

Cambridge University Press
0521605776 - Applied Choice Analysis: A Primer
David A. Hensher, John M. Rose and William H. Greene
Frontmatter
[More information](#)

Applied Choice Analysis

A Primer

Almost without exception, everything human beings undertake involves a choice. In recent years, there has been a growing interest in the development and application of quantitative statistical methods to study choices made by individuals with the purpose of gaining a better understanding both of how choices are made and of forecasting future choice responses. In this primer, the authors provide an unintimidating introduction to the main techniques of choice analysis and include detail on themes such as data collection and preparation, model estimation and interpretation, and the design of choice experiments. A companion website to the book provides practice data sets and software to estimate the main discrete choice models such as multinomial logit, nested logit, and mixed logit. This primer will be an invaluable resource to students as well of immense value to consultants/professionals, researchers, and anyone else interested in choice analysis and modeling.

Companion website www.cambridge.org/0521605776

DAVID A. HENSHER is Director of the Institute of Transport Studies and Professor of Management in the Faculty of Economics and Business at the University of Sydney

JOHN M. ROSE is a Lecturer at the Institute of Transport Studies at the University of Sydney

WILLIAM H. GREENE is Professor of Economics and Entertainment and Media Faculty Fellow in the Department of Economics at the Stern School of Business, New York University

Cambridge University Press
0521605776 - Applied Choice Analysis: A Primer
David A. Hensher, John M. Rose and William H. Greene
Frontmatter
[More information](#)

Applied Choice Analysis

A Primer

David A. Hensher
The University of Sydney

John M. Rose
The University of Sydney

William H. Greene
New York University



Cambridge University Press
0521605776 - Applied Choice Analysis: A Primer
David A. Hensher, John M. Rose and William H. Greene
Frontmatter
[More information](#)

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521605779

© David A. Hensher, John M. Rose, William H. Greene 2005

This book is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2005

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this book is available from the British Library

ISBN-13 978-0-521-84426-0 hardback
ISBN-10 0-521-84426-6 hardback
ISBN-13 978-0-521-60577-9 paperback
ISBN-10 0-521-60577-6 paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this book, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

	<i>List of figures</i>	page xiii
	<i>List of tables</i>	xviii
	<i>Preface</i>	xxiii
Part I	Basic topics	
1	In the beginning	3
2	Basic notions of statistics	8
	2.1 Introduction	8
	2.2 Data	8
	2.2.1 The importance of understanding data	10
	2.3 A note on mathematical notation	10
	2.3.1 Summation	11
	2.3.2 Product	12
	2.4 Probability	12
	2.4.1 Relative frequencies	13
	2.4.2 Defining random variables	14
	2.4.3 Probability distribution functions	14
	2.4.4 Cumulative distribution functions	16
	2.4.5 Multivariate probability density functions	17
	2.4.6 The multivariate probability function	18
	2.4.7 Marginal probability density functions	21
	2.4.8 Conditional probability density functions	21
	2.4.9 Defining statistical independence	23
	2.5 Properties of random variables	23
	2.5.1 Expected value	25
	2.5.1.1 Properties of expected values	26
	2.5.2 Variance	28
	2.5.2.1 Properties of variance	28

▼

vi	Contents	
	2.5.3 Covariance	30
	2.5.3.1 Properties of covariance	31
	2.5.4 The variance–covariance matrix	32
	2.5.5 Correlation	33
	2.5.5.1 Properties of the correlation coefficient	34
	2.5.6 Correlation and variances	36
	2.6 Sample population statistics	36
	2.6.1 The sample mean	36
	2.6.2 The sample variance	38
	2.6.3 The sample covariance	38
	2.6.4 The sample correlation coefficient	39
	2.7 Sampling error and sampling distributions	39
	2.8 Hypothesis testing	41
	2.8.1 Defining the null and alternative hypotheses	42
	2.8.2 Selecting the test-statistic	44
	2.8.3 Significance of the test and alpha	45
	2.8.4 Performing the test	51
	2.8.5 Example hypothesis test: the one sample <i>t</i> -test	51
	2.9 Matrix algebra	52
	2.9.1 Transposition	53
	2.9.2 Matrix addition and subtraction	53
	2.9.3 Matrix multiplication by a scalar	54
	2.9.4 Matrix multiplication	54
	2.9.5 Determinants of matrices	55
	2.9.6 The identity matrix	56
	2.9.7 The inverse of a matrix	57
	2.9.8 Linear and quadratic forms	58
	2.9.9 Positive definite and negative definite matrices	59
	2.10 Conclusion	59
	Appendix 2A Measures of correlation or similarity	59
3	Choosing	62
	3.1 Introduction	62
	3.2 Individuals have preferences, and they count	63
	3.3 Using knowledge of preferences and constraints in choice analysis	71
	3.4 Setting up a behavioral choice rule	74
	3.5 Deriving a basic choice model	82
	3.6 Concluding overview	86
4	Paradigms of choice data	88
	4.1 Introduction	88
	4.2 Data consistent with choice	89
	4.3 Revealed preference data	92
	4.3.1 Choice-based sampling	95

