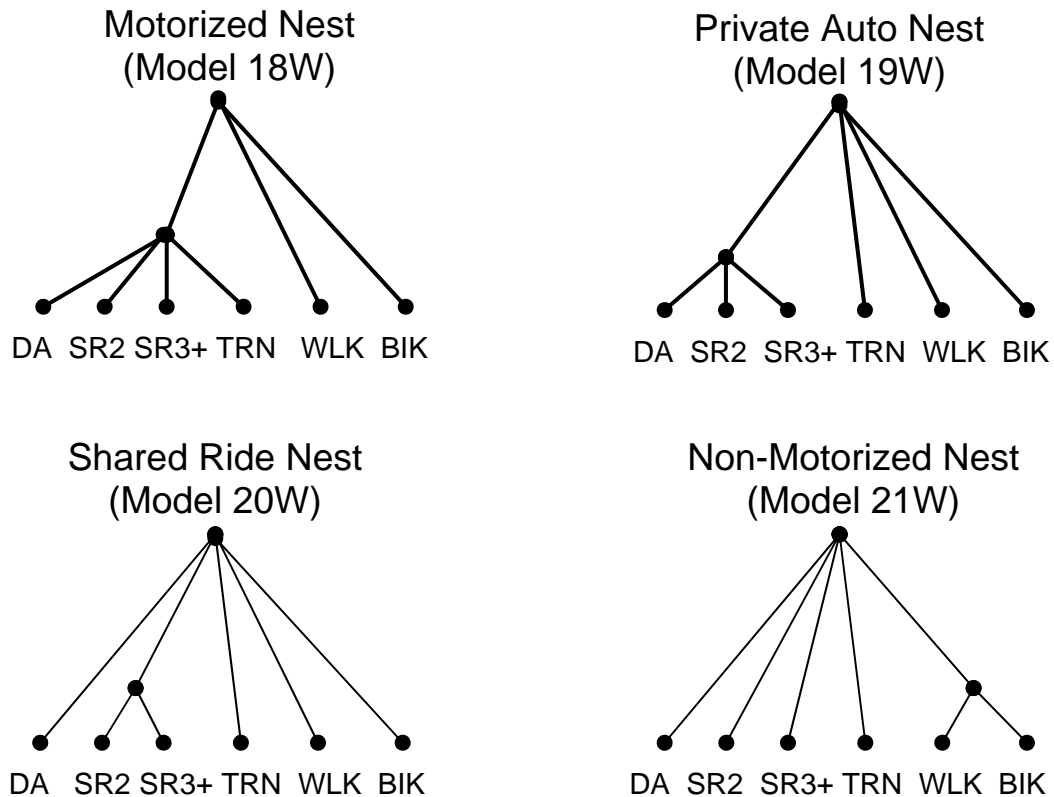


Figure 9.1 Single Nest Models

9.2 Nested Models for Work Trips

The estimation of nested logit models for work trips begins with consideration of the four single nest models mentioned above and illustrated in Figure 9.1. Each of these models obtains improved goodness-of-fit relative to the MNL model (see Table 9-1). Three of the four models; Models 18W, 20W, 21W; result in acceptable logsum parameters and two of these models (Models 18W and 20W) significantly reject the MNL model at close to the 0.05 level (see the chi-square test and significance in the last two rows of Table 9-1). However, the model with the private automobile nest (Model 19W) is rejected because the logsum parameter is greater than

one and the model with the non-motorized nest (Model 21W) cannot reject the MNL model at any reasonable level.

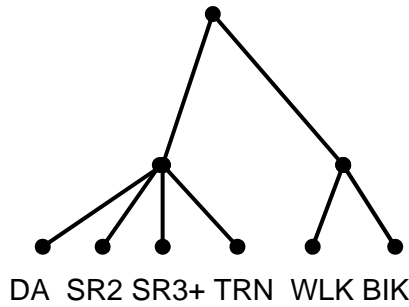
Table 9-1 Single Nest Work Trip Models

Variable	Model 17W	Model 18W	Model 19W	Model 20W	Model 21W
Nest	None	Motorized (M)	Private Auto (P)	Shared Ride (S)	Non-Motorized (NM)
Travel Cost by Income (1990 cents per 1000 1990 \$'s)	-0.0524 (-5.0)	-0.0388 (-5.1)	-0.0607 (-9.3)	-0.0455 (-4.6)	-0.0519 (-6.3)
Travel Time (minutes)					
Motorized Modes Only	-0.0202 (-5.3)	-0.0146 (-4.8)	-0.0202 (-5.6)	-0.0206 (-5.7)	-0.0199 (-5.2)
Non-Motorized Modes Only	-0.0454 (-7.9)	-0.0462 (-8.4)	-0.0462 (-8.2)	-0.0452 (-7.7)	-0.0454 (-8.6)
OVT by Distance (mi.)	-0.133 (-6.8)	-0.112 (-5.9)	-0.136 (-7.1)	-0.134 (-6.9)	-0.135 (-8.2)
Motorized Modes					
Household Income (1,000's of 1990 dollars)					
All Private Vehicle Modes	0.00	0.00	0.00	0.00	0.00
Transit	-0.0053 (-2.7)	-0.0039 (-3.5)	-0.0045 (-2.4)	-0.0054 (-2.8)	-0.0053 (-2.8)
Bike	-0.0086 (-1.7)	-0.0095 (-2.1)	-0.0075 (-1.7)	-0.0089 (-2.0)	-0.0092 (-2.8)
Walk	-0.0060 (-1.9)	-0.0066 (-2.3)	-0.0050 (-1.7)	-0.0062 (-2.1)	-0.0056 (-2.1)
Vehicles per Worker					
Drive Alone (base)					
Shared Ride (any)	-0.317 (-4.8)	-0.225 (-4.7)	-0.511 (-5.8)	-0.315 (-5.2)	-0.317 (-5.4)
Transit	-0.946 (-8.0)	-0.704 (-5.7)	-0.872 (-8.0)	-0.938 (-9.0)	-0.947 (-9.1)
Bike	-0.702 (-2.7)	-0.742 (-3.4)	-0.611 (-2.8)	-0.703 (-3.2)	-0.693 (-4.1)
Walk	-0.722 (-4.3)	-0.772 (-5.4)	-0.614 (-4.2)	-0.724 (-5.0)	-0.714 (-5.3)
Dummy Variable for Destination in CBD					
Drive Alone (base)	0.00	0.00	0.00	0.00	0.00
Shared Ride 2	0.260 (2.1)	0.192 (2.8)	0.398 (3.8)	0.396 (3.6)	0.260 (2.1)
Shared Ride 3+	1.07 (5.6)	0.778 (4.9)	1.59 (6.5)	0.641 (4.6)	1.07 (5.5)
Transit	1.31 (7.9)	0.918 (4.9)	1.37 (7.8)	1.32 (7.7)	1.31 (7.3)
Bike	0.489 (1.4)	0.478 (1.3)	0.524 (1.5)	0.501 (1.4)	0.414 (1.5)
Walk	0.102 (0.4)	0.113 (0.5)	0.117 (0.5)	0.114 (0.5)	0.104 (0.5)
Empl. Density - Work Zone (employees / square mile)					
Drive Alone (base)	0.00	0.00	0.00	0.00	0.00
Shared Ride 2	0.0016 (4.0)	0.0011 (4.3)	0.0025 (8.1)	0.0019 (5.5)	0.0016 (4.3)
Shared Ride 3+	0.0023 (5.0)	0.0016 (4.6)	0.0036 (7.7)	0.0019 (5.5)	0.0023 (5.0)
Transit	0.0031 (8.7)	0.0022 (5.3)	0.0035 (10.2)	0.0032 (9.0)	0.0031 (9.1)
Bike	0.0019 (1.6)	0.0014 (1.2)	0.0023 (1.9)	0.0020 (1.6)	0.0022 (2.4)
Walk	0.0029 (3.9)	0.0023 (3.0)	0.0033 (4.5)	0.0030 (3.8)	0.0028 (3.6)

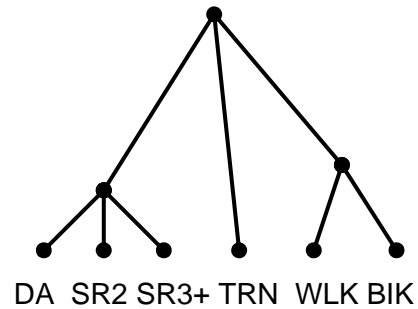
Variable	Model 17W	Model 18W	Model 19W	Model 20W	Model 21W
Nest	None	Motorized (M)	Private Auto (P)	Shared Ride (S)	Non-Motorized (NM)
Constants					
Drive Alone (base)	0.00	0.00	0.00	0.00	0.00
Shared Ride 2	-1.81 (-17.0)	-1.32 (-5.9)	-2.57 (-12.1)	-1.68 (-17.4)	-1.81 (-18.5)
Shared Ride 3+	-3.43 (-22.6)	-2.50 (-5.9)	-4.96 (-11.9)	-2.21 (-8.8)	-3.43 (-23.0)
Transit	-0.685 (-2.8)	-0.404 (-2.6)	-0.847 (-3.4)	-0.680 (-2.8)	-0.682 (-2.9)
Bike	-1.63 (-3.8)	-1.38 (-3.5)	-1.82 (-4.7)	-1.62 (-4.2)	-1.44 (-4.9)
Walk	0.0682 (0.2)	0.334 (0.9)	-0.116 (-0.3)	0.0732 (0.2)	0.0791 (0.2)
Nesting Coefficients / Dissimilarity Parameters					
Non-Motorized Nest					0.766 (-2.1)
Shared Ride Nest				0.329 (-5.4)	
Automobile Nest			1.47 (4.0)		
Motorized Nest		0.723 (-2.3)			
Log-likelihood at Zero	-7309.601	-7309.601	-7309.601	-7309.601	-7309.601
Log-likelihood at Constants	-4132.916	-4132.916	-4132.916	-4132.916	-4132.916
Log-likelihood at Convergence	-3444.185	-3442.315	-3435.996	-3442.415	-3443.554
Rho Squared w.r.t Zero	0.5288	0.5291	0.5299	0.5291	0.5289
Rho Squared w.r.t. Constants	0.1666	0.1671	0.1686	0.1671	0.1668
Chi-Squared vs. MNL		3.74	16.38	3.54	1.26
Rejection significance		.053	.000	.060	.261

Next we consider three models with the non-motorized nest in parallel with each of the three other nests (Figure 9.2). In each case, the new models in Table 9-2; Models 22W, 23W and 24W; have better goodness of fit than the corresponding single nest models. Model 23W is rejected because the nest parameter for private automobile is greater than one, as before. Models 22W and 23W both reject the single non-motorized nest (Model 21W) at close to the 0.05 level; however, neither rejects the corresponding Motorized and Shared Ride models (Models 18W and 20W, respectively). In such cases, the analyst can decide to include the non-motorized nest or not, for other than statistical reasons.

Motorized – Non-Motorized Nests
(Model 22W)



Private Auto – Non-Motorized Nests
(Model 23W)



Shared Ride – Non-Motorized Nest
(Model 24W)

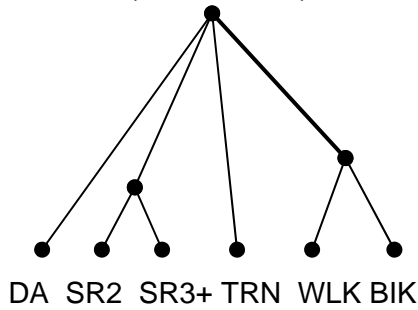


Figure 9.2 Non-Motorized Nest in Parallel with Motorized, Private Automobile and Shared Ride Nests

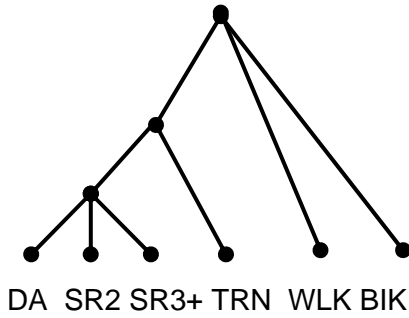
Table 9-2 Parallel Two Nest Work Trip Models

Variable	Model 17W		Model 22W		Model 23W		Model 24W	
Nest	None		M-NM		P-NM		S-NM	
Travel Cost by Income (1990 cents per 1000 1990 dollars)	-0.0524	(-5.0)	-0.0386	(-4.8)	-0.0600	(-9.0)	-0.0449	(-4.5)
Travel Time (minutes)								
Motorized Modes Only	-0.0204	(-5.3)	-0.0145	(-4.4)	-0.0199	(-5.5)	-0.0203	(-5.5)
Non-Motorized Modes Only	-0.0454	(-7.9)	-0.0462	(-9.3)	-0.0460	(-8.5)	-0.0452	(-8.6)
OVT by Distance (mi.) Motorized Modes	-0.133	(-6.8)	-0.114	(-6.0)	-0.138	(-7.3)	-0.136	(-8.1)
Household Income (1,000's of 1990 dollars)								
All Private Vehicle Modes	0.00		0.00		0.00		0.00	
Transit	-0.0053	(-2.7)	-0.0039	(-3.3)	-0.0046	(-2.4)	-0.0054	(-2.8)
Bike	-0.0086	(-1.7)	-0.0101	(-2.9)	-0.0081	(-2.4)	-0.0095	(-2.7)
Walk	-0.0060	(-1.9)	-0.0062	(-2.3)	-0.0046	(-1.7)	-0.0058	(-2.1)
Vehicles per Worker								
Drive Alone (base)								
Shared Ride (any)	-0.317	(-4.8)	-0.226	(-4.0)	-0.512	(-4.9)	-0.315	(-5.3)
Transit	-0.946	(-8.0)	-0.707	(-5.4)	-0.874	(-7.9)	-0.939	(-9.0)
Bike	-0.702	(-2.7)	-0.735	(-4.1)	-0.595	(-3.6)	-0.694	(-4.3)
Walk	-0.722	(-4.3)	-0.764	(-5.6)	-0.608	(-4.4)	-0.716	(-5.4)
Dummy Variable for Destination in CBD								
Drive Alone (base)	0.00		0.00		0.00		0.00	
Shared Ride 2	0.260	(2.1)	0.193	(2.8)	0.399	(3.5)	0.398	(3.6)
Shared Ride 3+	1.07	(5.6)	0.781	(4.6)	1.59	(6.4)	0.639	(4.6)
Transit	1.31	(7.9)	0.921	(4.8)	1.37	(7.8)	1.32	(7.5)
Bike	0.489	(1.4)	0.407	(1.5)	0.445	(1.6)	0.425	(1.5)
Walk	0.102	(0.4)	0.114	(0.5)	0.120	(0.5)	0.117	(0.5)
Empl. Density - Work Zone (employees / square mile)								
Drive Alone (base)	0.00		0.00		0.00		0.00	
Shared Ride 2	0.0016	(4.0)	0.0012	(4.1)	0.0025	(7.6)	0.0019	(5.6)
Shared Ride 3+	0.0023	(5.0)	0.0016	(4.5)	0.0036	(7.9)	0.0019	(5.6)
Transit	0.0031	(8.7)	0.0022	(5.2)	0.0035	(10.1)	0.0032	(9.1)
Bike	0.0019	(1.6)	0.0017	(1.8)	0.0026	(2.9)	0.0023	(2.3)
Walk	0.0029	(3.9)	0.0022	(2.9)	0.0032	(4.5)	0.0029	(3.9)
Constants								
Drive Alone (base)	0.00		0.00		0.00		0.00	
Shared Ride 2	-1.81	(-17.0)	-1.33	(-5.7)	-2.57	(-12.2)	-1.68	(-17.3)
Shared Ride 3+	-3.43	(-22.6)	-2.51	(-5.6)	-4.96	(-12.2)	-2.20	(-8.8)
Transit	-0.685	(-2.8)	-0.404	(-2.6)	-0.843	(-3.4)	-0.677	(-2.9)
Bike	-1.63	(-3.8)	-1.20	(-3.5)	-1.65	(-5.0)	-1.43	(-4.3)
Walk	0.0682	(0.2)	0.345	(1.0)	-0.110	(-0.3)	0.0842	(0.3)

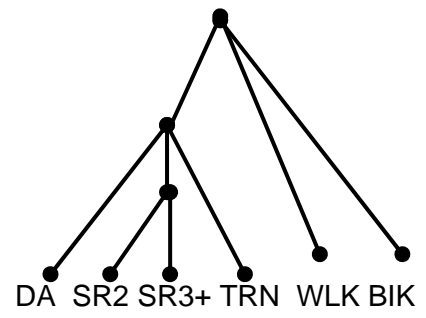
Variable	Model 17W	Model 22W	Model 23W	Model 24W
Nest	None	M-NM	P-NM	S-NM
Nesting Coefficients / Dissimilarity Parameters				
Non-Motorized Nest		0.769 (-1.5)	0.765 (-1.4)	0.762 (-1.3)
Shared Ride Nest				0.324 (-5.4)
Automobile Nest			1.46 (4.0)	
Motorized Nest		0.726 (-2.1)		
Log-likelihood at Zero	-7309.601	-7309.601	-7309.601	-7309.601
Log-likelihood at Constants	-4132.916	-4132.916	-4132.916	-4132.916
Log-likelihood at Convergence	-3444.185	-3441.673	-3435.358	-3441.761
Rho Squared w.r.t Zero	0.5288	0.5292	0.5300	0.5291
Rho Squared w.r.t. Constants	0.1666	0.1672	0.1688	0.1672
Chi-Squared vs. MNL		5.02	17.65	4.85
Rejection Significance		0.081	0.000	0.089
Chi-Squared vs. M, P and S nests		1.28	1.28	1.31
Rejection Significance		0.257	0.259	0.253
Chi-Sqr vs. NM nest (Model21W)		3.76	16.39	3.59
Rejection Significance		0.052	0.000	0.058

Another option is to consider ‘hierarchical’ two nest models with Motorized and Private Automobile, Motorized and Shared Ride or Private Automobile and Shared Ride as illustrated in Figure 9.3. Of the three models reported in Table 9-3, only Model 26W, Motorized and Shared Ride, produces an acceptable result. Model 27W is rejected because the private automobile nest coefficient is greater than one, as before. Model 25W is also rejected, despite the fact that the automobile nest parameter is less than one because it is greater than the logsum parameter for the motorized nest which is above it (see Section 8.4). Model 26W, however, represents an attractive model achieving the best goodness-of-fit of any of the models considered, with acceptable logsum parameters. Further, it rejects the MNL model (Model 17W), the Motorized (Model 18W) and the Shared Ride (Model 20W) nest models at roughly the 0.03, 0.07 and 0.06 levels, respectively.

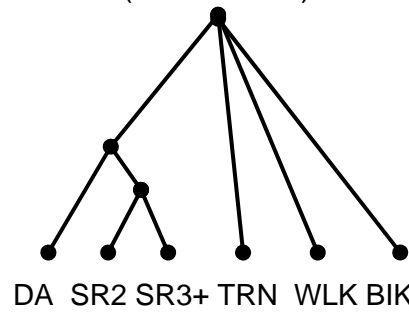
Motorized – Shared Ride Nest
(Model 25W)



Motorized - Private Auto Nest
(Model 26W)



Private Auto - Shared Ride Nest
(Model 27W)



Motorized - Private Auto - Shared Ride Nest
(Model 29W)

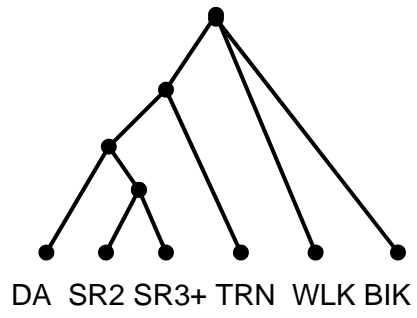


Figure 9.3 Hierarchically Nested Models

Table 9-3 Hierarchical Two Nest Work Trip Models

Variable	Model 17W		Model 25W		Model 26W		Model 27W	
Nest	None		M-P		M-S		P-S	
Travel Cost by Income (1990 cents per 1000 1990 dollars)	-0.0524	(-5.0)	-0.0363	(-5.2)	-0.0336	(-3.3)	-0.0526	(-7.2)
Travel Time (minutes)								
Motorized Modes Only	-0.0202	(-5.3)	-0.0106	(-4.6)	-0.0149	(-4.4)	-0.0209	(-5.9)
Non-Motorized Modes Only	-0.0454	(-7.9)	-0.0471	(-8.7)	-0.0461	(-8.3)	-0.046	(-8.0)
OVT by Distance (mi.) Motorized Modes	-0.133	(-6.8)	-0.0995	(-5.7)	-0.113	(-6.6)	-0.137	(-8.2)
Household Income (1,000's of 1990 dollars)								
All Private Vehicle Modes	0.00		0.00		0.00		0.00	
Transit	-0.0053	(-2.7)	-0.0022	(-2.2)	-0.0040	(-3.0)	-0.0046	(-2.4)
Bike	-0.0086	(-1.7)	-0.0091	(-2.0)	-0.00970	(-2.2)	-0.0077	(-1.7)
Walk	-0.0060	(-1.9)	-0.0061	(-2.1)	-0.0068	(-2.3)	-0.0053	(-1.8)
Vehicles per Worker								
Drive Alone (base)								
Shared Ride (any)	-0.317	(-4.8)	-0.321	(-5.2)	-0.224	(-4.2)	-0.515	(-4.9)
Transit	-0.946	(-8.0)	-0.463	(-5.2)	-0.700	(-5.1)	-0.86	(-7.9)
Bike	-0.702	(-2.7)	-0.688	(-3.1)	-0.742	(-3.4)	-0.611	(-2.8)
Walk	-0.722	(-4.3)	-0.705	(-4.9)	-0.774	(-5.4)	-0.616	(-4.2)
Dummy Variable for Destination in CBD								
Drive Alone (base)	0.00		0.00		0.00		0.00	
Shared Ride 2	0.260	(2.1)	0.275	(3.5)	0.291	(3.5)	0.62	(4.1)
Shared Ride 3+	1.07	(5.6)	1.02	(5.5)	0.471	(3.7)	0.889	(5.6)
Transit	1.31	(7.9)	0.731	(5.1)	0.927	(4.6)	1.38	(7.7)
Bike	0.489	(1.4)	0.500	(1.4)	0.486	(1.4)	0.539	(1.5)
Walk	0.102	(0.4)	0.142	(0.6)	0.122	(0.5)	0.133	(0.5)
Empl. Density - Work Zone (employees / square mile)								
Drive Alone (base)	0.00		0.00		0.00		0.00	
Shared Ride 2	0.0016	(4.0)	0.0016	(5.0)	0.0014	(4.4)	0.0030	(6.3)
Shared Ride 3+	0.0023	(5.0)	0.0024	(5.3)	0.0014	(4.4)	0.0030	(6.4)
Transit	0.0031	(8.7)	0.0019	(5.7)	0.0023	(4.9)	0.0036	(9.5)
Bike	0.0019	(1.6)	0.0014	(1.2)	0.0015	(1.2)	0.0024	(2.0)
Walk	0.0029	(3.9)	0.0022	(3.0)	0.0023	(3.0)	0.0034	(4.5)
Constants								
Drive Alone (base)	0.00		0.00		0.00		0.00	
Shared Ride 2	-1.81	(-17.0)	-1.63	(-6.2)	-1.23	(-5.4)	-2.38	(-12.7)
Shared Ride 3+	-3.43	(-22.6)	-3.15	(-6.3)	-1.62	(-4.5)	-2.96	(-14.6)
Transit	-0.685	(-2.8)	-0.407	(-3.0)	-0.401	(-2.3)	-0.837	(-3.6)
Bike	-1.63	(-3.8)	-1.38	(-3.5)	-1.38	(-3.5)	-1.81	(-4.7)
Walk	0.0682	(0.2)	0.339	(1.0)	0.337	(0.9)	-0.113	(-0.3)

Variable	Model 17W	Model 25W	Model 26W	Model 27W
Nest	None	M-P	M-S	P-S
Nesting Coefficients / Dissimilarity Parameters				
Non-Motorized Nest			0.242 (-6.6)	0.364 (-31.0)
Shared Ride Nest		0.923 (-0.5)		1.48 (4.2)
Automobile Nest		0.532 (-5.4)	0.725 (-2.0)	
Motorized Nest				
Log-likelihood at Zero	-7309.601	-7309.601	-7309.601	-7309.601
Log-likelihood at Constants	-4132.916	-4132.916	-4132.916	-4132.916
Log-likelihood at Convergence	-3444.185	-3427.166	-3440.601	-3433.910
Rho Squared w.r.t Zero	0.5288	0.5311	0.5293	0.5302
Rho Squared w.r.t. Constants	0.1666	0.1708	0.1675	0.1691
Chi-Squared vs. MNL		34.04	7.17	20.55
Rejection Significance		.000	.028	.000
Upper Nest		Motorized	Motorized	Private Auto
Chi-Squared vs. Upper Nest		30.30	3.43	4.17
Rejection Significance		.000	0.064	0.041
Lower Nest		Private Auto	Shared Ride	Shared Ride
Chi-Squared vs. Lower Nest		17.66	3.63	17.01
Rejection Significance		.000	0.057	.000

Table 9-4 presents four additional nested model structures for work trips. The first two models extend the nesting structures previously estimated. Model 28W adds the non-motorized nest to Model 26W (see Figure 9.4) with further improved goodness-of-fit, but cannot reject Model 26W at any reasonable level of significance. Model 29W which includes the Motorized, Private Automobile and Shared Ride Nests (see Figure 9.3) is infeasible as it obtains a logsum for the automobile nest that is greater than for the motorized nest, as before. However, Model 29W presents a structure that represents our expectation of the relationship among the motorized modes. Thus, we estimate Model 30W which is identical to Model 29W except that it constrains the automobile nest parameter to 0.75 times the value of the motorized nest parameter. While Model 30W results in a poorer goodness-of-fit than any of the other models including the simple MNL model; it is worth considering because it reflects increased substitution as we move from the Motorized Nest to the Private Automobile Nest to the Shared Ride Nest, as expected. The data does not support this structure (relative to the MNL) but if the analyst or policy makers are

convinced it is correct, the use of the relational constraint can produce a model consistent with these assumptions.

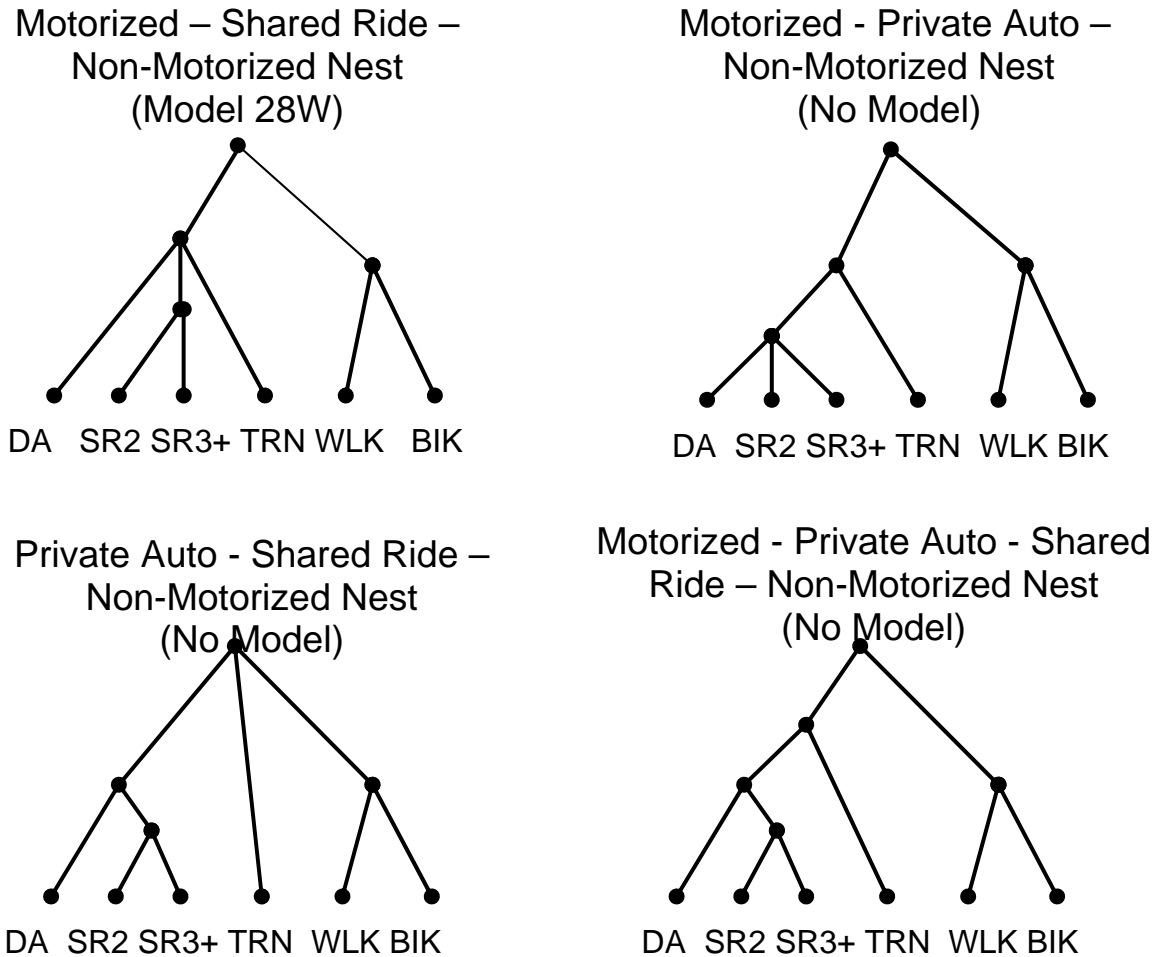


Figure 9.4 Complex Nested Models

The final model, Model 31W, extends Model 30W by adding the Non-Motorized nest to the structure. It therefore has the same issues regarding the constrained nest parameters, but it offers the advantage of including the substitution effects associated with all four nests previously selected. Again, the decision on the acceptance of this model may be based primarily on the judgment of analysts and policy makers.

Table 9-4 Complex and Constrained Nested Models for Work Trips

Variable	Model 28W		Model 29W		Model 30W		Model 31W	
Nest	M-S-NM		M-P-S		M-P-S Constrained		M-P-S-NM Constrained	
Travel Cost by Income (1990 cents per 1000 1990 dollars)	-0.0334	(-3.5)	-0.0317	(-4.4)	-0.0307	(-3.4)	-0.0305	(-3.4)
Travel Time (minutes)								
Motorized Modes Only	-0.0148	(-4.5)	-0.0111	(-4.8)	-0.0164	(-3.9)	-0.0163	(-3.8)
Non-Motorized Modes Only	-0.0460	(-9.4)	-0.0469	(-8.5)	-0.0456	(-8.2)	-0.0456	(-8.9)
OVT by Distance (mi.) Motorized Modes	-0.115	(-6.5)	-0.100	(-5.8)	-0.118	(-5.4)	-0.120	(-5.4)
Household Income (1,000's of 1990 dollars)								
All Private Vehicle Modes	0.00		0.00		0.00		0.00	
Transit	-0.0040	(-3.4)	-0.0022	(-2.2)	-0.0048	(-3.1)	-0.0049	(-2.7)
Bike	-0.0102	(-3.0)	-0.0092	(-2.0)	-0.0102	(-2.3)	-0.0107	(-3.0)
Walk	-0.0064	(-2.4)	-0.0063	(-2.1)	-0.0072	(-2.5)	-0.0068	(-2.5)
Vehicles per Worker								
Drive Alone (base)								
Shared Ride (any)	-0.225	(-4.3)	-0.323	(-4.8)	-0.187	(-3.8)	-0.187	(-3.3)
Transit	-0.703	(-5.3)	-0.457	(-5.2)	-0.815	(-4.6)	-0.817	(-4.7)
Bike	-0.735	(-4.2)	-0.688	(-3.1)	-0.787	(-3.6)	-0.785	(-4.5)
Walk	-0.765	(-5.6)	-0.706	(-4.7)	-0.830	(-5.8)	-0.821	(-5.9)
Dummy Variable for Destination in CBD								
Drive Alone (base)	0.00		0.00		0.00		0.00	
Shared Ride 2	0.293	(3.5)	0.416	(4.3)	0.230	(3.1)	0.231	(2.8)
Shared Ride 3+	0.472	(3.9)	0.578	(5.0)	0.404	(3.2)	0.404	(3.4)
Transit	0.930	(4.6)	0.737	(5.2)	1.02	(3.9)	1.02	(3.7)
Bike	0.416	(1.5)	0.507	(1.4)	0.485	(1.4)	0.414	(1.5)
Walk	0.123	(0.6)	0.150	(0.6)	0.121	(0.5)	0.122	(0.5)
Empl. Density - Work Zone (employees / square mile)								
Drive Alone (base)	0.00		0.00		0.00		0.00	
Shared Ride 2	0.0014	(4.5)	0.0019	(5.4)	0.0011	(3.7)	0.0011	(3.5)
Shared Ride 3+	0.0014	(4.5)	0.0020	(5.4)	0.0011	(3.7)	0.0011	(3.5)
Transit	0.0023	(5.0)	0.0020	(5.5)	0.0024	(4.3)	0.0024	(4.1)
Bike	0.0018	(2.0)	0.0015	(1.2)	0.0015	(1.2)	0.0018	(2.0)
Walk	0.0022	(3.0)	0.0022	(2.9)	0.0024	(3.0)	0.0023	(2.9)
Constants								
Drive Alone (base)	0.00		0.00		0.00		0.00	
Shared Ride 2	-1.24	(-5.5)	-1.51	(-6.1)	-1.00	(-4.5)	-1.01	(-4.6)
Shared Ride 3+	-1.62	(-4.9)	-1.86	(-5.7)	-1.37	(-3.8)	-1.37	(-4.3)
Transit	-0.400	(-2.3)	-0.397	(-3.0)	-0.377	(-1.8)	-0.376	(-1.5)
Bike	-1.19	(-3.6)	-1.38	(-3.5)	-1.33	(-3.3)	-1.14	(-3.5)
Walk	0.347	(1.0)	0.339	(0.9)	0.390	(1.1)	0.402	(1.1)

Variable	Model 28W	Model 29W	Model 30W	Model 31W
Nest	M-S-NM	M-P-S	M-P-S Constrained	M-P-S-NM Constrained
Nesting Coefficients / Dissimilarity Parameters				
Non-Motorized Nest	0.767 (-1.4)			0.767 (-1.3)
Shared Ride Nest	0.240 (-8.2)	0.217 (-11.0)	0.233 (-7.1)	0.231 (-8.4)
Automobile Nest		0.928 (-0.5)	0.598 (-3.1)	0.600 (-4.8)
Motorized Nest	0.728 (-2.1)	0.533 (-5.4)	0.798 " ⁴⁷	0.800 "
Log-likelihood at Zero	-7309.601	-7309.601	-7309.601	-7309.601
Log-likelihood at Constants	-4132.916	-4132.916	-4132.916	-4132.916
Log-likelihood at Convergence	-3439.943	-3425.159	-3453.824	-3453.179
Rho Squared w.r.t Zero	0.5294	0.5314	0.5275	0.5276
Rho Squared w.r.t. Constants	0.1677	0.1712	0.1643	0.1645
Chi-Squared vs. MNL	8.48	N/A	N/A	N/A
Rejection significance	.037	N/A	N/A	N/A
Chi-Squared vs. single NM nest	7.22	N/A	N/A	N/A
Rejection significance	.027	N/A	N/A	N/A
Chi-Squared vs. single S nest	4.94	N/A	N/A	N/A
Rejection significance	.084	N/A	N/A	N/A
Chi-Squared vs. single P nest	N/A	N/A	N/A	N/A
Rejection significance	N/A	N/A	N/A	N/A
Chi-Squared vs. single M nest	4.74	N/A	N/A	N/A
Rejection significance	.093	N/A	N/A	N/A
Chi-Squared vs. Model 22W	3.46	N/A	N/A	N/A
Rejection significance	.063	N/A	N/A	N/A
Chi-Squared vs. Model 24W	3.64	N/A	N/A	N/A
Rejection significance	.057	N/A	N/A	N/A
Chi-Squared vs. Model 26W	1.32	N/A	N/A	N/A
Rejection significance	.251	N/A	N/A	N/A

Based on these results, Models 26W, 28W, 30W and 31W are all potential candidates for a final model. Model 26W is not rejected by any other model and is the simplest structure of this group. Model 28W is slightly better than Model 26W but enough to statistically reject it. Models 30W and 31W have a poorer goodness of fit than Model 26W and 28W, respectively but they incorporate the private automobile intermediate nest. The decision about which of these models to use is largely a matter of judgment. It is possible that other models might also be considered.

⁴⁷ The quote (ditto) mark indicates that the t-statistic is identical to the one above due to the imposition of a constraint on the ratio of the corresponding parameters.

We further examine Model 26W, because of its simplicity, to demonstrate the differences in direct-elasticity and cross-elasticity between the MNL (17W) model and pairs of alternatives in different parts of the NL tree associated with 26W.

Table 9-5 MNL (17W) vs. NL Model 26W

Variable	Model 17W		Model 26W	
Nest	None		M-S	
Travel Cost by Income (1990 cents per 1000 1990 dollars)	-0.0524	(-5.0)	-0.0336	(-3.3)
Travel Time (minutes)				
Motorized Modes Only	-0.0202	(-5.3)	-0.0149	(-4.4)
Non-Motorized Modes Only	-0.0454	(-7.9)	-0.0461	(-8.3)
OVT by Distance (mi.) Motorized Modes	-0.133	(-6.8)	-0.113	(-6.6)
Household Income (1,000's of 1990 dollars)				
All Private Vehicle Modes	0.00		0.00	
Transit	-0.0053	(-2.7)	-0.0040	(-3.0)
Bike	-0.0086	(-1.7)	-0.0970	(-2.2)
Walk	-0.0060	(-1.9)	-0.0068	(-2.3)
Vehicles per Worker				
Drive Alone (base)				
Shared Ride (any)	-0.317	(-4.8)	-0.224	(-4.2)
Transit	-0.946	(-8.0)	-0.700	(-5.1)
Bike	-0.702	(-2.7)	-0.742	(-3.4)
Walk	-0.722	(-4.3)	-0.774	(-5.4)
Dummy Variable for Destination in CBD				
Drive Alone (base)	0.00		0.00	
Shared Ride 2	0.260	(2.1)	0.291	(3.5)
Shared Ride 3+	1.07	(5.6)	0.471	(3.7)
Transit	1.31	(7.9)	0.927	(4.6)
Bike	0.489	(1.4)	0.486	(1.4)
Walk	0.102	(0.4)	0.122	(0.5)
Empl. Density - Work Zone (employees / square mile)				
Drive Alone (base)	0.00		0.00	
Shared Ride 2	0.0016	(4.0)	0.0014	(4.4)
Shared Ride 3+	0.0023	(5.0)	0.0014	(4.4)
Transit	0.0031	(8.7)	0.0023	(4.9)
Bike	0.0019	(1.6)	0.0015	(1.2)
Walk	0.0029	(3.9)	0.0023	(3.0)

Variable	Model 17W	Model 26W
Nest	None	M-S
Constants		
Drive Alone (base)	0.00	0.00
Shared Ride 2	-1.81 (-17.0)	-1.23 (-5.4)
Shared Ride 3+	-3.43 (-22.6)	-1.62 (-4.5)
Transit	-0.685 (-2.8)	-0.401 (-2.3)
Bike	-1.63 (-3.8)	-1.38 (-3.5)
Walk	0.0682 (0.2)	0.337 (0.9)
Nesting Coefficients / Dissimilarity Parameters		
Non-Motorized Nest		
Shared Ride Nest		0.242 (-6.6)
Automobile Nest		
Motorized Nest		0.725 (-2.0)
Log-likelihood at Zero	-7309.601	-7309.601
Log-likelihood at Constants	-4132.916	-4132.916
Log-likelihood at Convergence	-3444.185	-3440.601
Rho Squared w.r.t Zero	0.5288	0.5293
Rho Squared w.r.t. Constants	0.1666	0.1675
Chi-Squared vs. MNL		7.17
Rejection Significance		.028

The structure of the nested logit model (26W) in Figure 9.3 is reproduced here so that each nest can be readily identified. The following table reports the elasticities for the MNL Model 17W and the NL Model 26W.