4.4	Stated preference (or stated choice) data	96
4.5	Further comparisons	97
4.6	Why not use both RP and SP data?	98
4.7	Socio-demographic characteristic data	98
5	Processes in setting up stated choice experiments	100
5.1	Introduction	100
5.2	What is an experimental design?	100
5.2.1	Stage 1: Problem definition refinement	103
5.2.2	Stage 2: Stimuli refinement	104
5.2.2.1	Refining the list of alternatives	104
5.2.2.2	Refining the list of attributes and attribute levels	105
5.2.3	Stage 3: Experimental design considerations	109
5.2.3.1	Labeled versus unlabeled experiments	112
5.2.3.2	Reducing the number of levels	114
5.2.3.3	Reducing the size of experimental designs	115
5.2.3.4	Dummy and effects coding	119
5.2.3.5	Calculating the degrees of freedom required	122
5.2.3.6	Blocking the design	126
5.2.4	Stage 4: Generating experimental designs	127
5.2.4.1	Assigning an attribute as a blocking variable	130
5.2.5	Stage 5: Allocating attributes to design columns	131
5.3	A note on unlabeled experimental designs	150
5.4	Optimal designs	152
	Appendix 5A Designing nested attributes	154
	Appendix 5B Assignment of quantitative attribute-level labels	156
6	Choices in data collection	161
6.1	Introduction	161
6.2	General survey instrument construction	161
6.3	Questionnaires for choice data	166
6.3.1	Stage 6: Generation of choice sets	166
6.3.2	Stage 7: Randomizing choice sets	170
6.3.3	Stage 8: Survey construction	172
6.3.3.1	Choice context	173
6.3.3.2	Use an example	174
6.3.3.3	Independence of choice sets	174
6.3.3.4	More than one choice	175
6.3.3.5	The no-choice or delay-choice alternative	176
6.4	Revealed preferences in questionnaires	177
6.5	Studies involving both RP and SP data	177
6.6	Using RP data in SP experiments: the “current alternative”	178
6.7	Sampling for choice data: the theory	184
6.7.1	Simple random samples	185

viii	Contents	
	6.7.2 Stratified random sampling	190
	6.7.3 Conclusion to the theory of calculating sample sizes	192
	6.8 Sampling for choice data: the reality	193
7	NLOGIT for applied choice analysis: a primer	197
	7.1 Introduction	197
	7.2 About the software	197
	7.2.1 About NLOGIT	197
	7.2.2 About NLOGIT/ACA	198
	7.2.3 Installing NLOGIT/ACA	198
	7.3 Starting NLOGIT/ACA and exiting after a session	198
	7.3.1 Starting the program	198
	7.3.2 Inputting the data	198
	7.3.3 Reading data	200
	7.3.4 The project file	200
	7.3.5 Leaving your session	201
	7.4 Using NLOGIT	201
	7.5 How to get NLOGIT to do what you want	202
	7.5.1 Using the Text Editor	202
	7.5.2 Command format	204
	7.5.3 Commands	205
	7.5.4 Using the Project File Box	206
	7.6 Useful hints and tips	206
	7.6.1 Limitations in NLOGIT (and NLOGIT/ACA)	207
	7.7 NLOGIT software	207
	7.7.1 Support	208
	7.7.2 The program installed on your computer	208
	7.7.3 Using NLOGIT/ACA in the remainder of the book	208
	Appendix 7A Diagnostic and error messages	208
8	Handling choice data	218
	8.1 Introduction	218
	8.2 The basic data setup	219
	8.2.1 Entering multiple data sets: stacking and melding	222
	8.2.2 Handling data on the non-chosen alternative in RP data	222
	8.2.3 Combining sources of data	224
	8.2.4 Weighting on an exogenous variable	226
	8.2.5 Handling rejection: the “no option”	227
	8.3 Entering data into NLOGIT	230
	8.3.1 Entering data directly into NLOGIT	230
	8.3.2 Importing data into NLOGIT	232
	8.3.2.1 The Text/Document Editor	232
	8.3.3 Reading data into NLOGIT	232
	8.3.4 Writing data into NLOGIT	235
	8.3.5 Saving data sets	235