Figure 9.5 Motorized – Shared Ride Nest (Model 26W)

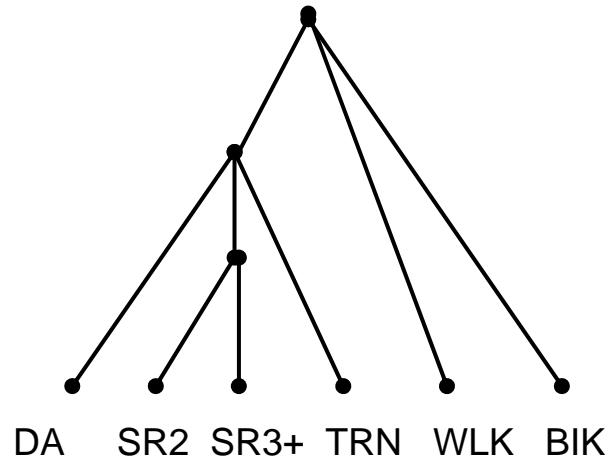


Figure 9.6 Elasticities for MNL (17W) and NL Model (26W)(All Values Multiplied by IVT_j)

Elasticity of Mode Probability	MNL Model (17W)	Nested Model (26W)
Direct Elasticity	$-(1 - P_j) \times 0.0202$	SR Nest $-\left[(1 - P_j) + 3.13 \times (1 - P_{j SR-Nest})\right] \times 0.0149 > -0.0615(1 - P_j)$ Mot. Nest $-\left[(1 - P_j) + 1.14 \times (1 - P_{j Mot-Nest})\right] \times 0.0149 > -0.319(1 - P_j)$ All Other $-0.0149(1 - P_j)$
Cross Elasticity	$-P_j \times 0.0202$	SR Nest $-\left(P_j + 3.13(P_{j SR-Nest})\right) \times 0.0149 > -0.0616P_j$ Mot. Nest $-\left(P_j + 1.14(P_{j SR-Nest})\right) \times 0.0149 > -0.319(1 - P_j)$ All Other $-0.0149P_j$

It is apparent from the above table that reduction in the magnitude of the utility parameter for the NL model results in a lower direct and cross elasticity for alternatives that are in neither of the nests depicted in Figure 9.5 than they would have in the corresponding MNL model while alternatives in the lowest and intermediate nests have increased elasticity. The magnitude in of the elasticity increases as alternatives or pairs of alternatives are in lower nests in the tree. Possibly, a better way to think about this is that adoption of the MNL model results in some type of averaged elasticities rather than the distinct elasticities for alternatives in a properly formed NL model.

9.3 Nested Models for Shop/Other Trips

The exploration of nested logit models for shop/other trips follows the same approach as used with work trips, beginning with the consideration of four single nest models depicted in Figure 9.4. For these trips, all four of the single nest structures produce valid models (Table 9-6). Further, Models 15 S/O, 16 S/O and 17 S/O strongly reject the MNL model; however, the non-motorized nest (Model 18 S/O) does not reject the MNL model at any reasonable level of significance.

Table 9-6 Single Nest Shop/Other Trip Models

Variables	Model 14 S/O	Model 15 S/O	Model 16 S/O	Model 17 S/O	Model 18 S/O
Nest	None	Motorized (M)	Private Auto (P)	Shared Ride (S)	Non-motorized (NM)
Constants					
Transit	0.530 (0.7)	-0.139 (-0.3)	-0.206 (-0.3)	0.188 (0.3)	0.561 (0.8)
Shared Ride 2	-1.19 (-9.0)	-0.612 (-3.2)	-0.287 (-1.9)	-0.743 (-5.1)	-1.19 (-8.9)
Shared Ride 3+	-3.20 (-13.7)	-1.63 (-3.4)	-0.764 (-2.0)	-1.45 (-4.0)	-3.19 (-14.0)
Shared Ride 2/3 & Drive Alone	-1.78 (-11.4)	-0.909 (-3.4)	-0.424 (-2.0)	-0.929 (-4.7)	-1.78 (-12.4)
Shared Ride 2/3	-4.11 (-11.7)	-2.09 (-3.3)	-0.970 (-2.0)	-1.70 (-3.7)	-4.11 (-11.5)
Bike	-4.41 (-7.3)	-4.22 (-7.5)	-4.57 (-7.8)	-4.30 (-7.7)	-3.56 (-5.8)
Walk	0.461 (1.2)	0.774 (2.1)	0.242 (0.6)	0.544 (1.5)	0.356 (0.9)
Drive Alone (base)	0.00	0.00	0.00	0.00	0.00

Variables	Model 14 S/O	Model 15 S/O	Model 16 S/O	Model 17 S/O	Model 18 S/O
Nest	None	Motorized (M)	Private Auto (P)	Shared Ride (S)	Non-motorized (NM)
Log of Persons per Household					
Transit	2.10 (5.4)	1.18 (3.2)	0.908 (2.2)	1.96 (5.4)	2.10 (5.7)
Shared Ride 2	1.14 (9.9)	0.582 (3.2)	0.272 (2.0)	1.57 (12.7)	1.14 (9.6)
Shared Ride 3+	3.19 (18.3)	1.63 (3.5)	0.764 (2.1)	2.26 (11.1)	3.18 (19.2)
Shared Ride 2/3 & Drive Alone	1.90 (14.5)	0.973 (3.4)	0.453 (2.1)	1.82 (16.8)	1.90 (15.6)
Shared Ride 2/3	2.89 (11.3)	1.48 (3.4)	0.691 (2.1)	2.16 (11.7)	2.89 (10.9)
Bike	2.27 (4.8)	1.65 (3.4)	1.32 (3.1)	2.20 (4.8)	1.94 (5.1)
Walk	1.32 (6.2)	0.741 (2.9)	0.428 (1.9)	1.27 (6.4)	1.37 (7.4)
Drive Alone (base)	0.00	0.00	0.00	0.00	0.00
Number of Vehicles					
Transit	-2.06 (-6.6)	-0.796 (-2.2)	-1.68 (-5.7)	-2.02 (-7.6)	-2.06 (-6.9)
Shared Ride 2	-0.248 (-4.7)	-0.124 (-2.8)	-0.0573 (-1.1)	-0.375 (-7.1)	-0.248 (-4.8)
Shared Ride 3+	-0.819 (-10.3)	-0.419 (-3.3)	-0.195 (-2.0)	-0.556 (-9.1)	-0.820 (-10.2)
Shared Ride 2/3 & Drive Alone	-0.456 (-7.5)	-0.231 (-3.2)	-0.107 (-2.0)	-0.438 (-9.2)	-0.456 (-8.0)
Shared Ride 2/3	-0.700 (-6.0)	-0.357 (-3.1)	-0.166 (-1.9)	-0.518 (-8.3)	-0.700 (-6.0)
Bike	-0.401 (-2.0)	-0.282 (-1.5)	-0.161 (-0.7)	-0.387 (-2.1)	-0.438 (-2.6)
Walk	-0.714 (-5.5)	-0.639 (-5.5)	-0.495 (-3.8)	-0.705 (-5.8)	-0.703 (-6.0)
Drive Alone (base)	0.00	0.00	0.00	0.00	0.00
Travel Time (minutes)					
Non-Motorized Modes Only	-0.0848 (-8.0)	-0.0895 (-9.1)	-0.0833 (-8.8)	-0.0839 (-9.3)	-0.0817 (-8.6)
Motorized Modes Only	-0.0344 (-3.1)	-0.0188 (-2.3)	-0.0252 (-2.3)	-0.0257 (-2.6)	-0.0338 (-3.4)
OVT by Distance (mi.) M. Modes	-0.208 (-3.9)	-0.115 (-2.4)	-0.201 (-3.9)	-0.201 (-4.5)	-0.215 (-4.1)
Travel Cost (1990 cents)					
Travel Cost by Log of Income (1990 cents per log of 1000 1990 \$'s)					
	-0.0111 (-3.6)	-0.0059 (-2.6)	-0.0028 (-1.3)	-0.0070 (-2.4)	-0.0110 (-4.1)
Zero Vehicle Household D.V.					
Transit, Bike, Walk	1.57 (4.2)	1.33 (3.7)	1.97 (5.3)	1.53 (4.5)	1.57 (4.5)
All Private Vehicle Modes	0.00	0.00	0.00	0.00	0.00
D.V. for Destination in Core					
Transit	2.01 (3.2)	1.10 (2.5)	1.78 (2.9)	2.19 (3.8)	1.99 (3.5)
Shared Ride (any)	0.716 (1.8)	0.368 (1.7)	0.157 (0.4)	0.793 (2.1)	0.717 (1.9)
Bike, Walk	0.301 (0.2)	0.134 (0.2)	0.128 (0.7)	0.488 (0.4)	0.282 (0.6)
Drive Alone (base)	0.00	0.00	0.00	0.00	0.00
Nesting Coefficients / Dissimilarity Parameters					
Non-Motorized Nest					0.720 (-1.8)
Shared Ride Nest				0.308 (-6.3)	
Private Automobile Nest			0.231 (-7.1)		
Motorized Nest		0.501 (-3.4)			
Log-likelihood at Zero	-6201.516	-6201.516	-6201.516	-6201.516	-6201.516
Log-likelihood at Constants	-4962.194	-4962.194	-4962.194	-4962.194	-4962.194
Log-likelihood at Convergence	-4457.235	-4453.332	-4448.477	-4450.545	-4456.373
Rho Squared w.r.t Zero	0.2813	0.2819	0.2827	0.2823	0.2814
Rho Squared w.r.t. Constants	0.1018	0.1025	0.1035	0.1031	0.1019
Chi-Squared vs. MNL		7.8060	17.5160	13.3800	1.7240
Rejection significance		0.005	0.000	0.000	0.189

Despite the marginal significance of the non-motorized nest, the possible two parallel nest models including non-motorized and one other nest, were estimated and are reported in Table 9-7. All three specifications result in valid models with improved goodness-of-fit over the MNL and the single nest NM model (Model 18 S/O) but Models 19 S/O, 20 S/O and 21 S/O do not reject the corresponding Motorized (Model 15 S/O), Private Automobile (Model 16 S/O) and Shared Ride (Model 17 S/O) nest models, which are restricted forms of these models, at a high level of confidence.

It is important to note that a variety of different initial parameters were required to find the solution for Model 20 S/O reported here. The selection of different starting values for the model parameters, especially the nest parameter(s) may lead to convergence at different values in or out of the acceptable range as discussed in CHAPTER 10.

Table 9-7 Parallel Two Nest Models for Shop/Other Trips

Variables	Model 14 S/O		Model 19 S/O		Model 20 S/O		Model 21 S/O	
Nests	None		M-NM		P-NM		S-NM	
Constants								
Transit	0.530	(0.7)	-0.115	(-0.3)	-0.17	(-0.2)	0.228	(0.3)
Shared Ride 2	-1.19	(-9.0)	-0.623	(-3.8)	-0.282	(-2.4)	-0.739	(-5.3)
Shared Ride 3+	-3.20	(-13.7)	-1.66	(-3.9)	-0.75	(-2.4)	-1.43	(-4.0)
Shared Ride 2/3 & Drive Alone	-1.78	(-11.4)	-0.925	(-3.9)	-0.416	(-2.4)	-0.922	(-4.7)
Shared Ride 2/3	-4.11	(-11.7)	-2.12	(-3.9)	-0.952	(-2.3)	-1.69	(-3.6)
Bike	-4.41	(-7.3)	-3.44	(-5.2)	-3.69	(-6.0)	-3.41	(-5.7)
Walk	0.461	(1.2)	0.679	(1.9)	0.129	(0.3)	0.433	(1.2)
Drive Alone (base)	0.00		0.00		0.0		0.00	
Log of Persons per Household								
Transit	2.10	(5.4)	1.20	(3.6)	0.899	(2.2)	1.95	(5.3)
Shared Ride 2	1.14	(9.9)	0.592	(3.8)	0.266	(2.4)	1.57	(12.7)
Shared Ride 3+	3.19	(18.3)	1.66	(4.0)	0.75	(2.4)	2.26	(10.8)
Shared Ride 2/3 & Drive Alone	1.90	(14.5)	0.990	(3.9)	0.445	(2.4)	1.82	(18.1)
Shared Ride 2/3	2.89	(11.3)	1.51	(3.9)	0.678	(2.3)	2.15	(11.3)
Bike	2.27	(4.8)	1.38	(3.6)	0.996	(2.7)	1.87	(5.8)
Walk	1.32	(6.2)	0.790	(3.3)	0.465	(2.3)	1.32	(7.1)
Drive Alone (base)	0.00		0.00		0.0		0.00	
Number of Vehicles								
Transit	-2.06	(-6.6)	-0.812	(-2.7)	-1.67	(-5.7)	-2.02	(-7.8)
Shared Ride 2	-0.248	(-4.7)	-0.127	(-3.2)	-0.0562	(-2.3)	-0.376	(-7.8)
Shared Ride 3+	-0.819	(-10.3)	-0.426	(-3.8)	-0.192	(-2.4)	-0.555	(-8.4)
Shared Ride 2/3 & Drive Alone	-0.456	(-7.5)	-0.235	(-3.7)	-0.105	(-2.4)	-0.439	(-10.3)
Shared Ride 2/3	-0.700	(-6.0)	-0.364	(-3.4)	-0.163	(-2.4)	-0.517	(-8.6)
Bike	-0.401	(-2.0)	-0.328	(-2.0)	-0.208	(-1.3)	-0.428	(-2.6)
Walk	-0.714	(-5.5)	-0.627	(-5.3)	-0.479	(-4.0)	-0.693	(-5.4)
Drive Alone (base)	0.00		0.00		0.0		0.00	

Variables	Model 14 S/O		Model 19 S/O		Model 20 S/O		Model 21 S/O	
Nests	None		M-NM		P-NM		S-NM	
Travel Time (minutes)								
Non-Motorized Modes Only	-0.0848	(-8.0)	-0.0867	(-9.8)	-0.08	(-7.9)	-0.0805	(-8.7)
Motorized Modes Only	-0.0344	(-3.1)	-0.0190	(-2.7)	-0.0246	(-2.4)	-0.0252	(-2.8)
OVT by Distance (mi.) Motorized Modes	-0.208	(-3.9)	-0.120	(-2.8)	-0.209	(-4.0)	-0.209	(-4.1)
Travel Cost by Log of Income (1990 cents per log of 1000 1990 dollars)	-0.0111	(-3.6)	-0.0060	(-2.9)	-0.0028	(-2.3)	-0.0069	(-3.0)
Zero Vehicle Household Dummy Variable								
Transit, Bike, Walk	1.57	(4.2)	1.35	(4.1)	1.98	(5.6)	1.54	(4.3)
All Private Vehicle Modes	0.00		0.00		0.0		0.00	
Dummy Variable for Destination in Core								
Transit	2.01	(3.2)	1.11	(2.9)	1.76	(3.4)	2.17	(3.5)
Shared Ride (any)	0.716	(1.8)	0.374	(1.7)	0.154	(1.8)	0.794	(2.0)
Bike, Walk	0.301	(0.2)	0.127	(0.2)	0.103	(0.2)	0.465	(0.3)
Drive Alone (base)	0.00		0.00		0.0		0.00	
Nesting Coefficients/Dissimilarity Parameters								
Non-Motorized Nest			0.746	(-1.5)	0.226	(8.1)	0.708	(-2.0)
Shared Ride Nest							0.304	(-5.7)
Private Automobile Nest					0.703	(1.9)		
Motorized Nest			0.510	(-3.8)				
Log-likelihood at Zero	-6201.516		-6201.516		-6201.516		-6201.516	
Log-likelihood at Constants	-4962.194		-4962.194		-4962.194		-4962.194	
Log-likelihood at Convergence	-4457.235		-4452.593		-4447.480		-4449.592	
Rho Squared w.r.t Zero	0.2813		0.2820		0.2828		0.2825	
Rho Squared w.r.t. Constants	0.1018		0.1027		0.1037		0.1033	
Chi-Squared vs. MNL			9.2840		19.5100		15.2860	
Rejection significance			0.010		.000		.000	
Chi-Squared vs. M, P and S nests			1.4780		1.9940		1.9060	
Rejection significance			.224		.158		.167	
Chi-Sqrd vs. NM nest (Model 18 S/O)			7.5600		17.7860		13.5620	
Rejection significance			.006		.000		.000	

As in the case of work trips, hierarchical two-nest models, depicted in Figure 9.2, are considered next, Table 9-8. Model 22 S/O obtains a satisfactory parameter for the automobile logsum but the motorized logsum is not significantly different from one and the model does not reject the single nest Private Automobile structure (Model 16 S/O). However, Models 23 S/O (with shared ride under the motorized nest) and 24 S/O (with shared ride under the automobile nest) provide excellent results. The similarity interpretation of these nests is reasonable and both

models strongly reject the MNL model and the constrained single nest models reported in Table 9-7.

Table 9-8 Hierarchical Two Nest Models for Shop/Other Trips

Variables	Model 14 S/O		Model 22 S/O		Model 23 S/O		Model 24 S/O	
Nest	None		M-P		M-S		P-S	
Constants								
Transit	0.530	(0.7)	-0.271	(-0.4)	-0.263	(-0.5)	-0.420	(-0.5)
Shared Ride 2	-1.19	(-9.0)	-0.258	(-2.4)	-0.450	(-3.2)	-0.308	(-2.0)
Shared Ride 3+	-3.20	(-13.7)	-0.686	(-2.4)	-0.900	(-2.9)	-0.686	(-2.0)
Shared Ride 2/3 & Drive Alone	-1.78	(-11.4)	-0.381	(-2.4)	-0.569	(-3.1)	-0.409	(-2.0)
Shared Ride 2/3	-4.11	(-11.7)	-0.871	(-2.4)	-1.06	(-2.8)	-0.830	(-2.0)
Bike	-4.41	(-7.3)	-4.48	(-7.9)	-4.19	(-7.6)	-4.50	(-7.6)
Walk	0.461	(1.2)	0.388	(1.0)	0.767	(2.1)	0.320	(0.8)
Drive Alone (base)	0.00		0.0		0.00		0.00	
Log of Persons per Household								
Transit	2.10	(5.4)	0.795	(2.3)	1.21	(3.3)	1.02	(2.2)
Shared Ride 2	1.14	(9.9)	0.244	(2.4)	0.881	(3.4)	0.491	(1.7)
Shared Ride 3+	3.19	(18.3)	0.686	(2.5)	1.33	(3.5)	0.871	(2.0)
Shared Ride 2/3 & Drive Alone	1.90	(14.5)	0.407	(2.5)	1.04	(3.5)	0.628	(1.9)
Shared Ride 2/3	2.89	(11.3)	0.621	(2.5)	1.26	(3.5)	0.814	(2.0)
Bike	2.27	(4.8)	1.29	(3.0)	1.69	(3.7)	1.43	(2.8)
Walk	1.32	(6.2)	0.397	(1.9)	0.786	(3.0)	0.534	(1.9)
Drive Alone (base)	0.00		0.0		0.00		0.00	
Number of Vehicles								
Transit	-2.06	(-6.6)	-1.2	(-2.1)	-0.906	(-2.3)	-1.74	(-5.5)
Shared Ride 2	-0.248	(-4.7)	-0.0514	(-2.4)	-0.207	(-3.2)	-0.112	(-1.6)
Shared Ride 3+	-0.819	(-10.3)	-0.175	(-2.5)	-0.325	(-3.4)	-0.214	(-2.0)
Shared Ride 2/3 & Drive Alone	-0.456	(-7.5)	-0.0957	(-2.4)	-0.249	(-3.4)	-0.148	(-1.8)
Shared Ride 2/3	-0.700	(-6.0)	-0.149	(-2.4)	-0.300	(-3.4)	-0.193	(-1.9)
Bike	-0.401	(-2.0)	-0.163	(-0.8)	-0.286	(-1.4)	-0.191	(-0.9)
Walk	-0.714	(-5.5)	-0.504	(-4.2)	-0.637	(-5.2)	-0.522	(-3.8)
Drive Alone (base)	0.00		0.0		0.00		0.00	
Travel Time (minutes)								
Non-Motorized Modes Only	-0.0848	(-8.0)	-0.0855	(-8.8)	-0.0885	(-9.7)	-0.0833	(-7.9)
Motorized Modes Only	-0.0344	(-3.1)	-0.022	(-2.4)	-0.0162	(-2.2)	-0.0187	(-1.6)
OVT by Distance (mi.) Motorized Modes	-0.208	(-3.9)	-0.169	(-2.9)	-0.123	(-2.6)	-0.193	(-3.7)
Travel Cost by Log of Income (1990 cents per log of 1000 1990 dollars)								
	-0.0111	(-3.6)	-0.0025	(-2.4)	-0.0044	(-2.4)	-0.0034	(-1.7)
Zero Vehicle Household Dummy Variable								
Transit, Bike, Walk	1.57	(4.2)	1.84	(5.6)	1.38	(4.5)	1.89	(5.0)
All Private Vehicle Modes	0.00		0.0		0.00		0.00	
Dummy Variable for Destination in Core								
Transit	2.01	(3.2)	1.46	(2.6)	1.31	(2.7)	1.88	(3.5)
Shared Ride (any)	0.716	(1.8)	0.142	(2.0)	0.449	(1.9)	0.240	(1.1)
Bike, Walk	0.301	(0.2)	0.0879	(0.4)	0.248	(0.3)	0.183	(0.2)
Drive Alone (base)	0.00		0.0		0.00		0.00	

Variables	Model 14 S/O	Model 22 S/O	Model 23 S/O	Model 24 S/O
Nest	None	M-P	M-S	P-S
Nesting Coefficients/Dissimilarity Parameters				
Non-Motorized Nest			0.197 (-9.6)	0.169 (-9.4)
Shared Ride Nest				
Private Automobile Nest		0.207 (9.3)		0.331 (-3.6)
Motorized Nest		0.791 (0.9)	0.563 (-2.7)	
Log-likelihood at Zero	-6201.516	-6201.516	-6201.516	-6201.516
Log-likelihood at Constants	-4962.194	-4962.194	-4962.194	-4962.194
Log-likelihood at Convergence	-4457.235	-4448.234	-4448.221	-4446.677
Rho Squared w.r.t Zero	0.2813	0.2827	0.2827	0.2830
Rho Squared w.r.t. Constants	0.1018	0.1036	0.1036	0.1039
Chi-Squared vs. MNL		18.0020	18.0280	21.1160
Rejection significance		.001	.000	.001
Chi-Squared vs. Upper Nest Only		10.1960	10.2220	3.6000
Rejection significance		.001	.001	.058
Chi-Squared vs. Lower Nest Only		0.4860	4.6480	7.7360
Rejection significance		.486	.031	.005

Unlike the case of the home-based work models, the home-based shop/other data set supports complex nested models which combine the hierarchical and parallel nests to capture the different substitutability between several groups of alternatives. Models 25 S/O and 26 S/O improve goodness-of-fit over both the best hierarchical (Model 24 S/O) and parallel (Model 20 S/O) two-nest models. These more complex models can reject all of the single nest models at the .05 level or higher, but cannot reject the best two-nest models at a high level of confidence. Even so, they may still be preferred for their potentially more realistic representation of trade-offs between pairs of alternatives. The same rationale could be applied between the two models, as Model 26 S/O cannot reject 25 S/O, either, but allows for greater substitutability between the non-motorized modes (for which the statistical evidence is moderately significant). The nesting coefficient for the motorized nest is not particularly significant in either model, but it can be retained given the reasonableness of its value.

Table 9-9 Complex Nested Models for Shop/Other Trips

Variables	Model 25 S/O		Model 26 S/O	
Nest	M-P-S		M-P-S-NM	
Constants				
Transit	-0.439	(-0.8)	-0.406	(-0.8)
Shared Ride 2	-0.280	(-2.1)	-0.282	(-2.1)
Shared Ride 3+	-0.621	(-2.1)	-0.624	(-2.1)
Shared Ride 2/3 & Drive Alone	-0.372	(-2.1)	-0.374	(-2.1)
Shared Ride 2/3	-0.750	(-2.0)	-0.753	(-2.0)
Bike	-4.39	(-7.4)	-3.55	(-6.0)
Walk	0.490	(1.2)	0.359	(0.9)
Drive Alone (base)	0.00		0.00	
Log of Persons per Household				
Transit	0.891	(2.2)	0.905	(2.6)
Shared Ride 2	0.451	(1.8)	0.454	(2.0)
Shared Ride 3+	0.793	(2.1)	0.798	(2.2)
Shared Ride 2/3 & Drive Alone	0.575	(2.0)	0.578	(2.1)
Shared Ride 2/3	0.742	(2.1)	0.747	(2.1)
Bike	1.39	(3.2)	1.09	(2.9)
Walk	0.498	(1.9)	0.542	(2.3)
Drive Alone (base)	0.00		0.00	
Number of Vehicles				
Transit	-1.17	(-1.9)	-1.24	(-2.2)
Shared Ride 2	-0.104	(-1.7)	-0.104	(-1.8)
Shared Ride 3+	-0.195	(-2.1)	-0.196	(-2.1)
Shared Ride 2/3 & Drive Alone	-0.136	(-1.9)	-0.137	(-2.0)
Shared Ride 2/3	-0.176	(-2.0)	-0.177	(-2.0)
Bike	-0.193	(-0.8)	-0.239	(-1.5)
Walk	-0.534	(-4.1)	-0.518	(-4.1)
Drive Alone (base)	0.00		0.00	
Travel Time (minutes)				
Non-Motorized Modes Only	-0.0859	(-8.7)	-0.0824	(-9.0)
Motorized Modes Only	-0.0167	(-1.8)	-0.0167	(-1.9)
OVT by Distance (mi.) Motorized Modes	-0.155	(-2.6)	-0.167	(-3.1)
Travel Cost by Log of Income (1990 cents per log of 1000 1989 dollars)				
	-0.0031	(-1.7)	-0.0031	(-1.8)
Zero Vehicle Household Dummy Variable				
Transit, Bike, Walk	1.74	(5.0)	1.78	(5.2)
All Private Vehicle Modes	0.00		0.00	
Dummy Variable for Destination in Core				
Transit	1.49	(3.0)	1.53	(3.1)
Shared Ride (any)	0.222	(1.3)	0.223	(1.2)
Bike, Walk	0.138	(0.2)	0.124	(0.0)
Drive Alone (base)	0.00		0.00	
Nesting Coefficients/Dissimilarity Parameters				
Non-Motorized Nest			0.715	(-2.2)
Shared Ride Nest	0.152	(-10.7)	0.153	(-10.6)
Private Automobile Nest	0.304	(-4.5)	0.306	(-4.6)
Motorized Nest	0.755	(-0.9)	0.789	(-0.8)
Log Likelihood at Zero	-6201.516		-6201.516	
Log Likelihood at Constants	-4962.194		-4962.194	
Log Likelihood at Convergence	-4446.355		-4445.434	

Variables	Model 25 S/O	Model 26 S/O
Nest	M-P-S	M-P-S-NM
Rho Squared w.r.t Zero	0.2830	0.2832
Rho Squared w.r.t. Constants	0.1040	0.1041
Chi-Squared vs. MNL	21.7600	23.6020
Rejection Significance	0.000	0.000
Chi-Squared vs. Model S/O20	2.2500	4.0920
Rejection Significance	0.134	0.129
Chi-Squared vs. Model S/O24	0.6440	2.4860
Rejection Significance	0.422	0.289
Chi-Squared vs. Model S/O25		1.8420
Rejection Significance		0.175

9.4 Practical Issues and Implications

A large number of nested logit structures can be proposed for any context in which the number of alternatives is not very small. As seen, even in the case of four alternatives, there are 25 possible nest structures and this number increases to 235 for five alternatives and beyond 2000 for six or more alternatives. Since searching across this many nest structures is generally infeasible, it is the responsibility of the analyst to propose a subset of these structures for primary consideration or, alternatively, to eliminate a subset of such structures from consideration. At the same time, it is desirable for the analyst to remain open to other possible structures that may be suggested by others who have familiarity with the behavior under study.

It is not uncommon for some or all of the proposed structures to be infeasible due to obtaining an estimated logsum or nesting parameter that is greater than one or greater than the parameter in a higher level nest. It is a matter of judgment whether to eliminate the proposed structure based on the estimation results or to constrain selected nest parameters to fixed values or to relative values that ensure that the structure is consistent with utility maximization. This ‘imposition’ of a fixed value or other constraints on nest parameters requires careful judgment, discussion with other modeling and policy analysts and open reporting so that potential users of the model are aware of the basis for the proposed model.

An important part of the decision process is the differential interpretation associated with any nested logit model relative to the multinomial logit model or other nested logit models. The central element of such interpretation is the way in which a change in any of the characteristics

of an alternative affects the probability of that alternative and each other alternative being chosen. This is the essential element of competition/substitution between pairs of alternatives and may dramatically influence the predicted impact of selected policy changes.

CHAPTER 10: Multiple Maxima in the Estimation of Nested Logit Models

10.1 Multiple Optima

One of the most important differences between the Multinomial Logit (MNL) and Nested Logit (NL) models is that there exists a unique optimum for the set of parameters in a MNL model but not necessarily in a NL model. This means that the estimation of an MNL model will result in identical estimation parameters independent of the initial values adopted for such parameters at the start of estimation. On the other hand, the parameters of an NL model may include multiple optima, some or all (or none) of which may be consistent with utility maximization. This imposes an additional responsibility on the analyst; to ensure that the best feasible solution (consistent with utility maximization) be found if one exists⁴⁸.

This issue is illustrated for four cases; one based on Work Mode Choice and three based on Shop/Other Mode Choice. Multiple results are reported for work mode choice with Private Automobile and Non-Motorized Nests in Table 10-1. The private automobile nest parameter is consistently greater than one; however, the shared ride nest parameter varies dramatically from large negative to positive but less than one depending on the starting parameters used. In those cases where the shared ride nest parameter is negative, model parameters for the SR2 and SR3+ alternatives take on extreme values raising questions about the acceptability of the model.

⁴⁸ The probability of finding the best feasible solution can be increased by using a variety of heuristics that take account of both feasibility and maximization. This problem is distinctly different from the use of algorithms designed to find the global optimum as the objective in estimating NL models is not finding the global optimum, but a feasible optimum.

Table 10-1 Multiple Solutions for Model 27W (See Table 9-3)

	Model 27W	Model 27W	Model 27W	Model 27W
Nests	P-S	P-S	P-S	P-S
Initial P-S Nest Parameters	1.0, 0.1	0.1, 0.1	0.5, 0.5	0.3, 0.5 / 1.0, 1.0
Travel Cost by Income (1990 cents per 1000 1989 dollars)	-0.0592 (-4.8)	-0.062 (-13.4)	-0.059 (-4.8)	-0.053 (-7.2)
Travel Time (minutes)				
Motorized Modes Only	-0.019 (-5.0)	-0.0194 (-5.0)	-0.019 (-4.9)	-0.021 (-5.9)
Non-Motorized Modes Only	-0.0462 (-8.1)	-0.0461 (-8.1)	-0.046 (-8.1)	-0.046 (-8.0)
OVT by Distance (mi.)				
Motorized Modes	-0.135 (-6.9)	-0.136 (-8.5)	-0.136 (-7.9)	-0.137 (-8.2)
Household Income (1,000's of 1989 dollars)				
All Private Vehicle Modes	0.0	0.0	0.0	0.0
Transit	-0.0044 (-2.3)	-0.0044 (-2.4)	-0.005 (-2.3)	-0.005 (-2.4)
Bike	-0.0075 (-1.6)	-0.0074 (-1.7)	-0.007 (-1.7)	-0.008 (-1.7)
Walk	-0.005 (-1.7)	-0.005 (-1.7)	-0.005 (-1.7)	-0.005 (-1.8)
Vehicles per Worker				
Drive Alone (base)	0.0	0.0	0.0	0.0
Shared Ride (any)	-0.518 (-4.7)	-0.581 (-13.6)	-0.518 (-4.8)	-0.515 (-4.9)
Transit	-0.872 (-7.6)	-0.9 (-8.3)	-0.875 (-7.8)	-0.86 (-7.9)
Bike	-0.607 (-2.7)	-0.628 (-2.8)	-0.608 (-2.8)	-0.611 (-2.8)
Walk	-0.615 (-4.0)	-0.63 (-4.1)	-0.61 (-4.2)	-0.616 (-4.2)
Dummy Variable for Destination in CBD				
Drive Alone (base)	0.0	0.0	0.0	0.0
Shared Ride 2	136 (10.9)	7.18 (13.7)	1.90 (1.4)	0.620 (4.1)
Shared Ride 3+	-371 (-10.7)	-17.8 (-13.7)	-3.05 (-0.8)	0.889 (5.6)
Transit	1.39 (8.4)	1.4 (9.1)	1.39 (7.7)	1.38 (7.7)
Bike	0.533 (1.5)	0.546 (1.6)	0.528 (1.5)	0.539 (1.5)
Walk	0.118 (0.5)	0.142 (0.6)	0.119 (0.5)	0.133 (0.5)
Empl. Density - Work Zone (employees / square mile)				
Drive Alone (base)	0.0	0.0	0.0	0.0
Shared Ride 2	-0.0167 (-2.7)	0.003 (12.9)	0.0039 (4.3)	0.0030 (6.3)
Shared Ride 3+	0.0338 (3.4)	0.001 (12.7)	0.0013 (0.9)	0.0030 (6.4)
Transit	0.00363 (8.3)	0.0032 (12.4)	0.0036 (8.6)	0.0036 (9.5)
Bike	0.0024 (2.0)	0.002 (1.6)	0.0024 (1.9)	0.0024 (2.0)
Walk	0.0034 (4.3)	0.003 (4.1)	0.0034 (4.4)	0.0034 (4.5)
Constants				
Drive Alone (base)	0.0	0.0	0.0	0.0
Shared Ride 2	63.9 (10.1)	1.38 (13.6)	-1.39 (-1.4)	-2.38 (-12.7)
Shared Ride 3+	722 (10.3)	36.9 (13.8)	7.13 (0.7)	-2.96 (-14.6)
Transit	-0.911 (-3.73)	-0.84 (-3.8)	-0.895 (-3.6)	-0.837 (-3.6)
Bike	-1.81 (-4.68)	-1.79 (-4.7)	-1.81 (-4.7)	-1.81 (-4.7)
Walk	-0.103 (-0.297)	-0.095 (-0.3)	-0.118 (-0.3)	-0.113 (-0.3)
Nesting Coefficients / Dissimilarity Parameters				
Shared Ride Nest	-389. (-10.3)	-21.2 (-14.4)	-5.26 (-1.2)	0.364 (-31.0)
Private Automobile Nest	1.48 (-4.1)	1.46 (-4.3)	1.48 (4.3)	1.48 (4.2)

	Model 27W	Model 27W	Model 27W	Model 27W
Nests	P-S	P-S	P-S	P-S
Initial P-S Nest Parameters	1.0, 0.1	0.1, 0.1	0.5, 0.5	0.3, 0.5 / 1.0, 1.0
Log Likelihood at Zero	-7309.601	-7309.601	-7309.601	-7309.601
Log Likelihood at Constants	-4132.916	-4132.916	-4132.916	-4132.916
Log Likelihood at Convergence	-3435.211	-3442.284	-3440.200	-3433.910
Rho Squared w.r.t Zero	0.5300	0.5291	0.5294	0.5302
Rho Squared w.r.t. Constants	0.1688	0.1671	0.1676	0.1691

This issue is illustrated again in Table 10-2 which reports three estimations for Model 20 S/O with nests for private automobile and non-motorized alternatives. The example of Model 20 S/O not only illustrates the existence of multiple optima, but also that the shape of the log likelihood function may not be easy to intuit. Very different starting points can eventually settle to the same solution while other, nearby starting points can diverge to different results. The first estimate, using initial parameter of 0.5 for both nests obtained a result with a number of very small parameter estimates; those in the utility function have no significance while the private automobile nest parameter has an extremely high level of significance. The other two estimations obtain essentially identical estimates with reasonable parameters for the utility function and nesting parameters. Interestingly, the goodness of fit for the second and third estimates is superior to that for the first estimate.