8.3.6 Loading data into NLOGIT	236
8.3.6.1 Changing the maximum default size of the Data Editor	236
8.4 Data entered into a single line	237
8.5 Data cleaning	241
8.5.1 Testing for multicollinearity using NLOGIT	246
Appendix 8A Design effects coding	248
Appendix 8B Converting single-line data commands	250
9 Case study: mode-choice data	254
9.1 Introduction	254
9.2 Study objectives	254
9.3 The pilot study	256
9.3.1 Pilot sample collection	263
9.3.1.1 Interviewer briefing	263
9.3.1.2 Interviewing	264
9.3.1.3 Analysis of contacts	264
9.3.1.4 Interviewer debriefing	265
9.4 The main survey	265
9.4.1 The mode-choice experiment	267
9.4.1.1 Detailed description of attributes	274
9.4.1.2 Using the showcards	276
9.4.2 RP data	276
9.4.3 The household questionnaire	277
9.4.4 The commuter questionnaire	277
9.4.5 The sample	278
9.4.5.1 Screening respondents	282
9.4.5.2 Interviewer briefing	283
9.4.5.3 Interviewing	283
9.4.5.4 Analysis of total contacts	283
9.4.5.5 Questionnaire check edit	284
9.4.5.6 Coding and check edit	284
9.4.5.7 Data entry	286
9.4.5.8 SPSS setup	286
9.5 The case study data	286
9.5.1 Formatting data in NLOGIT	289
9.5.2 Getting to know and cleaning the data	292
Appendix 9A The contextual statement associated with the travel choice experiment	296
Appendix 9B Mode-choice case study data dictionary	298
Appendix 9C Mode-choice case study variable labels	302
10 Getting started modeling: the basic MNL model	308
10.1 Introduction	308
10.2 Modeling choice in NLOGIT: the MNL command	308

x	Contents	
	10.3 Interpreting the MNL model output	316
	10.3.1 Maximum likelihood estimation	317
	10.3.2 Determining the sample size and weighting criteria used	323
	10.3.3 Interpreting the number of iterations to model convergence	324
	10.3.4 Determining overall model significance	326
	10.3.5 Comparing two models	335
	10.3.6 Determining model fit: the pseudo- R^2	337
	10.3.7 Type of response and bad data	339
	10.3.8 Obtaining estimates of the indirect utility functions	339
	10.3.8.1 Matrix: LastDsta/LastOutput	343
	10.4 Interpreting parameters for effects and dummy coded variables	344
	10.5 Handling interactions in choice models	352
	10.6 Measures of willingness to pay	357
	10.7 Obtaining choice probabilities for the sample	360
	10.8 Obtaining the utility estimates for the sample	366
	Appendix 10A Handling unlabelled experiments	371
11	Getting more from your model	374
	11.1 Introduction	374
	11.2 Adding to our understanding of the data	375
	11.2.1 Show	375
	11.2.2 Descriptives	378
	11.2.3 Crosstab	381
	11.3 Adding to our understanding of the model parameters	383
	11.3.1 ;Effects: elasticities	384
	11.3.2 Calculating arc elasticities	392
	11.3.3 ;Effects: marginal effects	393
	11.4 Simulation	399
	11.4.1 Marginal effects for categorical coded variables	407
	11.4.2 Reporting marginal effects	411
	11.5 Weighting	413
	11.5.1 Endogenous weighting	413
	11.5.2 Weighting on an exogenous variable	418
	11.6 Calibrating the alternative-specific constants of choice models estimated on SP data	420
	11.6.1 Example (1) (the market shares of all alternatives are known <i>a priori</i>)	423
	11.6.2 Example (2) (the market shares for some alternatives are unknown)	424
	Appendix 11A Calculating arc elasticities	426
12	Practical issues in the application of choice models	437
	12.1 Introduction	437
	12.2 Calibration of a choice model for base and forecast years	438
	12.3 Designing a population data base: synthetic observations	439