Table 10-2 Multiple Solutions for Model 20 S/O (See Table 9-7)

Variables	Model 20 S/O		Model 20 S/O		Model 20 S/O	
Nests	P-NM		P-NM		P-NM	
Initial Parameters	0.5, 0.5		0.2, 0.7		1.0, 1.0	
Constants						
Transit	-1.34	(-2.3)	-0.17	(-0.2)	-0.169	(-0.2)
Shared Ride 2	-0.0001	(0.0)	-0.282	(-2.4)	-0.282	(-1.9)
Shared Ride 3+	-0.0003	(0.0)	-0.75	(-2.4)	-0.751	(-2.0)
Shared Ride 2/3 & Drive Alone	-0.0002	(0.0)	-0.416	(-2.4)	-0.417	(-2.0)
Shared Ride 2/3	-0.0004	(0.0)	-0.952	(-2.3)	-0.953	(-2.0)
Bike	-3.34	(-6.6)	-3.69	(-6.0)	-3.69	(-5.0)
Walk	0.141	(0.4)	0.129	(0.3)	0.129	(0.3)
Drive Alone (base)	0.0		0.0		0.0	
Log of Persons per Household						
Transit	0.545	(1.5)	0.899	(2.2)	0.9	(2.2)
Shared Ride 2	0.0001	(0.0)	0.266	(2.4)	0.267	(2.0)
Shared Ride 3+	0.0003	(0.0)	0.75	(2.4)	0.75	(2.0)
Shared Ride 2/3 & Drive Alone	0.0002	(0.0)	0.445	(2.4)	0.445	(2.0)
Shared Ride 2/3	0.0003	(0.0)	0.678	(2.3)	0.679	(2.0)
Bike	0.667	(2.3)	0.996	(2.7)	0.997	(2.2)
Walk	0.234	(1.4)	0.465	(2.3)	0.465	(2.1)
Drive Alone (base)	0.0		0.0		0.0	
Number of Vehicles						
Transit	-1.61	(-6.4)	-1.67	(-5.7)	-1.67	(-5.6)
Shared Ride 2	-2E-05	(0.0)	-0.0562	(-2.3)	-0.0563	(-1.1)
Shared Ride 3+	-7E-05	(0.0)	-0.192	(-2.4)	-0.192	(-2.0)
Shared Ride 2/3 & Drive Alone	-4E-05	(0.0)	-0.105	(-2.4)	-0.105	(-2.0)
Shared Ride 2/3	-6E-05	(0.0)	-0.163	(-2.4)	-0.164	(-1.9)
Bike	-0.173	(-1.1)	-0.208	(-1.3)	-0.209	(-0.9)
Walk	-0.421	(-3.4)	-0.479	(-4.0)	-0.479	(-3.7)
Drive Alone (base)	0.0		0.0		0.0	
Travel Time (minutes)						
Non-Motorized Modes Only	-0.0762	(-8.9)	-0.08	(-7.9)	-0.08	(-8.2)
Motorized Modes Only	-9E-07	(0.0)	-0.0246	(-2.4)	-0.0246	(-2.2)
OVT by Distance (mi.) Motorized Modes	-0.183	(-3.8)	-0.209	(-4.0)	-0.208	(-4.1)
Travel Cost by Log of Income (1990 cents per log of 1000 1989 dollars)						
	-1E-06	(0.0)	-0.0028	(-2.3)	-0.0028	(-1.3)
Zero Vehicle Household Dummy Variable						
Transit, Bike, Walk	2.05	(5.8)	1.98	(5.6)	1.98	(5.4)
All Private Vehicle Modes			0.0			
Dummy Variable for Destination in Core						
Transit	1.85	(3.8)	1.76	(3.4)	1.76	(3.0)
Shared Ride (any)	6E-05	(0.0)	0.154	(1.8)	0.155	(0.5)
Bike, Walk	0.0902	(0.6)	0.103	(0.2)	0.103	(0.1)
Drive Alone (base)			0.0			

Variables	Model 20 S/O	Model 20 S/O	Model 20 S/O
Nests	P-NM	P-NM	P-NM
Initial Parameters	0.5, 0.5	0.2, 0.7	1.0, 1.0
Nesting Coefficients/Dissimilarity Parameters			
Non-Motorized Nest	0.579 (-5.6)	0.226 (8.1)	0.227 (7.2)
Shared Ride Nest			
Private Automobile Nest	9E-05 (-242.7)	0.703 (1.9)	0.703 (1.8)
Motorized Nest			
Log Likelihood at Zero	-6201.516	-6201.516	-6201.516
Log Likelihood at Constants	-4962.194	-4962.194	-4962.194
Log Likelihood at Convergence	-4450.573	-4447.480	-4447.480
Rho Squared w.r.t Zero	0.2823	0.2828	0.2828
Rho Squared w.r.t. Constants	0.1031	0.1037	0.1037

The complexity of the log likelihood objective function is also highlighted in Table 10-3. Model 22 S/O converges to a reasonable pair of nest parameters and the highest goodness of fit when the MNL solution parameters and a wide range of initial nesting parameters (including 0.25/0.75; 0.5/0.5; 1.0/1.0; 0.3/0.7; 0.4/0.6; 0.6/0.8; and 0.2/0.5) are used (first column). However, a variety of other initial nest parameters (columns 2, 3 and 4) produce unacceptable nest parameters and inferior goodness of fit. Perhaps surprisingly, if the model is started with nest parameters very close to the preferred values; *e.g.*, 0.2 and 0.8, the correct solution is not found. Therefore, even *a priori* knowledge of the nest parameters is not guaranteed to avoid searching from multiple initial nest parameter values.

Table 10-3 Multiple Solutions for Model 22 S/O (See Table 9-8)

Variables	Model 22 S/O	Model 22 S/O	Model 22 S/O	Model 22 S/O
Initial Parameters	Multiple Pairs	0.1/1.0	0.1/0.8	0.2/0.8
Constants				
Transit	-0.269 (-0.5)	-1.55 (-2.3)	-1.44 (-2.9)	-1.49 (-1.9)
Shared Ride 2	-0.258 (-1.6)	-0.0492 (-1.3)	-0.0241 (-8.9)	-0.0104 (-3.7)
Shared Ride 3+	-0.685 (-1.7)	-0.134 (-1.3)	-0.0635 (-13.4)	-0.0274 (-4.0)
Shared Ride 2/3 & Drive Alone	-0.38 (-1.7)	-0.0681 (-1.3)	-0.0353 (-11.6)	-0.0152 (-3.8)
Shared Ride 2/3	-0.87 (-1.8)	-0.157 (-1.3)	-0.0806 (-11.8)	-0.0348 (-4.0)
Bike	-4.48 (-8.0)	-4.5 (-7.2)	-4.48 (-7.6)	-4.5 (-7.2)
Walk	0.388 (0.9)	0.262 (0.6)	0.306 (0.7)	0.278 (0.6)
Drive Alone (base)	0.0	0.0	0.0	0.0
Log of Persons per Household				
Transit	0.795 (2.1)	0.632 (1.4)	0.554 (1.7)	0.555 (0.9)
Shared Ride 2	0.244 (1.7)	0.0495 (1.3)	0.0226 (9.6)	0.0097 (3.6)
Shared Ride 3+	0.685 (1.8)	0.128 (1.3)	0.0635 (17.1)	0.0274 (4.0)
Shared Ride 2/3 & Drive Alone	0.406 (1.7)	0.077 (1.3)	0.0376 (14.0)	0.0162 (3.9)
Shared Ride 2/3	0.62 (1.8)	0.112 (1.3)	0.0575 (11.3)	0.0248 (4.0)
Bike	1.29 (3.0)	1.08 (2.5)	1.05 (2.6)	1.04 (2.3)
Walk	0.396 (1.7)	0.199 (1.0)	0.175 (1.0)	0.162 (0.9)
Drive Alone (base)	0.0	0.0	0.0	0.0
Number of Vehicles				
Transit	-1.2 (-1.6)	-2.09 (-1.7)	-1.77 (-2.1)	-1.88 (-1.2)
Shared Ride 2	-0.0514 (-1.4)	-0.0112 (-1.4)	-0.0047 (-4.5)	-0.002 (-2.7)
Shared Ride 3+	-0.175 (-1.7)	-0.0314 (-1.3)	-0.0162 (-9.9)	-0.007 (-3.7)
Shared Ride 2/3 & Drive Alone	-0.0956 (-1.7)	-0.0194 (-1.4)	-0.0088 (-7.7)	-0.0038 (-3.4)
Shared Ride 2/3	-0.149 (-1.7)	-0.0271 (-1.4)	-0.0138 (-6.0)	-0.0059 (-3.3)
Bike	-0.163 (-0.8)	-0.105 (-0.5)	-0.101 (-0.5)	-0.0955 (-0.4)
Walk	-0.504 (-4.0)	-0.439 (-3.4)	-0.437 (-3.4)	-0.432 (-3.4)
Drive Alone (base)	0.0	0.0	0.0	0.0
Travel Time (minutes)				
Non-Motorized Modes Only	-0.0855 (-8.5)	-0.0867 (-9.8)	-0.0819 (-7.0)	-0.0826 (-8.1)
Motorized Modes Only	-0.022 (-1.7)	-0.019 (-2.7)	-3E-05 (-0.8)	-0.0008 (-0.6)
OVT by Distance (mi.) Motorized Modes	-0.169 (-2.7)	-0.12 (-2.8)	-0.183 (-2.0)	-0.176 (-2.6)
Travel Cost by Log of Income (1990 cents per log of 1000 1989 dollars)				
	-0.0025 (-1.3)	-0.0006 (-1.4)	-0.0003 (-3.7)	-0.0001 (-2.7)
Zero Vehicle Household Dummy Variable				
Transit, Bike, Walk	1.84 (5.0)	2.11 (4.9)	2.08 (5.0)	2.11 (5.1)
Personal Vehicle Modes	0.0	0.0	0.0	0.0
Dummy Variable for Destination in Core				
Transit	1.46 (2.5)	1.11 (2.9)	2.09 (2.1)	2 (2.9)
Shared Ride (any)	0.142 (0.8)	0.374 (1.7)	0.0015 (4.3)	0.0127 (1.8)
Bike, Walk	0.0884 (0.4)	0.127 (0.2)	0.137 (0.2)	0.128 (0.2)
Drive Alone (base)	0.0	0.0	0.0	0.0

Variables	Model 22 S/O	Model 22 S/O	Model 22 S/O	Model 22 S/O
Initial Parameters	Multiple Pairs	0.1/1.0	0.1/0.8	0.2/0.8
Nesting Coefficients/Dissimilarity Parameters				
Non-Motorized Nest				
Shared Ride Nest				
Automobile Nest	0.207 (-6.7)	0.0387 (32.5)	0.019 (3715.9)	0.0082 (503.4)
Motorized Nest	0.791 (-0.7)	1.17 (-0.4)	1.05 (-0.1)	1.09 (-0.1)
Log Likelihood at Zero	-6201.516	-6201.516	-6201.516	-6201.516
Log Likelihood at Constants	-4962.194	-4962.194	-4962.194	-4962.194
Log Likelihood at Convergence	-4448.234	-4454.884	-4451.034	-4451.070
Rho Squared w.r.t Zero	0.2827	0.2816	0.2823	0.2823
Rho Squared w.r.t. Constants	0.1036	0.1022	0.1030	0.1030

Throughout CHAPTER 9, 0.5 was used as the starting value for all nest parameters, except for Models 27W and 20 S/O above. Although there is no theoretical reason for using this starting point, some limited experimentation lead to adopting this practice over using 1.0 (the MNL solution) as the starting point. When all the nest parameters are started at 1.0, estimation frequently resulted in infeasible nesting coefficients. Table 10-4 illustrates two examples of this result. The result that 0.5 worked well with these models and data sets may or may not generalize to other specifications or data sets, and should not necessarily be taken as a general rule. Rather, similar experimentation should be undertaken in any case where nesting parameters fall outside the desired zero to one range.

Table 10-4 Multiple Solutions for Complex S/O Models (See Table 9-9)

Variables	Model 25-A S/O	Model 25 S/O	Model 26-A S/O	Model 26 S/O
Nest	M-P-S	M-P-S	M-P-S-NM	M-P-S-NM
Initial Search Value for Nesting Coefficients	1.0	0.5	1.0	0.5
Constants				
Transit	0.137 (0.2)	-0.439 (-0.8)	0.196 (0.2)	-0.406 (-0.8)
Shared Ride 2	0.0405 (0.6)	-0.280 (-2.1)	0.0409 (0.6)	-0.282 (-2.1)
Shared Ride 3+	0.796 (1.2)	-0.621 (-2.1)	0.804 (1.2)	-0.624 (-2.1)
Shared Ride 2/3 & Drive Alone	0.310 (1.1)	-0.372 (-2.1)	0.313 (1.1)	-0.374 (-2.1)
Shared Ride 2/3	1.10 (1.2)	-0.750 (-2.0)	1.11 (1.2)	-0.753 (-2.0)
Bike	-4.57 (-7.7)	-4.39 (-7.4)	-3.72 (-5.0)	-3.55 (-6.0)
Walk	0.256 (0.6)	0.490 (1.2)	0.131 (0.3)	0.359 (0.9)
Drive Alone (base)	0.00	0.00	0.00	0.00

Variables	Model 25-A S/O	Model 25 S/O	Model 26-A S/O	Model 26 S/O
Nest	M-P-S	M-P-S	M-P-S-NM	M-P-S-NM
Initial Search Value for Nesting Coefficients	1.0	0.5	1.0	0.5
Log of Persons per Household				
Transit	0.760 (1.7)	0.891 (2.2)	0.769 (1.7)	0.905 (2.6)
Shared Ride 2	0.582 (1.3)	0.451 (1.8)	0.588 (1.3)	0.454 (2.0)
Shared Ride 3+	-0.179 (-0.8)	0.793 (2.1)	-0.180 (-0.9)	0.798 (2.2)
Shared Ride 2/3 & Drive Alone	0.312 (1.2)	0.575 (2.0)	0.315 (1.3)	0.578 (2.1)
Shared Ride 2/3	-0.0577 (-0.4)	0.742 (2.1)	-0.0587 (-0.4)	0.747 (2.1)
Bike	1.20 (2.8)	1.39 (3.2)	0.906 (2.0)	1.09 (2.9)
Walk	0.342 (1.4)	0.498 (1.9)	0.383 (1.6)	0.542 (2.3)
Drive Alone (base)	0.00	0.00	0.00	0.00
Number of Vehicles				
Transit	-1.39 (-2.1)	-1.17 (-1.9)	-1.45 (-2.0)	-1.24 (-2.2)
Shared Ride 2	-0.138 (-1.3)	-0.104 (-1.7)	-0.139 (-1.3)	-0.104 (-1.8)
Shared Ride 3+	0.0481 (0.8)	-0.195 (-2.1)	0.0484 (0.8)	-0.196 (-2.1)
Shared Ride 2/3 & Drive Alone	-0.0869 (-1.2)	-0.136 (-1.9)	-0.0877 (-1.2)	-0.137 (-2.0)
Shared Ride 2/3	-0.0017 (0.0)	-0.176 (-2.0)	-0.0017 (0.0)	-0.177 (-2.0)
Bike	-0.130 (-0.5)	-0.193 (-0.8)	-0.181 (-1.0)	-0.239 (-1.5)
Walk	-0.474 (-3.6)	-0.534 (-4.1)	-0.458 (-3.5)	-0.518 (-4.1)
Drive Alone (base)	0.00	0.00	0.00	0.00
Travel Time (minutes)				
Non-Motorized Modes Only	-0.0844 (-8.7)	-0.0859 (-8.7)	-0.0809 (-8.0)	-0.0824 (-9.0)
Motorized Modes Only	-0.0337 (-2.4)	-0.0167 (-1.8)	-0.0339 (-2.5)	-0.0167 (-1.9)
OVT by Distance (mi.) Motorized Modes	-0.199 (-3.2)	-0.155 (-2.6)	-0.210 (-3.3)	-0.167 (-3.1)
Travel Cost by Log of Income (1990 cents per log of 1000 1989 dollars)	-0.0007 (-0.6)	-0.0031 (-1.7)	-0.0007 (-0.6)	-0.0031 (-1.8)
Zero Vehicle Household Dummy Variable				
Transit, Bike, Walk	1.97 (4.6)	1.74 (5.0)	1.98 (4.5)	1.78 (5.2)
All Private Vehicle Modes	0.00	0.00	0.00	0.00
Dummy Variable for Destination in Core				
Transit	1.61 (2.3)	1.49 (3.0)	1.63 (2.3)	1.53 (3.1)
Shared Ride (any)	0.160 (1.0)	0.222 (1.3)	0.161 (1.0)	0.223 (1.2)
Bike, Walk	0.160 (0.2)	0.138 (0.2)	0.141 (0.2)	0.124 (0.0)
Drive Alone (base)	0.00	0.00	0.00	0.00
Nesting Coefficients/Dissimilarity Parameters				
Non-Motorized Nest			0.711 (-1.6)	0.715 (-2.2)
Shared Ride Nest	-0.332 (-4.9)	0.152 (-10.7)	-0.335 (-4.9)	0.153 (-10.6)
Private Automobile Nest	0.188 (-5.4)	0.304 (-4.5)	0.190 (-5.5)	0.306 (-4.6)
Motorized Nest	0.906 (-0.3)	0.755 (-0.9)	0.932 (-0.2)	0.789 (-0.8)
Log Likelihood at Zero	-6201.516	-6201.516	-6201.516	-6201.516
Log Likelihood at Constants	-4962.194	-4962.194	-4962.194	-4962.194
Log Likelihood at Convergence	-4447.275	-4446.355	-4446.320	-4445.434
Rho Squared w.r.t Zero	0.2829	0.2830	0.2830	0.2832
Rho Squared w.r.t. Constants	0.1038	0.1040	0.1040	0.1041
Chi-Squared vs. MNL	19.9200	21.7600	21.8300	23.6020
Confidence	100.0%	100.0%	100.0%	100.0%

CHAPTER 11: Aggregate Forecasting, Assessment, and Application

11.1 Background

Discrete choice models explain the choice behavior of individuals as a function of individual characteristics and attributes of the alternatives in the individual's choice set. However, an important objective of discrete choice analysis from is to predict the group behavior of individuals due to changes in socio-demographic characteristics over time and/or changes in attributes of alternatives.

The previous chapters have discussed the specification and estimation of travel mode choice models. This chapter describes an aggregation approach to predicting the mode choice of a group of individuals from the estimated choice parameters and from relevant information regarding the current or future values (due to socio-demographic changes or policy actions) of exogenous variables. This chapter also discusses issues related to the aggregate assessment of the performance of mode choice models and the application of the models to evaluate policy actions.

11.2 Aggregate Forecasting

The first issue in aggregate forecasting is to define the population for which we are seeking aggregate predictions. In the context of travel mode choice, the population will be all (or a subset of) residents of the metropolitan region of interest. If the focus of the study is region-wide, the predictions are required for the behavior of the full population. However, if the focus of the study is on a specific corridor or socioeconomic subgroup, it may be satisfactory to consider only the population that is relevant to the study. Finally, even if the focus is on a sub-group of the population, it may be necessary to predict more population groups to account for interaction between groups.

Once the population of interest is defined and if we know the values of the exogenous variables x_{ni} for each individual n and alternative i in the desired population, the expected

number of individuals choosing travel mode i can be obtained in a conceptually straight-forward fashion:

$$\widehat{N}_i = \sum_{n=1}^N P_i(x_{ni}, \widehat{\theta}) = \sum_{n=1}^N \widehat{P}_{ni} \quad 11.1$$

Where $\widehat{\theta}$ is the expected value of the vector of parameters obtained in the estimation phase and \widehat{P}_{ni} is the estimated probability of choosing mode i for individual n . The expected share of the population using mode i can be obtained from equation 11.1 as:

$$\widehat{S}_i = \frac{1}{N} \sum_{n=1}^N P_i(x_{ni}, \widehat{\theta}) = \frac{1}{N} \sum_{n=1}^N \widehat{P}_{ni} \quad 11.2$$

There are two sources of variation associated with the expressions in equations 11.1 and 11.2. The first source is the probabilistic form of the discrete choice model at the individual-level. Even if the true value of the parameter vector θ is known, the probabilistic form of choice at the individual level implies the presence of sampling variance in the aggregate number or share of individuals choosing mode i . This sampling variance declines with the size N of the population for which an aggregate prediction is desired⁴⁹. If we are actually making predictions for the entire population, N is likely to be large enough so that the sampling variance can be ignored (see Cramer, 1991; page 86 for a detailed discussion). Of course, we never know the true value of the vector θ . We only have an estimate $\widehat{\theta}$ of the true value. This gives rise to the second source of variation, which we label as the estimation variance. An estimate of the asymptotic variance of the aggregate share prediction due to this estimation variance can be written as:

$$Var(\widehat{S}_i) = d_i' Var(\widehat{\theta}) d_i, \text{ where } d_i = \frac{1}{N} \sum_{n=1}^N \left[\widehat{P}_{ni} \left(x_{ni} - \sum_j \widehat{P}_{nj} \cdot x_{nj} \right) \right] \quad 11.3$$

⁴⁹ It is important to recognize that a distinct N is associated with each subgroup of the population so that while it may be possible to get a fairly accurate estimate of, for example, mode shares for the entire region; it may not be possible to do so for distinct socio-economic or spatially defined sub-groups.

$Var(\hat{\theta})$ is the asymptotic variance-covariance matrix of the parameters obtained in estimation and d_i' is the transpose matrix of d_i . If the dimension of the vector x_{ni} is $K \times 1$, then the dimension of d_i is also $K \times 1$ and the dimension of $Var(\hat{\theta})$ is $K \times K$.

The discussion thus far has assumed that we know the exogenous variable vector x_{ni} for each individual n in the population for which we desire aggregate predictions. In reality, this is virtually impossible. Practical methods for aggregation attempt to approximate either equation 11.1 or 11.2 while reducing the data needs and the computational burden of a full enumeration procedure which involves the use of data on all individuals in the desired population. Historically, this involved using a sample of the population for prediction in a procedure called sample enumeration. As its name suggests, sample enumeration uses a random sample of the population of interest and then applies equation 11.1 or equation 11.2 to the sample. The sample itself may be obtained in several ways. One approach is to use the estimation sample as the base and change the characteristics of the estimation sample to make it representative of the future population. For example, one can "age" the estimation sample over time based on information about birth and death rates or use temporal income trends to update the estimation sample so it is representative of future conditions. Alternatively, for policy analysis, one can change the level-of-service variables for affected individuals in the estimation sample to obtain a prediction sample for use in sample enumeration.

An alternative approach, which has become more widely used, is to construct a synthetic prediction sample for the entire population through micro-simulation. The synthetic population can be generated for a specific point in time based on a combination of data sources and 'aged' over the prediction period to provide an extensive representation of the full population.

The central issue in the use of this approach is the generation or micro-simulation of the synthetic population and application of the above equations to that population, M .

$$\tilde{S}_i = \frac{1}{M} \sum_{n=1}^M P_i(x_{ni}, \hat{\theta}) = \frac{1}{M} \sum_{n=1}^M \hat{P}_{ni} \quad 11.4$$

$$Var(\hat{S}_i) = d_i' Var(\hat{\theta}) d_i, \text{ where } d_i = \frac{1}{M} \sum_{n=1}^M \left[\hat{P}_{ni} \left(x_{ni} - \sum_j \hat{P}_{nj} \cdot x_{nj} \right) \right] \quad 11.5$$

Different procedures can be used to generate the synthetic population for the region in such a manner that individuals and households with specific characteristics are located in specific spatial locations (see Miller, 1996 and Beckman *et al*, 1966 for a more detailed discussion). Generally, these locations are defined by census tracts as this is the smallest area for which data is commonly available. The approach is based on the use of two levels of census data. These are the marginal distributions of household characteristics for each census tract and multi-way distributions of data included in the Public Use Microdata Sample (PUMS) where each Public Use Microdata Area (PUMA) consists of a set of census tracts. “A two-step iterative proportional fitting (IPF) procedure is used to estimate simultaneously the multi-way distribution for each census tract within a PUMA (Miller, 1996)”, in such a way as to match the marginal distribution of each census tract and the multi-way correlation of the PUMA. Once this has been accomplished, individual households with relevant characteristics can be drawn from the generated multi-way sample for the census tract.

11.3 Aggregate Assessment of Travel Mode Choice Models

In CHAPTER 5 we discussed disaggregate measures of fit that can be used to evaluate the performance of alternative model structures/specifications. The testing of alternative model structures and specifications should be conducted at the disaggregate level to select a preferred model structure/specification. This testing and selection process should not be pursued at the aggregate level. There are several reasons for this. First, in some model structures such as the multinomial logit, the predicted aggregate share of each modal alternative in the estimation sample will be the same regardless of the model specification as long as a full set of alternative-specific constants are included. Thus, aggregate testing of alternative models using the estimation sample is futile. Second, a model that performs poorly at the disaggregate level may perform as well or even better than a model that performs well at the disaggregate level when both models are applied to a hold-out sample to obtain aggregate predictions. For example,

consider the situation when the sample shares in the hold-out sample are exactly the same as in the estimation sample. In this situation, a multinomial logit model with just the alternative specific constants will fit the aggregate shares in the hold-out sample exactly. However, a disaggregate model with additional explanatory variables will not be able to do so. On the other hand, if the characteristics of the hold-out sample are quite different from that of the estimation sample, the latter model is likely to provide a better aggregate fit than the former model. Thus, the aggregate comparison of two models on a hold-out sample is heavily influenced by the composition of the hold-out sample *vis-à-vis* that of the estimation sample. Third, errors in individual-level predictions tend to average out in the aggregate, and so aggregate-level testing does not discriminate much among alternative models. Fourth, the same problems mentioned earlier are applicable (though to a lesser extent) even for semi-aggregate comparisons among alternative models. For example, consider predictions for sub-groups of the estimation sample. While the sub-groups will be different based on the values of one or two exogenous variables, they are likely to be relatively homogenous with respect to other exogenous variables. Again, this will affect any model comparisons conducted at the semi-aggregate level. In fact, the best level for comparison is the natural extension of the semi-aggregate approach to the situation where each sub-group represents an individual; that is, the disaggregate level.

The discussion above emphasizes the need to pursue model testing and selection at the disaggregate level. However, once a model has been selected, it becomes necessary to validate the models at the aggregate level. This validation is used to ensure that the prediction of aggregate behavior, which drives the modeling process, is consistent with the historically observed aggregate data and by extension future aggregate behavior. Effective validation requires matching prediction to observations at easily observable and measurable locations. These include screen-line crossings, total crossings of a major barrier such as a river, vehicle travel on selected major roadways and passenger ridership on selected transit routes. The predicted shares can be compared to the actual shares and summary measures such as the root-mean square error or average absolute error percentage may be computed (see Ben-Akiva and Lerman, 1985, page 210 for an example) to assess the degree to which the predictions

correspond to the aggregate observed data. This requires aggregate predictions based on the micro-simulation procedures described in the preceding section.

Any systematic errors in prediction would require review and revision of the underlying models to ensure that the models satisfy aggregate observations while retaining the most important elements of the underlying disaggregate models.

CHAPTER 12: Recent Advances in Discrete Choice Modeling

12.1 Background

In this chapter, we provide an overview of the motivation for, and structure of, advanced discrete choice models. The discussion is intended to familiarize readers with structural alternatives to the multinomial logit and nested logit models. It is not intended to provide the detailed mathematical formulations or the estimation techniques for these advanced models. Appropriate references are provided for readers interested in these details. This chapter draws from a resource paper presented by Bhat at the 2003 International Association of Travel Behavior Research held in Lucerne, Switzerland (see Bhat, 2005).

Before proceeding to review advanced discrete choice models, we first discuss two important assumptions of the multinomial logit (MNL) formulation. The first assumption in the MNL model is that the random components of the utilities of the different alternatives are independent and identically distributed (IID). The assumption of independence implies that there are no common unobserved factors affecting the utilities of the various alternatives. This assumption is violated, for example, if a decision-maker assigns a higher utility to all transit modes (bus, train, *etc.*) because of the opportunity to socialize or if the decision maker assigns a lower utility to all the transit modes because of the lack of privacy. In such situations, the same underlying unobserved factor (opportunity to socialize or lack of privacy) impacts the utilities of multiple modes. As indicated in CHAPTER 8, the presence of such common underlying factors across modal utilities has implications for competitive structure. The assumption of identically distributed (across alternatives) random utility terms implies that the variation in unobserved factors affecting modal utility is the same across all modes. In general, there is no theoretical reason to believe that this will be the case. For example, if comfort is an unobserved variable whose values vary considerably for the train mode (based on, say, the degree of crowding on different train routes) but little for the automobile mode, then the random components for the automobile and train modes will have different variances. Unequal error variances have significant implications for competitive structure, as discussed in detail by Bhat (1995).

A second assumption of the MNL model is that it maintains homogeneity in responsiveness to attributes of alternatives across individuals (*i.e.*, an assumption of response homogeneity). More specifically, the MNL model does not allow sensitivity variations to an attribute (for example, travel cost or travel time in a mode choice model) due to unobserved individual characteristics. However, unobserved individual characteristics can and generally will affect responsiveness. For example, some individuals by their intrinsic nature may be extremely time-conscious while other individuals may be "laid back" and less time-conscious. Ignoring the effect of unobserved individual attributes can lead to biased and inconsistent parameter and choice probability estimates (see Chamberlain, 1980).

In the rest of this Chapter, we discuss three types of advanced discrete choice model structures that relax one or both of the MNL assumptions discussed above. These three structures correspond to: (1) The GEV class of models, (2) The mixed multinomial logit (MMNL) class of models, and (3) The mixed GEV (MGEV) class of models.

12.2 The GEV Class of Models

The GEV-class of models relaxes the independence from irrelevant alternatives (IIA) property of the multinomial logit model by relaxing the independence assumption between the error terms of alternatives. In other words, a generalized extreme value error structure is used to characterize the unobserved components of utility as opposed to the univariate and independent extreme value error structure used in the multinomial logit model. There are three important characteristics of all GEV models: (1) the overall variances of the alternatives (*i.e.*, the scale of the utilities of alternatives) are assumed to be identical across alternatives, (2) the choice probability structure takes a closed-form expression, and (3) all GEV models collapse to the MNL model when the parameters generating correlation take values that reduce the correlations between *each pair* of alternatives to zero. With respect to the last point, it has to be noted that the MNL model is also a member of the GEV class, though we will reserve the use of the term "GEV class" to models that constitute generalizations of the MNL model.

The general structure of the GEV class of models was derived by McFadden (1978) from the random utility maximization hypothesis, and generalized by Ben-Akiva and Francois (1983). Several specific GEV structures have been formulated and applied within the GEV class, including the Nested Logit (NL) model (Williams, 1977; McFadden, 1978; Daly and Zachary, 1978), the Paired Combinatorial Logit (PCL) model (Chu, 1990; Koppelman and Wen, 2000), the Cross-Nested Logit (CNL) model (Vovsha, 1997), the Ordered GEV (OGEV) model (Small, 1987), the Multinomial Logit-Ordered GEV (MNL-OGEV) model (Bhat, 1998a), the ordered GEV-nested logit (OGEV-NL) model (Whelan *et al.*, 2002) and the Product Differentiation Logit (PDL) model (Breshanan *et al.*, 1997). More recently, Wen and Koppelman (2001) proposed a general GEV model structure, which they referred to as the Generalized Nested Logit (GNL) model. Swait (2001), independently, proposed a similar structure, which he refers to as the choice set Generation Logit (GenL) model; Swait's derivation of the GenL model is motivated from the concept of latent choice sets of individuals, while Wen and Koppelman's derivation of the GNL model is motivated from the perspective of flexible substitution patterns across alternatives. Wen and Koppelman (2001) illustrate the general nature of the GNL model formulation by deriving the other GEV model structures mentioned earlier as special restrictive cases of the GNL model or as approximations to restricted versions of the GNL model. Swait (2001) presents a network representation for the GenL model, which also applies to the GNL model.

Researchers are not restricted to the GEV structures identified above, and can generate new GEV model structures customized to their specific empirical situation. In fact, only a handful of possible GEV model structures appear to have been implemented, and there are likely to be several, yet undiscovered, model structures within the GEV class. For example, Karlstrom (2001) has proposed a GEV model that is quite different in form from all other GEV models derived in the past. Also, Bierlaire (2002) proposed the network GEV structure which provides a high degree of flexibility in the formulation of GEV models.

Of course, GEV models, while allowing flexibility in substitution patterns, can also entail the estimation of a substantial number of dissimilarity and allocation parameters. The net result is that the analyst will have to impose informed restrictions on these GEV models, customized to the application context under investigation.

An important point to note here is that GEV models are consistent with utility maximization only under rather strict conditions on model structure parameters. The origin of these restrictions can be traced back to the requirement that the variance of the joint alternatives be identical in the GEV models. Also, GEV models do not relax assumptions related to taste homogeneity in response to an attribute (such as travel time or cost in a mode choice model) due to unobserved decision-maker characteristics, and cannot be applied to panel data with temporal correlation in unobserved factors within the choices of the same decision-making agent. However, it is indeed refreshing to note the renewed interest and focus on GEV models today, since such models do offer computational tractability, and provide a theoretically sound measure for benefit valuation.

12.3 The MMNL Class of Models

The MMNL class of models, like the GEV class of models, generalizes the MNL model. However, unlike the closed form of the GEV class, the MNL class involves the analytically intractable integration of the multinomial logit formula over the distribution of unobserved random parameters. It takes the structure shown below:

$$P_{qi}(\theta) = \int_{-\infty}^{+\infty} L_{qi}(\beta) f(\beta | \theta) d(\beta), \quad \text{where} \quad 12.1$$

$$L_{qi}(\beta) = \frac{e^{\beta' x_{qi}}}{\sum_j e^{\beta' x_{qj}}}.$$

P_{qi} is the probability that individual q chooses alternative i , x_{qi} is a vector of observed variables specific to individual q and alternative i , β represents parameters which are random realizations from a density function $f(\cdot)$, and θ is a vector of underlying moment parameters characterizing $f(\cdot)$.

The first known applications of the mixed logit structure of Equation 12.1 appear to have been by Boyd and Mellman (1980) and Cardell and Dunbar (1980). However, these were not individual-level models and, consequently, the integration inherent in the mixed logit

formulation had to be evaluated only once for the entire market. Train (1986) and Ben-Akiva *et al.* (1993) applied the mixed logit to customer-level data, but considered only one or two random coefficients in their specifications. Thus, they were able to use quadrature techniques for estimation. The first applications to realize the full potential of mixed logit by allowing several random coefficients simultaneously include Revelt and Train (1998) and Bhat (1998b), both of which were originally completed in the early 1996 and exploited the advances in simulation methods (for a detailed discussion of these recent advances, please see Train (2003) and Bhat (2005)).

The MMNL model structure of Equation 12.1 can be motivated from two very different (but formally equivalent) perspectives (see Bhat, 2000a). Specifically, a MMNL structure may be generated from an intrinsic motivation to allow flexible substitution patterns across alternatives (error-components structure) or from a need to accommodate unobserved heterogeneity across individuals in their sensitivity to observed exogenous variables (random-coefficients structure) or a combination of the two. Examples of the error-components motivation in the literature include Brownstone and Train (1999), Bhat (1998c), Jong *et al.* (2002a and b), Whelan *et al.* (2002), and Batley *et al.* (2001a and b). The reader is also referred to the work of Walker and her colleagues (Ben-Akiva *et al.*, 2001; Walker, 2002) and Munizaga and Alvarez-Daziano (2002) for important identification issues in the context of the error components MMNL model. Examples of the random-coefficients structure include Revelt and Train (1998), Bhat, (2000b), Hensher (2001), and Rizzi and Ortúzar (2003).

A normal distribution is assumed for the density function $f(\cdot)$ in Equation 12.1 when an error-components structure forms the basis for the MMNL model. However, while a normal distribution remains the most common assumption for the density function $f(\cdot)$ for a random-coefficients structure, other density functions may be more appropriate. For example, a log-normal distribution may be used if, from a theoretical perspective, an element of β has to take the same sign for every individual (such as a negative coefficient on the travel cost parameter in a travel mode choice model). Other distributions that have been used in the literature include triangular and uniform distributions (see Revelt and Train, 2000; Train, 2001; Hensher and Greene, 2003) and the Rayleigh distribution (Siikamaki and Layton, 2001). The triangular and

uniform distributions have the nice property that they are bounded on both sides, thus precluding the possibility of very high positive or negative coefficients for some decision-makers as would be the case if normal or log-normal distributions are used. By constraining the mean and spread to be the same, the triangular and uniform distributions can also be customized to cases where all decision-makers should have the same sign for one or more coefficients. The Rayleigh distribution, like the lognormal distribution, assures the same sign of coefficients for all decision-makers.⁵⁰

The MMNL class of models can approximate any discrete choice model derived from random utility maximization (including the multinomial probit) as closely as one pleases (see McFadden and Train, 2000) subject to computation limits on the number of points used to represent the mixture distributions. The MMNL model structure is also conceptually appealing and easy to understand since it is the familiar MNL model mixed with the multivariate distribution (generally multivariate normal) of the random parameters (see Hensher and Greene, 2003). In the context of relaxing the IID error structure of the MNL, the MMNL model represents a computationally efficient structure when the number of error components (or factors) needed to generate the desired error covariance structure across alternatives is much smaller than the number of alternatives (see Bhat, 2003a,b). The MMNL model structure also serves as a comprehensive framework for relaxing both the IID error structure as well as the response homogeneity assumption.

A few notes are in order here about the MMNL model vis-à-vis the MNP model. First, both these models are very flexible in the sense of being able to capture random taste variations and flexible substitution patterns. Second, both these models are able to capture temporal correlation over time, as would normally be the case with panel data. Third, the MMNL model is able to accommodate non-normal distributions for random coefficients, while the MNP model can handle only normal distributions. Fourth, researchers and practitioners familiar with the traditional MNL model might find it conceptually easier to understand the structure of the MMNL model compared to the MNP. Fifth, both the MMNL and MNP model, in general,

⁵⁰ The reader is referred to Hess and Axhausen (2005) for a review of alternative distribution forms and the ability of these distributed forms to approximate several different types of true distributional forms.

require the use of simulators to estimate the multidimensional integrals in the likelihood function (the reader is referred to Bhat, 2003a, for a detailed discussion of the advances in simulation methods that have made it very practical to estimate the MMNL and MNP models). Sixth, the MMNL model can be viewed as arising from the use of a logit-smoothed Accept-Reject (AR) simulator for an MNP model (see Bhat 2000c, and Train 2003; page 124). Seventh, the simulation techniques for the MMNL model are conceptually simple, and straightforward to code. They involve simultaneous draws from the appropriate density function with unrestricted ranges for all alternatives. Overall, the MMNL model is very appealing and broad in scope, and there appears to be little reason to prefer the MNP model over the MMNL model subject to resolution of identification problems still under study (Walker 2002). However, there is at least one exception to this general rule, corresponding to the case of normally distributed random taste coefficients. Specifically, if the number of normally distributed random coefficients is substantially more than the number of alternatives, the MNP model offers advantages because the dimensionality is of the order of the number of alternatives (in the MMNL, the dimensionality is of the order of the number of random coefficients)⁵¹.

12.4 The Mixed GEV Class of Models

The MMNL class of models is very general in structure and can accommodate both relaxations of the IID assumption as well as unobserved response homogeneity within a simple unifying framework. Consequently, the need to consider a mixed GEV class may appear unnecessary. However, there are instances when substantial computational efficiency gains may be achieved using a MGEV structure. Consider, for instance, Bhat and Guo's (2004) model for household residential location choice. It is possible, if not very likely, that the utility of spatial units that are close to each other will be correlated due to common unobserved spatial elements. A common specification in the spatial analysis literature for capturing such spatial correlation is to allow contiguous alternatives to be correlated. In the MMNL structure, such a correlation structure may

⁵¹ The reader is also referred to Munizaga and Alvarez-Daziano (2002) for a detailed discussion comparing the MMNL model with the nested logit and MNP models.

be imposed through the specification of a multivariate MNP-like error structure, which will then require multidimensional integration of the order of the number of spatial units (see Bolduc *et al.*, 1996). On the other hand, a carefully specified GEV model can accommodate the spatial correlation structure within a closed-form formulation.⁵² However, the GEV model structure of Bhat and Guo cannot accommodate unobserved random heterogeneity across individuals. One could superimpose a mixing distribution over the GEV model structure to accommodate such random coefficients, leading to a parsimonious and powerful MGEV structure. Thus, in a case with 1000 spatial units (or zones), the MMNL model would entail a multidimensional integration of the order of 1000 plus the number of random coefficients, while the MGEV model involves multidimensional integration only of the order of the number of random coefficients (a reduction of dimensionality of the order of 1000!).

In addition to computational efficiency gains, there is another more basic reason to prefer the MGEV class of models when possible over the MMNL class of models. This is related to the fact that closed-form analytic structures should be used whenever feasible, because they are always more accurate than the simulation evaluation of analytically intractable structures (see Train, 2003; pg. 191). In this regard, superimposing a mixing structure to accommodate random coefficients over a closed form analytic structure that accommodates a particular desired inter-alternative error correlation structure represents a powerful approach to capture random taste variations and complex substitution patterns.

Clearly, there are valuable gains to be achieved by combining the state-of-the-art developments in closed-form GEV models with the state-of-the-art developments in open-form mixed distribution models. With the recent advances in simulation techniques, there appears to be a feeling among some discrete choice modelers that there is no need for any further consideration of closed-form structures for capturing correlation patterns. But, as Bhat and Guo (2004) have demonstrated in their paper, the developments in GEV-based structures and open-form mixed models are not as mutually exclusive as may be the impression in the field; rather

⁵² The GEV structure used by Bhat and Guo is a restricted version of the GNL model proposed by Wen and Koppelman. Specifically, the GEV structure takes the form of a paired GNL (PGNL) model with equal dissimilarity parameters across all paired nests (each paired nest includes a spatial unit and one of its adjacent spatial units).

these developments can, and are, synergistic, enabling the estimation of model structures that cannot be estimated using GEV structures alone or cannot be efficiently estimated (from a computational standpoint) using a mixed multinomial logit structure.