12.4	The concept of synthetic households	440
12.4.1	Synthetic households' generation framework	441
12.5	The population profiler	443
12.5.1	The synthetic household specification	444
12.6	The sample profiler	448
12.7	Establishing attribute levels associated with choice alternatives in the base year and in a forecast year	451
12.8	Bringing the components together in the application phase	452
12.9	Developing a decision support system	453
12.9.1	Using the data sample averages in creating the DSS	455
12.9.2	Using the choice data in creating the DSS	461
12.9.3	Improving the look of the DSS	472
12.9.4	Using the DSS	475
12.10	Conclusion	475
Part II Advanced topics		
13	Allowing for similarity of alternatives	479
13.1	Introduction	479
13.2	Moving away from IID between all alternatives	481
13.3	Setting out the key relationships for establishing a nested logit model	482
13.4	The importance of a behaviorally meaningful linkage mechanism between the branches on a nested structure	486
13.5	The scale parameter	487
13.6	Bounded range for the IV parameter	490
13.7	Searching for the "best" tree structure	494
	Appendix 13A Technical details of the nested logit model	495
14	Nested logit estimation	518
14.1	Introduction	518
14.2	The Hausman-test of the IIA assumption	519
14.3	The nested logit model commands	530
14.3.1	Normalizing and constraining IV parameters	534
14.3.2	RU1 and RU2	538
14.3.3	Specifying start values for the NL model	538
14.3.4	A quick review of the NL model	539
14.4	Estimating an NL model and interpreting the output	541
14.4.1	Estimating the probabilities of a two-level NL model	549
14.4.2	Comparing RU1 to RU2	556
14.5	Specifying utility functions at higher levels of the NL tree	564
14.6	Handling degenerate branches in NL models	570
14.7	Three-level NL models	574
14.8	Searching for the best NL tree structure: the degenerate nested logit	577
14.9	Combining sources of data: SP-RP	580

xii	Contents	
	14.10 Additional commands	592
	Appendix 14A The Hausman-test of the IIA assumption for models with alternative-specific parameter estimates	595
	Appendix 14B Three-level NL model system of equations	601
15	The mixed logit model	605
	15.1 Introduction	605
	15.2 Mixed logit choice models	606
	15.3 Conditional distribution for sub-populations with common choices	610
	15.4 Model specification issues	611
	15.4.1 Selecting the random parameters	611
	15.4.2 Selecting the distribution of the random parameters	612
	15.4.3 Imposing constraints on a distribution	614
	15.4.4 Selecting the number of points for the simulations	614
	15.4.5 Preference heterogeneity around the mean of a random parameter	617
	15.4.6 Accounting for observations drawn from the same individual correlated choice situations	617
	15.4.7 Accounting for correlation between parameters	619
	15.5 Willingness-to-pay challenges	620
	15.6 Conclusions	621
16	Mixed logit estimation	623
	16.1 Introduction	623
	16.2 The mixed logit model basic commands	623
	16.3 NLOGIT output: interpreting the mixed logit model	627
	16.4 How can we use random parameter estimates?	635
	16.4.1 A note on using the lognormal distribution	641
	16.5 Imposing constraints on a distribution	645
	16.6 Revealing preference heterogeneity around the mean of a random parameter	650
	16.6.1 Using the non-stochastic distribution	656
	16.6.2 Handling insignificant heterogeneity around the mean parameter estimates	661
	16.7 Correlated parameters	667
	16.8 Common-choice-specific parameter estimates: conditional parameters	679
	16.9 Presenting the distributional outputs graphically using a kernel density estimator	684
	16.10 Willingness-to-pay issues and the mixed logit model	686
	<i>Glossary</i>	695
	<i>References</i>	710
	<i>Index</i>	714