12.5 Summary

The field of discrete choice has seen a quantum jump in recent years. There is a sense today of absolute control over the behavioral structures one wants to estimate in empirical contexts and renewed excitement in the field, thanks to recent conceptual and simulation developments. However, analysts need to be careful not to get carried away with these new developments in choice modeling. The fundamental idea of discrete choice models will always continue to be the identification of systematic variations in the population. The advanced methods presented in this Chapter should be viewed as formulations that recognize the inevitable presence of unobserved heterogeneity across individuals and/or interactions among unobserved components affecting the utility of alternatives even after adopting the best systematic specifications there can be. In fact, a valuable contribution of recent developments in the field is precisely that they enable the confluence of careful structural specification with the ability to accommodate flexible substitution patterns and unobserved heterogeneity profiles.

References

- Ben-Akiva, M. and S.R. Lerman (1985) Discrete Choice Analysis: Theory and Application to Travel Demand, The MIT Press, Cambridge.
- Ben-Akiva, M. and D. Bolduc (1996) Multinomial probit with a logit kernel and a general parametric specification of the covariance structure, The 3rd Invitational Choice Symposium, Columbia University.
- Bhat, C.R., A. Govindarajan and V. Pulugurta (1998) "Disaggregate attraction-end choice modeling: formulation and empirical analysis," Transportation Research Record 1645, 60-68.
- Bhat, C.R. and V. Pulugurta (1998) "A comparison of two alternative behavioral mechanisms for household auto ownership decisions," Transportation Research, 32B, 10-24.
- Bhat, C.R. (1995) "A heteroscedastic extreme-value model of intercity mode choice," Transportation Research, 29B, 6, 471-483.
- Bhat, C.R. (1997b) "An endogenous segmentation mode choice model with an application to intercity travel," Transportation Science, 31, 34-48.
- Bhat, C.R. (1997c) "A nested logit model with covariance heterogeneity," Transportation Research, 31B, 11-21.
- Bhat, C.R. (1998a) "An analysis of travel mode and departure time choice for urban shopping trips," Transportation Research, 32B, 387-400.
- Bhat, C.R. (1998b) "Accommodating flexible substitution patterns in multidimensional choice modeling: formulation and application to travel mode and departure time choice," Transportation Research, 32B, 455-466.
- Bhat, C.R. (1998c) "Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling," Transportation Research, 32A, 495-507.
- Bhat, C.R. (2000) "Flexible Model Structures for Discrete Choice," Handbook of Transportation Modeling, Hensher, D.A. and K. Button (eds.), 71-90.
- Bierlaire, M. (2002). "The network GEV model," Proceedings, The 2nd Swiss Transportation Research Conference, Ascona, Switzerland.

- Bierlaire, M. and Y. Vandevyvere (1995) HieLow: The Interactive User's Guide.
Transportation Research Group, FUNDP, Namur
- Brownstone, D. and K. Train (1999) "Forecasting new product penetration with flexible substitution patterns," Journal of Econometrics, 89, 109.
- Bunch, D.S. and R. Kitamura (1991) Probit model estimation revisited: trinomial models of household car ownership, University of California Transportation Center, University of California, Berkeley, CA.
- Börsch-Supan, A. (1987) Econometric Analysis of Discrete Choice, Lecture Notes in Economics and Mathematical Systems, 296, Springer-Verlag, Berlin, Germany, 202-211.
- Borsch-Supan, A. (1990) HLOGIT Documentation.
- Bowman, J. (2004). Heteroscedastic nested logit kernel (or mixed logit) models for large multidimensional choice problems: Identification and estimation. Proceedings, 83rd Annual Meeting, Transportation Research Board, Washington, D.C.
- Bucklin, R.E., S. Gupta and S. Han (1995) "A brand's eye view of response segmentation in consumer brand choice behavior," Journal of Marketing Research, 32, 66-74.
- Cascetta, E., F. Russo and F. Viola (2002) "A Model of route perception in urban road networks," Transportation Research, 36A, 577-592.
- Chamberlain, G. (1980) "Analysis of covariance with qualitative data," Review of Economic Studies, 47, 225-238.
- Chiou, L. & Walker, J (2006). Identification and estimation of mixed logit models under simulation methods. Paper presented at the 85th Annual Meeting of the Transportation Research Board, Washington D.C.
- Chu, C. (1989) "A paired combinatorial logit model for travel demand analysis," Proceedings of the Fifth World Conference on Transportation Research, 295-309, Ventura, CA.
- Corres, S., V.A. Hajivassiliou and Y.M. Ioannides (1993) An empirical analysis of the dynamics of qualitative decisions of firms, Economic Working Papers Series, Yale University.
- Cramer, J.S. (1991) The Logit Model: An Introduction for Economists, Edward Arnold, London, England.

Daganzo, C. (1979) Multinomial Probit: The Theory and its Application to Demand Forecasting, Academic Press, New York, NY.

Econometric Software Inc., LIMDEP V7.0, 1996.

Erhardt, G.D., Koppelman, F.S., Freedman, J., Davidson, W.A., Mullins, A. (2004) "Modeling the Choice to use Toll and High-Occupancy Vehicle Facilities," Transportation Research Record 1854, 135-143.

Evers, G.H.M. (1990) "The residential location and workplace choice: a nested multinomial logit model," in Spatial Choices and Processes, eds: M.M. Fischer, P. Nijkamp and Y.Y. Papageorgiou, Elsevier Science Publishers, B.V. (North-Holland), 313-329.

Forinash C.V. and Koppelman F.S. (1993) "Application and interpretation of nested logit models of intercity mode choice," Transportation Research Record 1413, 98-106.

Gliebe, J.P., F.S. Koppelman and A. Ziliaskopoulos (1999) "Route choice using a paired combinatorial logit model," Proceedings, 78th meeting of the Transportation Research Board, Washington, D.C.

Gliebe, J.P. and F.S. Koppelman (2002) "A Model of Joint Activity Participation," Transportation, 29, 49-72.

Grayson, Alan (1979) "Disaggregate Model of mode choice in intercity travel," Transportation Research Record 835, 36-42.

Greene, William H. (2000) Econometric Analysis, 4th ed., Prentice Hall, Upper Saddle River, NJ.

Hague Consulting Group (1995), ALOGIT: Users Guide, Version 3.8, 1995.

Hensher, D.A., N.C. Smith, F.W. Milthorpe and P.O. Barnard (1992) "Dimensions of Automobile Demand: A Longitudinal Study of Household Automobile Ownership and Use," Studies in Regional Science and Urban Economics, 22, Elsevier Science Publishers, Amsterdam.

Hensher, D.A. (2005) The signs of the times: Imposing a globally signed condition on willingness to pay distributions. Transportation, 33, 5.

Hess, S., Bierlaire, M. & Polak, J.W. (2005). Estimation of value of travel-time savings using Mixed Logit models, Transportation Research, Part A, 2005, Pages 221-236.

- Horowitz, J.L., F.S. Koppelman and S.R. Lerman (1986) A self-instructing course in disaggregate mode choice modeling, Final report, prepared for the U.S. Department of Transportation, University Research and Training Program, Washington, D.C.
- Iglesias, M. (1997) Estimation of home-based social/recreational mode choice models, technical report HBSRMC #1, Planning Section, Metropolitan Transportation Commission, 101 Eight Street, Oakland, California.
- Kalyanam, K. and D.S. Putler (1997) "Incorporating demographic variables in brand choice models: an indivisible alternatives framework," Marketing Science, 16, 166-181.
- Koppelman, F.S. (1975) Travel Prediction with Models of Individualistic Choice Behavior, Ph.D. Dissertation, Department of Civil Engineering, MIT, Cambridge, MA.
- Koppelman, F.S. (1989) "Multidimensional Model System for Intercity Travel Choice Behavior," Transportation Research Record 1241, 1-8.
- Koppelman, F.S. and C-H. Wen (1998) "Nested logit models: Which are you using?" Transportation Research Record 1645, 1-7.
- Koppelman, F.S. and C-H Wen (2000) "The paired combinatorial logit model: properties, estimation and application," Transportation Research, 34B, 75-89.
- Koppelman, F.S. and V. Sethi (2000) "Closed Form Discrete Choice Models," Handbook of Transportation Modeling, Hensher, D.A. and K. Button (eds.), 211-227.
- KPMG Peat Marwick in association with ICF Kaiser Engineers, Inc., Midwest System Sciences, Resource Systems Group, Comsis Corporation and Transportation Consulting Group (1993) Florida High Speed and Intercity Rail Market and Ridership Study: Final Report, submitted to Florida Department of Transportation, July.
- Lancaster, K. (1971) Consumer Demand: A New Approach, New York, Columbia University Press.
- Lam, S-H (1991) Multinomial probit model estimation: computational procedures and applications, Unpublished Ph.D. dissertation, Department of Civil Engineering, The University of Texas at Austin.
- Lam, S-H. and H.S. Mahmassani (1991) "Multinomial probit model estimation: computational procedures and applications," in Methods for Understanding Travel Behavior in the 1990s, Proceedings of the International Association of Travel Behavior, 229-242.

- Lawton, K.T. (1989) Travel forecasting methodology report, Metropolitan Service District, Portland, OR.
- Marshall, N.L. and K.Q. Ballard (1998) "New distribution and mode choice models for the Chicago region," presented at the Annual TRB Meeting, Washington, D.C., January.
- McFadden, D. and K. Train (2000) "Mixed MNL models for discrete response," Journal of Applied Econometrics, 15, 447-470.
- McMillen, D.P. (1995) "Spatial effects in probit models: a Monte Carlo investigation," in L. Anselin and R.J.G.M. Florax (editors) New Directions in Spatial Econometrics, Springer-Verlag, New York.
- Ortuzar, J. de D. and L.G. Willumsen (1997), Modelling Transport, John Wiley & Sons, New York, NY.
- Prousaloglou, K.E. and Koppelman, F.S. (1999) "The Choice of Carrier, Flight and Fare Class," Journal of Air Transport Management, 5, 193-201.
- Purvis, C.L. (1996) Disaggregate estimation and validation of home-to-work departure time choice model, technical report HBWDT #1, Planning Section, Metropolitan Transportation Commission, 101 Eight Street, Oakland, California.
- Purvis, C.L. (1997) Disaggregate estimation and validation of a home-based work mode choice model, technical report HBWMC #2, Planning Section, Metropolitan Transportation Commission, 101 Eight Street, Oakland, California.
- Recker, W.W. (1995) "Discrete choice with an oddball alternative," Transportation Research, 29B, 201-212.
- Sermons, M.W. and F.S. Koppelman (1998) "Factor Analytic Approach to Incorporating Systematic Taste Variation into Models of Residential Mode Choice," Transportation Research Record 1617, 194-202.
- Small, K.A. (1987) "A discrete choice model for ordered alternatives," Econometrica, 55(2), 409-424.
- Steckel, J.H. and W.R. Vanhonacker (1988) "A heterogeneous conditional logit model of choice," Journal of Business and Economic Statistics, 6, 391-398.

- Swait, J. and W. Adamowicz (2001) "Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer," Organizational Behavior and Human Decision Processes, V. 86, N. 2, Pages 141-167
- Swait, J. and E.C. Stacey (1996) Consumer brand assessment and assessment confidence in models of longitudinal choice behavior, presented at the 1996 INFORMS Marketing Science Conference, Gainesville, FL, March 7-10.
- Train, K. (1986) Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand, The MIT Press, Cambridge, MA.
- Train, K. (1998) "Recreation demand models with taste differences over people," Land Economics, 74, 230-240.
- Train, K. & Sonnier, G. (2004). Mixed logit with bounded distributions of correlated partworths, in A. Alberini & R. Scarpa, Eds., *Applications of Simulation Methods in Environmental and resource Economics*, Kluwer Academic Publishers, Boston, MA.
- Vovsha, P. (1997) "Application of cross-nested logit model to mode choice in Tel-Aviv, Israel, metropolitan area," Transportation Research Record 1607, 6-15.
- Waddell, P. (1993) "Exogenous workplace choice in residential location models; Is the assumption valid?" Geographical Analysis, 25, 65-82.
- Wen, C-H and Koppelman, F.S. (1999), "Integrated Model System of Stop Generation and Tour Formation for Analysis of Activity and Travel Patterns," Transportation Research Record 1676, 136-144.
- White, E.H. and Company, Inc. (1991) 1990 Bay Area Travel Survey: Final Report, Metropolitan Transportation Commission, Oakland, CA.
- Williams, H. C. W. L. (1977) "On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit," Environment and Planning, Part A, V.9, 285-344.
- Yai, T., S. Iwakura and S. Morichi (1997) "Multinomial robit with structured covariance for route choice behavior," Transportation Research, 31B, 195-207.

Appendix A: ALOGIT, LIMDEP and ELM

The command files and estimation results from ALOGIT and LIMDEP for the base model specification reported in Table 5-2 are presented in Figure A.1 through Figure A.4. The estimation results from ELM are reported in Figure A.7⁵³. The outputs from these and other software packages typically include, at least, the following estimation results:

- Variable names, parameter estimates, standard errors of these estimates and the corresponding t-statistics for each variable/parameter;
- Log-likelihood values at zero (equal probability model), constants only (market shares model) and at convergence and
- Rho-Squared and other indicators of goodness of fit.

ALOGIT, LIMDEP and ELM also provide additional information either as part of a general log file or by optional request. This information varies among these and other software packages. However, it should be noted that two of the important outputs, the log-likelihood at zero and/or the log-likelihood at constants only (market shares) may be based on simplifying assumption that do not apply in all cases. In particular, it is not uncommon for software to compute these values based on the assumption that all alternatives are available to all users. Since this may not be the case, the user must be careful to validate this information. In any case, accurate estimates of these measures can be obtained simply by estimating models with no variables (or all variables constrained to zero) or with alternative specific constants only.

Effectively, most software packages produce essentially the same information but in different formats. For purposes of increased clarity and to simplify comparisons between models with different specifications, the estimation results for the base model and two reference models (zero coefficients and constants only) may be transformed to a format in which parameter estimates and their t-statistics are grouped by variable and model goodness of fit

⁵³ No command file is provided for ELM as the primary input is through a Graphic User Interface (GUI). However, the interface produces an intermediate command file that can be used to direct model estimation if desired by the user.

statistics are grouped together (Table 5-2) to facilitate comparison among models. This output format is standard for models estimated in a single batch in ELM.

Further information about each of these software packages can be obtained by going to their websites as follows:

- ALOGIT, http://www.hpgholding.nl/software/alo_intr.htm
- LIMDEP, <http://www.limdep.com/>
- ELM, <http://www.elm-works.com/>


```

$title MNL Model 1: Cost,Time,Income
$subtitle @SFM1.alo

$gen.stats all

$estimate

$nest root() da sr2 sr3 transt bik wak

01 Cost
02 Ivtt
03 Ovtt
04 Tott
05 IncSh2
06 IncSh3
07 IncTrn
08 IncBik
09 IncWlk
20 Sh2Cnst
30 Sh3Cnst
40 TrnCnst
50 BikCnst
60 WlkCnst

file(name = D:\KK\sf.dat, handle = sf)

Id
Persid
WrkZone
HmZone
AutCost
Sh2Cost
Sh3Cost
TrnCnst
AutTott
Sh2Tott
Sh3Tott
TrnTott
BikTott
WlkTott
DAAlone
ShRide2
ShRide3
ShRide
Transit
Bike
Walk
Income
alt
Choice

avail(da) = ifgt(DA_Av, 0)
avail(sr2) = ifgt(Sh2Tott, 0)
avail(sr3) = ifgt(Sh3Tott, 0)
avail(transt) = ifgt(Trn_Av, 0)
avail(bik) = ifgt(Bike_Av, 0)
avail(wak) = ifgt(Walk_Av, 0)

CHOSEN = Choice
choice = recode(CHOSEN, da, sr2, sr3, transt, bik, wak)

-Utility Functions

U(da) = p01*AutCost + p04*AutTott
U(sr2) = p20 + p01*Sh2Cost + p04*Sh2Tott + p05*Income
U(sr3) = p30 + p01*Sh3Cost + p04*Sh3Tott + p06*Income
U(transt) = p40 + p01*TrnCnst + p04*TrnTott + p07*Income
U(bik) = p50 + p04*BikTott + p08*Income
U(wak) = p60 + p04*WlkTott + p09*Income