Figures

2.1	The PDF of a continuous random variable	<i>page</i> 16
2.2	The CDF of a discrete random variable	18
2.3	The CDF of a continuous random variable	18
2.4	Plotting the home–work, work–home trip times	35
2.5	Sampling distribution for 1000 draws	41
2.6	One-tailed test distribution	43
2.7	Two-tailed test distribution	44
2.8	The relationship between α and β	47
2.9	The rejection region for one- or two-tailed tests	49
2.10	Calculating the determinant of a matrix in Excel	56
2.11	A singular matrix	56
2.12	The inverse of a matrix in Microsoft Excel	58
2.13	Calculating the inverse of a matrix in Microsoft Excel	58
3.1	The identification of an individual’s preferences for bus use	65
3.2	The budget or resource constraint	67
3.3	Changes to the budget or resource constraint	68
3.4	Individual preferences subject to a budget constraint	68
3.5	Indifference curves with budget constraints	69
3.6	Demand curve construction	70
3.7	Changes in demand and changes in quantity demanded	71
4.1	The technological frontier and the roles of RP and SP data	97
5.1	The experimental design process	102
5.2	Mapping part-worth utility	107
5.3	Stages in deriving fractional factorial designs	115
5.4	Estimation of linear vs quadratic effects	121
5.5	Generating designs using SPSS	128
5.6	Specifying the number of attribute levels per attribute	129
5.7	Specifying the minimum number of treatment combinations to generate	130
5.8	Calculating interaction design codes using Microsoft Excel	133

xiv	List of figures	
5.9	Microsoft Excel commands to generate correlations	133
5.10	Microsoft Excel Data Analysis and Correlation dialog boxes	140
6.1	Stages 6–8 of the experimental design process	166
6.2	Example choice sets	169
6.3	The meaning of grey	169
6.4	Example shoe choice set	170
6.5	Mode-choice example	174
6.6	Example of choice set with more than two choices	175
6.7	Collecting RP data for inclusion in an SP experiment	178
6.8	Choice sets using fixed attribute-level labels	180
6.9	Choice sets using percentage attribute-level labels	183
6.10	Calculating Z^2 using Microsoft Excel	186
6.11	Calculating sample sizes using Microsoft Excel	187
6.12	Calculating the allowable error using Microsoft Excel	189
6.13	Within-stratum acceptable error using overall population proportions	192
6.14	Within-stratum acceptable error using strata population proportions	193
7.1	Initial NLOGIT desktop	199
7.2	File Menu on Main Desktop and Open Project . . . Explorer	199
7.3	NLOGIT desktop after Project File Input	2000
7.4	Dialog for Exiting NLOGIT and Saving the Project File	201
7.5	Dialog for Opening the Text Editor	203
7.6	Text Editor Ready for Command Entry	203
7.7	Commands in the Text Editor	204
8.1	Choice set with the “no-travel” alternative	228
8.2	Creating new variables using the toolbar commands	230
8.3	Naming new variables using the <i>New Variable</i> dialog box	231
8.4	<i>Project</i> dialog box	231
8.5	The New Text/Document Editor pushbutton	232
8.6	The Go pushbutton	232
8.7	Changing the default workspace available in NLOGIT	237
8.8	The <i>Options</i> dialog box	237
8.9	Differences between the data and experimental design setups	247
9.1	Instructions and format of the initial mode-choice survey	258
9.2	Example of the format of the mode-choice experiment showcard	268
10.1	Likelihood surface	319
10.2	A complex negative LL surface with local sub-optimal solutions	321
10.3	The sigmoid curve	328
10.4	Using Excel to perform the $-2LL$ test	331
10.5	The $-2LL$ Chi-square test	331
10.6	Using Excel to calculate a pseudo- R^2	338
10.7	Mapping the pseudo- R^2 to the linear R^2	338
10.8	<i>Matrix: LastDsta</i> button	343
10.9	<i>Matrix: LastOutp</i> example	344
10.10	Linear and non-linear in parameters marginal utilities	346

10.11	Marginal utility estimates for the “number of vehicles–number of drivers” interaction	354
10.12	Marginal utility estimates for the “fare–time” interaction	357
10.13	Model probabilities saved within NLOGIT	361
10.14	Calculating choice probabilities using Excel	362
10.15	Using Excel to perform “what-if” scenarios on the choice probabilities	364
10.16	The <i>Project</i> dialog box and <i>New Scalar</i> dialog box	365
10.17	Estimated utilities saved within NLOGIT	368
10.18	Using Excel to test “what-if” scenarios on the utility estimates	369
10.19	Obtaining the average utility using the Calc; command	370
10A.1	Two scenarios from an unlabeled choice experiment	372
11.1	Proportions of correct and incorrect predictions	383
11.2	Marginal effects as the slopes of the tangent lines to the cumulative probability curve	395
11.3	Marginal effects for a categorical (dummy coded) variable	396
11A.1	Data Editor for calculating arc elasticities	432
11A.2	Within-choice-set elasticities	433
12.1	Synthetic household generation process	442
12.2	Constructing the front end of a DSS	445
12.3	Using the Microsoft Excel Data Validation option	456
12.4	Creating the back end of a DSS	456
12.5	Setting up the output for the DSS	457
12.6	Creating a graph to visually represent predicted market adjustments due to policy changes	458
12.7	Using the attribute-level data averages as the initial DSS attribute levels	458
12.8	Creating a front end using percentage changes	462
12.9	Copying the raw choice data into the DSS	463
12.10	Linking the back end to the front end of the DSS	464
12.11	Using the <i>vlookup</i> command in Microsoft Excel	464
12.12	Using the <i>if statement</i> to restrict the possible attribute-level ranges in the data	465
12.13	Inserting the parameter estimates into the back end of the DSS	466
12.14	Calculating the utilities for each alternative	467
12.15	Calculating the choice probabilities	467
12.16	PivotTable and PivotChart Report command menu	468
12.17	PivotTable and PivotChart Wizard box	468
12.18	Selecting the pivot table data using the PivotTable Wizard	469
12.19	The third PivotTable and PivotChart Wizard	469
12.20	Creating the pivot table through the Layout Wizard	470
12.21	Calculating the choice shares using the Microsoft pivot table	471
12.22	Refreshing the data of the pivot table	471
12.23	The Introduction screen buttons	472
12.24	The Microsoft Excel Forms toolbar	472

xvi	List of figures	
12.25	Recording macros	473
12.26	The DSS Help screen	474
12.27	Run Model and Reset Attribute Levels for DSS2	475
13.1	A tree diagram to recognize the (potential) linkages between choice sets	485
13A.1	A four-level NL tree structure	496
13A.2	Relationship between the scale parameters at levels 1 and 2 of an NL tree structure	500
13A.3	Relationship between the scale parameters at levels 2 and 3 of an NL tree structure	507
13A.4	Relationship between the scale parameters at levels 3 and 4 of an NL tree structure	511
14.1	Example NL tree structure one	531
14.2	Example tree structure two	532
14.3	Example tree structure three	533
14.4	Example tree structure four	534
14.5	Calculating probability and utilities of an NL model using NLOGIT	554
14.6	Example calculations of the probabilities and utilities of an NL model using NLOGIT	556
14.7	An NL tree structure with a degenerate alternative	571
14.8	An NL tree structure with two degenerate alternatives	572
14.9	A three-level NL tree structure with degenerate branches	574
14.10	Pooling RP and SP data sources	583
14.11	Saving predicted probabilities, conditional probabilities, and IV parameters	594
14.12	Saving the utility estimates in NL models	595
15.1	1000 draws on the <i>unit square</i>	615
16.1	Testing for statistical differences in means of triangular and normal distributions	632
16.2	Testing dispersion of the <i>toll</i> random parameter	633
16.3	Testing dispersion of the <i>cartc</i> random parameter	634
16.4	Histogram of a randomly drawn normal distribution with mean zero and standard deviation one	637
16.5	Histogram of a hypothetical sample for a lognormal distributed random parameter	638
16.6	Hypothetical individual-specific parameter estimates derived from a lognormal distribution	638
16.7	Hypothetical parameter distribution for a standard normal distribution	639
16.8	Histogram of hypothetical parameter estimates for a uniform distribution	640
16.9	Transforming a uniform distribution into a triangular distribution	642
16.10	Histogram of hypothetical parameter estimates for a triangular distribution	643
16.11	Individual-specific parameter estimates derived from an unconditional lognormal random parameter	645

16.12	Placing a constraint upon the dispersion of the <i>cartc</i> random parameter estimate	646
16.13	Means and spreads of <i>ctc</i> , <i>ctnc</i> , <i>ctp</i> , and <i>cntp</i> random parameters	671
16.14	Plotting the marginal utilities from unconditional parameter estimates	680
16.15	Locating the conditional parameter estimates	682
16.16	A matrix with the stored conditional random parameter estimates	683
16.17	Kernel density function for <i>fuel</i> and <i>toll cost</i> parameters	686
16.18	WTP-I matrix with WTP ratio	689
16.19	Calculating the VTTS for conditional parameter estimates (unconstrained distribution)	689
16.20	Calculating the VTTS for conditional parameter estimates (constrained distribution)	692
16.21	The <i>car</i> VTTS distribution	693
16.22	The <i>public transport</i> VTTS distribution	693
16.23	Plotting the VTTS derived from triangular distributions	695