```

Figure A.1 ALOGIT Input Command File

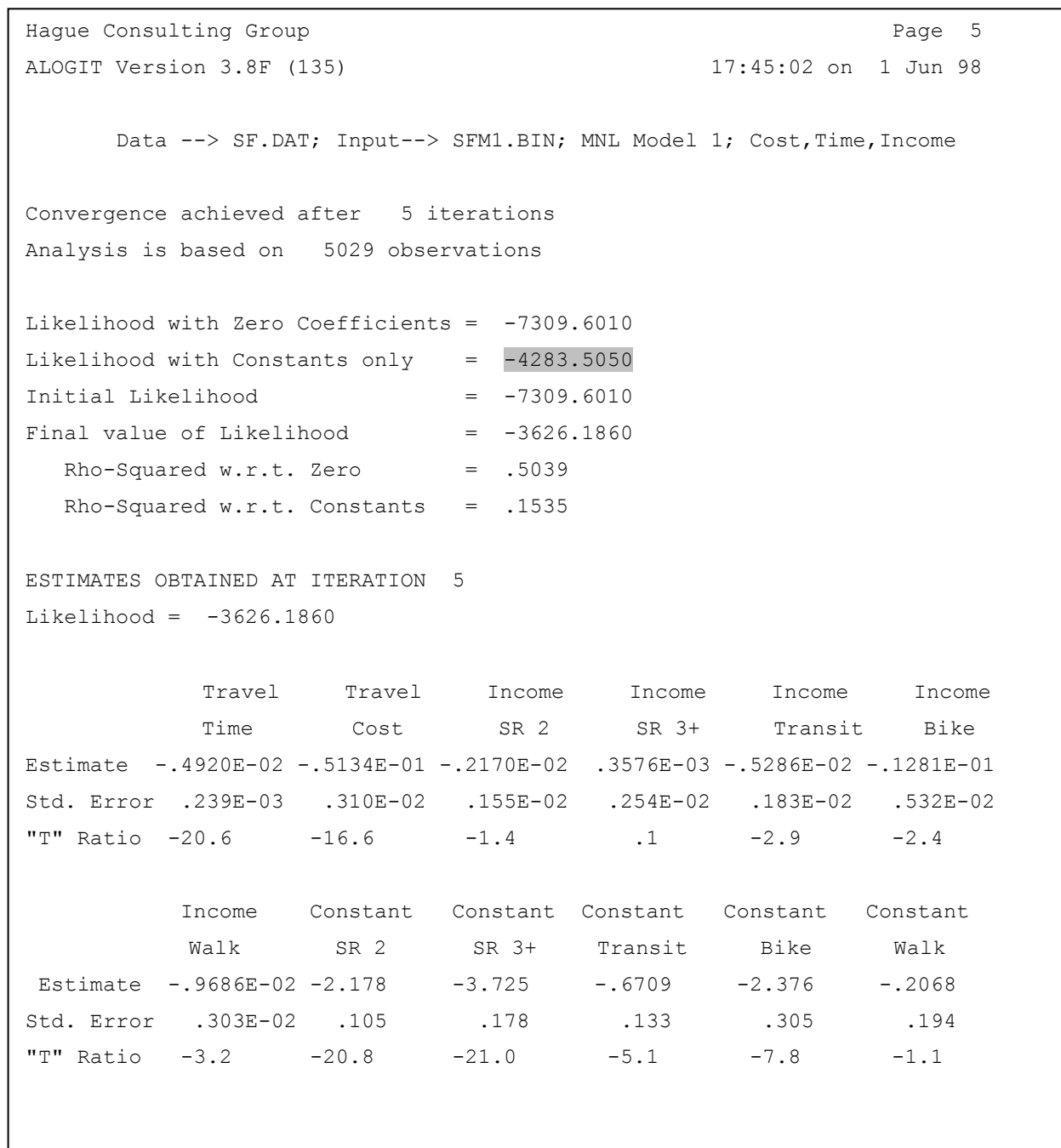


Figure A.2 Estimation Results for Basic Model Specification using ALOGIT⁵⁴

⁵⁴The "Likelihood with Constants Only" is calculated as if all alternatives are available to all cases this version of ALOGIT. As a result, the "Rho-Squared w.r.t. Constants" is incorrect.

```

? Open data file
read ; nobs = 22033 ; nvar = 37;
  file = SFLIM.PRN
  ; names = HHId, PerId, GrpSize, WkZone, HmZone, Dist, RsPopDen,
    RsEmpDen, WkPopDen, WkEmpDen, VehAvDum, FemDum, Age,
    DrLicDum, NonCaDum, NumVeh, HHSIZE, HHInc, FamType,
    NumEmpHH, NumAdlt, HHOwnDum, NmLt5, Nm5to11, Nm12to16,
    WkCCBD, WkNCCBD, CorReDis, VehbyWrk, Alt,
    Cost, IVTT, OVTT, TVTT, AltNum, Chosen, NumAlts $

? Open output file
open ; output = SFLIMRUN.OUT $

? Create alternative specific constants
sample ; all $
create ; if (altnum = 2) Sh2Cnst = 1 $
create ; if (altnum = 3) Sh3Cnst = 1 $
create ; if (altnum = 4) TrnCnst = 1 $
create ; if (altnum = 5) BikCnst = 1 $
create ; if (altnum = 6) WlkCnst = 1 $

? Income as an alternative sepcific variable.
create ; if (altnum = 2) IncSh2 = HHInc $
create ; if (altnum = 3) IncSh3 = HHInc $
create ; if (altnum = 4) IncTrn = HHInc $
create ; if (altnum = 5) IncBik = HHInc $
create ; if (altnum = 6) IncWlk = HHInc $

? ***** Model Estimation *****
? Compute log-likelihood at zero model
title ; *** Model 0: No coefficient model *** $
samp
le ; all $
nlogit ; LHS = Chosen, NumAlts, AltNum; maxit = 0
  ; RHS = Sh2Cnst, Sh3Cnst, TrnCnst, BikCnst, WlkCnst
  ; choices = DrAlone, ShRide2, ShRide3, Transit, Bike, Walk $
? Compute log likelihood at market share
title ; *** Model C: Constants only model--DA reference *** $
sample ; all $
nlogit ; LHS = Chosen, NumAlts, AltNum
  ; RHS = Sh2Cnst, Sh3Cnst, TrnCnst, BikCnst, WlkCnst
  ; choices = DrAlone, ShRide2, ShRide3, Transit, Bike, Walk $
? MNL
title ; *** Base Model *** $
sample ; all $
nlogit ; LHS = Chosen, NumAlts, AltNum
  ; RHS = Sh2Cnst, Sh3Cnst, TrnCnst, BikCnst, WlkCnst,
    Tvtt, Cost, IncSh2, IncSh3, IncTrn,
    IncBik, IncWlk
  ; choices = DrAlone, ShRide2, ShRide3, Transit, Bike, Walk $

```

Figure A.3 LIMDEP Input Command File

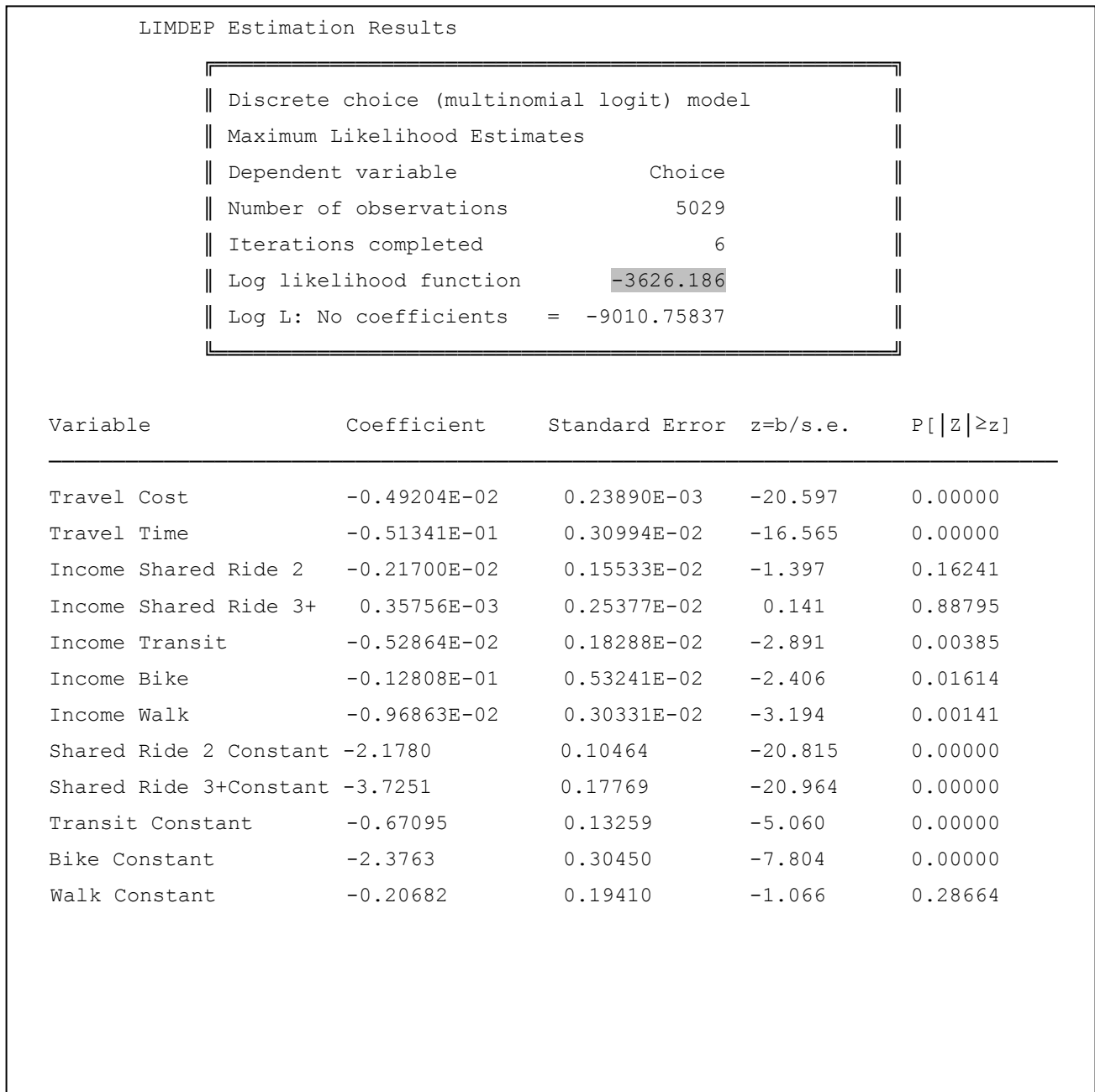


Figure A.4 Estimation Results for Basic Model Specification using LIMDEP⁵⁵

⁵⁵ The “Log L: No Coefficients” value in this version of Limdep (NLogit) is calculated incorrectly.

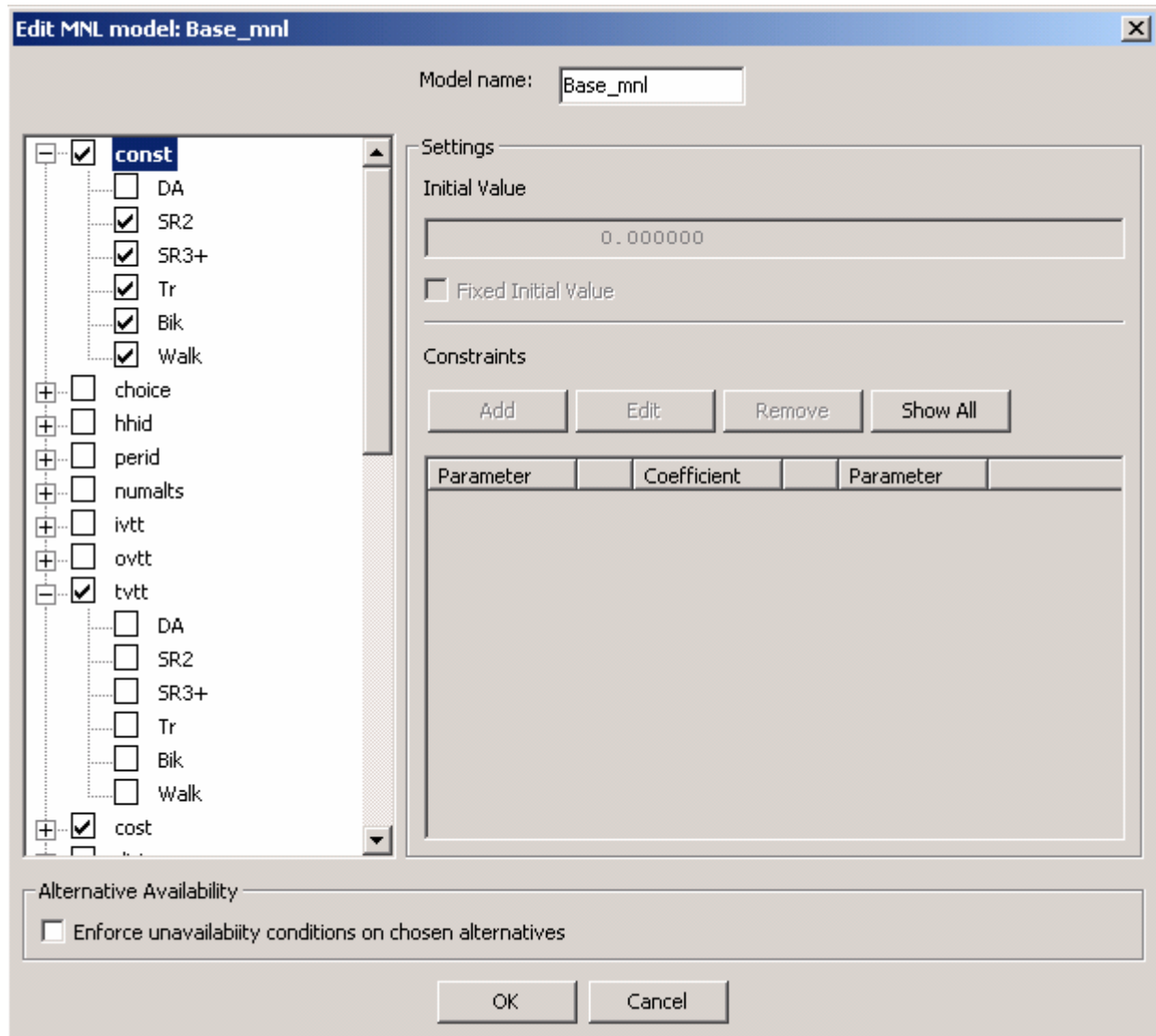


Figure A.5 ELM Model Specification⁵⁶

⁵⁶ ELM Model Specification Screen allows selection of variables (generic as shown for the constants or alternative specific as shown for total travel time) to be included in a model, selection of starting values, imposition of parameter constraints and imposition of ratios among parameters.

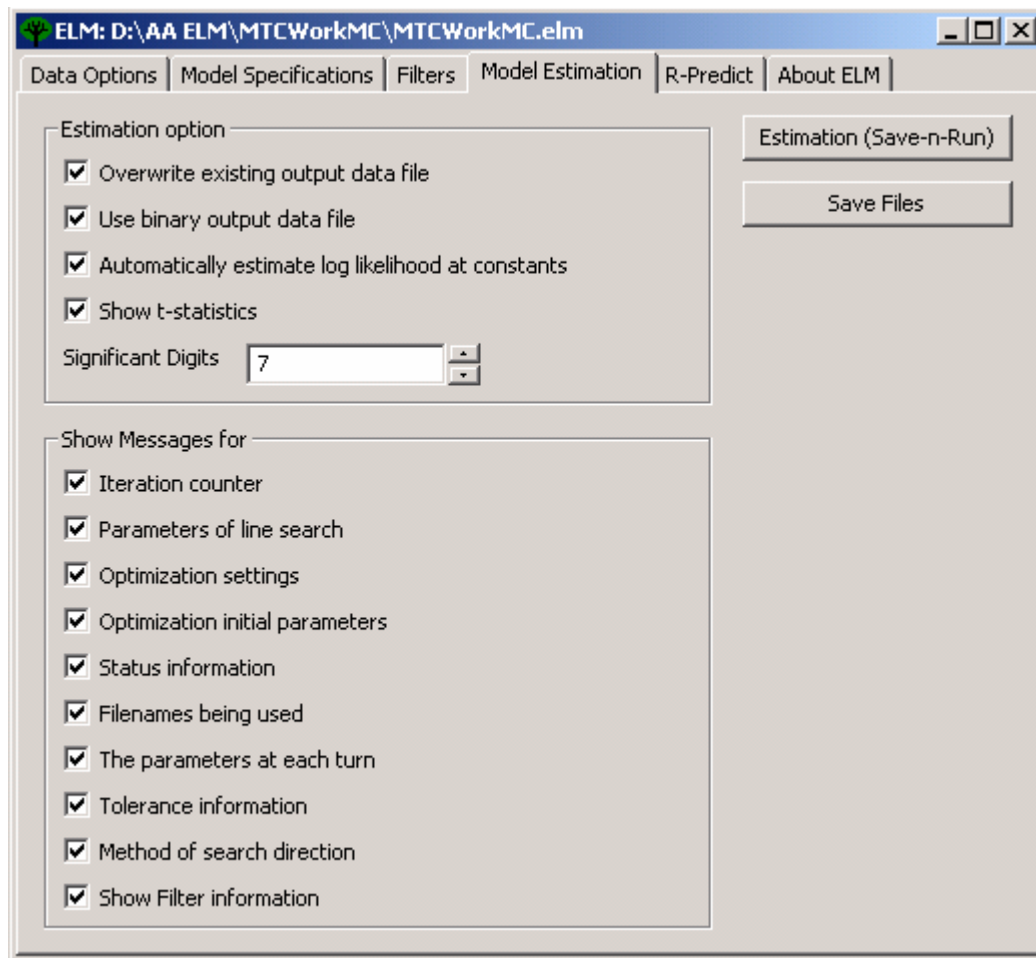


Figure A.6 ELM Model Estimation⁵⁷

⁵⁷ ELM Model Estimation Screen allows specification of a range of estimation and output options.

___ Summary of Models ___				
Base Model				
Log Likelihood	-3626.186			
LL @ Constants	-4132.916			
->Rho Squared @ Constants		0.1226		
Log Like @ Zero	-7309.601			
->Rho Squared @ Zero		0.5039		
Variable	Parameter	Std. Error	t Statistic	Signif.
tvtt	-5.134E-02	3.099E-03	-16.56	<0.001
cost	-4.921E-03	2.389E-04	-20.60	<0.001
hhinc*Shared_Ride_2	-2.170E-03	1.553E-03	-1.40	0.081
hhinc*Shared_Ride_3+	3.574E-04	2.538E-03	0.14	0.444
hhinc*Transit	-5.286E-03	1.829E-03	-2.89	0.002
hhinc*Bike	-1.281E-02	5.324E-03	-2.41	0.008
hhinc*Walk	-9.686E-03	3.033E-03	-3.19	0.001
*Shared_Ride_2	-2.178E+00	1.046E-01	-20.82	<0.001
*Shared_Ride_3+	-3.725E+00	1.777E-01	-20.96	<0.001
*Transit	-6.710E-01	1.326E-01	-5.06	<0.001
*Bike	-2.376E+00	3.045E-01	-7.80	<0.001
*Walk	-2.069E-01	1.941E-01	-1.07	0.143
Number of Cases	5029			
Alternative	Available	Chosen		
Drive_Alone	4755	3637		
Shared_Ride_2	5029	517		
Shared_Ride_3+	5029	161		
Transit	4003	498		
Bike	1738	50		
Walk	1479	166		

Figure A.7 ELM Estimation Results Reported in Excel⁵⁸

⁵⁸ ELM Estimation Results for multiple models included in a run are reported in a single Excel file as well as model specific log files.

Appendix B: Example Matlab Files on CD

The CD accompanying this manual includes a copy of the manual, the estimation data for both work and shop/other trips, MATLAB code, commands, log files and reports (together with relevant documentation) associated with six example mode choice model estimations based on the 1990 household survey of the San Francisco bay area used for examples throughout this manual. Two of the six models are multinomial logit models for the home-based work trip purpose and correspond to Models 15 W and 17 W from CHAPTER 6. The other four are additional models for the home-based shop/other trip purpose, two multinomial logit models and two nested logit models. These models are different from those presented earlier in this manual.

The contents of the CD are summarized in Table B-1. This documentation, the Matlab code itself, and two compilations of the results (Reports.doc and Reports.xls) reside in the parent directory. The data, data dictionaries, command files, log (complete report on estimation iterations and extended details on the estimation results and output reports (summary of key features of each estimation) are stored in two separate sub-directories corresponding to the trip purposes.

Table B-1 Files / Directory Structure

LM_MATLAB_Ex		Parent Directory
	Documentation.doc	Documentation
	Logit46.m	MATLAB code
	Reports.doc	Reports in Word
	Reports.xls	Reports in Excel
SF_HBShO		Sub-directory
	Log of SFHBSHO 04 MNL L46.txt	
	Log of SFHBSHO 09 MNL L46.txt	
	Log of SFHBSHO 09 NL AC L46.txt	
	Log of SFHBSHO 09 NL ASCc2 L46.txt	
	MATLAB cmds.txt	
	Report SFHBSHO 04 MNL L46.txt	Tab-delimited
	Report SFHBSHO 09 MNL L46.txt	Tab-delimited
	Report SFHBSHO 09 NL AC L46.txt	Tab-delimited
	Report SFHBSHO 09 NL ASCc2 L46.txt	Tab-delimited
	SFHBSHO.sav	SPSS data

LM MATLAB Ex		Parent Directory
	SFHBSHOW5.mat	MATLAB data
	SFHBSHO_DataDictionary.xls	
SF HBW		Sub-directory
	Log of SFW15_MNL_L46.txt	
	Log of SFW17_MNL_L46.txt	
	MATLAB_cmds.txt	
	Report SFW15_MNL_L46.txt	Tab-delimited
	Report SFW17_MNL_L46.txt	Tab-delimited
	SF MTC Work MC Data.sav	SPSS data
	SFMTCWork6.mat	MATLAB data
	SFMTCWork_DataDictionary.xls	

The MATLAB code⁵⁹ used for the estimations includes its own documentation/instruction for use. These can be accessed from MATLAB by (first checking that the working directory includes the logit46.m file and then) typing the command “help logit46”. The MATLAB_cmds.txt files, which include the commands used to execute each of the six model estimations, can be used directly or can serve as examples for modifying the specification of any of the models.

The two data sets, corresponding to the two trip purposes are included in both SPSS and MATLAB binary format. The MATLAB files contain all the variables in the SPSS files, as well as manipulations of the variables (to make them alternative specific, for example) for purposes of model estimation. Each dataset is accompanied by a data dictionary in Excel format.

⁵⁹ The MATLAB code is open source and is not supported.