Tables

2.1	Calculating the relative frequencies for travel to work	<i>page</i> 13
2.2	The PDF of a discrete random variable	15
2.3	PDF and CDF for a discrete random variable	17
2.4	Frequency counts for home–work and work–home trips	19
2.5	The bivariate distribution for home–work and work–home trips	20
2.6	Marginal probability distributions	22
2.7	Demonstrating statistical independence of two random variables	23
2.8	Demonstrating that the home–work and work–home travel times are not statistically independent	24
2.9	The expected value of rolling two dice	25
2.10	Calculating the average home–work and work–home travel times	26
2.11	The expected values of two separate rolls of a die	26
2.12	The expected value of a random variable multiplied by a constant	27
2.13	Calculating the variance of rolling two dice	29
2.14	The variance of two independent throws of a die	29
2.15	Calculating $X_1X_2f(X_1, X_2)$ for home–work and work–home trips	31
2.16	Calculating $X_1X_2f(X_1, X_2)$ for the roll of two dice	32
2.17	Estimating the variances for home–work and work–home trips	35
2.18	Respondent ratings of two soft drinks	37
2.19	Calculating the sample covariance between the ratings of two soft drinks	39
2.20	The relationships between H_0 , H_1 , Type I errors, and Type II errors	47
2.21	Comparing alpha to the probability value	50
2A.1	Appropriate correlation formula	60
5.1	Full factorial design	110
5.2	Full factorial design coding	110
5.3	Comparison of design codes and orthogonal codes	111
5.4	Choice treatment combination	112
5.5	Labeled choice experiment	112
5.6	Attribute levels for expanded number of alternatives	114
5.7	Dummy coding	119

5.8	Effects coding structure	120
5.9	Effects coding formats	121
5.10	Minimum treatment combination requirements for main effects only fractional factorial designs	123
5.11	Enumeration of all two-way interactions	125
5.12	Orthogonal coding of fractional factorial design	132
5.13	Orthogonal codes for main effects plus all two-way interaction columns	134
5.14	Design correlation	136
5.15	Attributes assigned to design columns	143
5.16	Using blocking variables to determine allocation of treatment combinations	145
5.17	Effects coding design of table 5.15	146
5.18	Correlation matrix for effects coded design	147
5.19	3^4 Fractional factorial design	148
5.20	Randomizing treatment combinations to use for additional design columns	148
5.21	Correlation matrix for randomizing treatment combinations	149
5.22	Using the foldover to generate extra design columns	149
5.23	Correlation matrix for designs using foldovers to generate additional columns	150
5A.1	<i>Price–quality</i> combinations	154
5A.2	Fractional factorial of a $6^1 \times 3^2$ design with <i>price–quality</i> nested attribute	155
5A.3	Correlation matrix of a fractional factorial of a $6^1 \times 3^2$ design with nested <i>price–quality</i> attribute	155
5B.1	Comparison of coding formats	157
5B.2	Estimating linear effects using quantitative attributes	158
5B.3	Correlation matrix for design shown in table 5B.2	159
5B.4	Estimating linear effects using different quantitative attribute-level labels	159
5B.5	Correlation matrix for the design shown in table 5B.3	160
6.1	Attaching cognitively meaningful attribute labels to attribute levels	167
6.2	A reproduction of the first two treatment combinations for the table 5.11 design	168
6.3	A 2^4 orthogonal design with two blocking variables	171
6.4	Randomization of choice sets across surveys	172
6.5	Reported attribute levels and SP attribute-level labels, different approaches	179
6.6	Attribute-level labels for a 2^4 orthogonal design	179
6.7	Pivoting SP attribute levels from reported attribute levels using percentages	181
6.8	Attribute-level labels as percentage changes for a 2^4 orthogonal design	182

xx	List of tables	
6.9	Pivoting SP attribute levels from reported attribute levels using wider percentages	182
6.10	Differences between relative and absolute treatment of sample proportions	187
6.11	Breakdown of intercity travellers, by purpose of trip	190
6.12	Strata sample sizes calculated using the overall population proportions	191
6.13	Strata sample sizes calculated using the strata population proportions	191
8.1	Most general choice data format in NLOGIT	220
8.2	Varying alternatives within choice sets	220
8.3	Varying the number of alternatives within choice sets: (1)	221
8.4	Varying the number of alternatives within choice sets: (2)	223
8.5	Entering socio-demographic characteristics	223
8.6	Combining SP and RP data	225
8.7	Exogenous weights entered	227
8.8	Adding the “no-choice” or “delay-choice” alternative	229
8.9	Data entered into a single line	239
8.10	RP data in a single-line data format	239
8A.1	Design effects coding	248
9.1	Attributes and levels used for the initial mode-choice showcards	257
9.2	Instructions and format of the initial mode-choice survey	259
9.3	Orthogonal fractional factorial mode-choice pilot experimental design	260
9.4	The set of attributes and attribute levels in the mode-choice experiment	269
9.5	Mode-choice experimental design	270
9.6	Correlation matrix for mode-choice experimental design	273
9.7	Number of interviews for each city	278
9.8	Targeted number of interviews, by location	279
9.9	Summary at urban area-wide level of profile of households, by fleet size	281
9.10	Final questionnaire totals, by state (after ITS editing)	285
9.11	Breakdown, by city, for the SP and RP data sets	285
9.12	Profile of RP modal share, chosen main mode	287
9.13	Profile of RP modal share, alternative mode	287
9.14	Effects coding for the <i>wkremply</i> variable	290
10.1	Reproduction of table 8.8	314
10.2	Searching over a range of values to maximize L*	318
10.3	Calculating log likelihood values	320
10.4	Part-worth utility estimates for the “number of vehicles–number of drivers” interaction	354
10.5	Part-worth utility estimates for the fare–time interaction	354

11.1	Relationship between elasticity of demand, change in price, and revenue	387
11.2	Reporting marginal effects	412
12.1	Household distribution at population level: indexing 25 core cells	443
12.2	Cross-tabulation of households at core-cell level	444
12.3	SHD1 method	446
12.4	SHD2: uniform distribution method	447
12.5	Uniform distribution allocation	447
12.6	SHD3	448
12.7	Household characteristics: a comparison of synthetic household projection to Census data	449
12.8	Resident characteristics: a comparison of synthetic household projection to Census data	450
14.1	<i>SPRP alti2</i> values	583
16.1	Comparison of the VTTS derived from unconditional parameter estimates, conditional (unconstrained), and conditional (constrained) parameter estimates	692
16.2	Statistical outputs for VTTS estimated using triangular distributions	695

Preface

I'm all in favor of keeping dangerous weapons out of the hands of fools. Let's start with typewriters. (Frank Lloyd Wright, 1868–1959)

Almost without exception, everything human beings undertake involves a *choice* (consciously or sub-consciously), including the choice not to choose. Some choices are the result of habit while others are fresh decisions made with great care, based on whatever information is available at the time from past experiences and/or current inquiry.

Since the 1970s, there has been a steadily growing interest in the development and application of quantitative statistical methods to study choices made by individuals (and, to a lesser extent, groups of individuals). With an emphasis on both understanding how choices are made and forecasting future choice responses, a healthy literature has evolved. Reference works by Louviere, Hensher, and Swait (2000), and Train (2003) synthesize the contributions. However while these two sources represent the state of the art (and practice), they are technically advanced and often a challenge for the beginner and practitioners.

Discussions with colleagues over the last few years have revealed a gap in the literature of choice analysis – a book that assumes very little background and offers an entry point for individuals interested in the study of choice regardless of their background. Writing such a book increasingly became a challenge for us. It is often more difficult to explain complex ideas in very simple language than to protect one's knowledge-base with complicated deliberations.

There are many discussion topics in this primer that are ignored in most books on the subject, yet are issues which students have pointed out in class as important in giving them a better understanding of what is happening in choice modeling. The lament that too many books on discrete choice analysis are written for the well informed is common and is sufficient incentive to write this book.

This primer for beginners is our attempt to meet the challenge. We agreed to try and write the first draft without referring to any of the existing material as a means (hopefully) of encouraging a flow of explanation. Pausing to consult can often lead to terseness in the code (as writers of novels can attest). Further draft versions leading to the final product did, however, cross-reference to the literature to ensure we had acknowledged appropriate

Cambridge University Press
0521605776 - Applied Choice Analysis: A Primer
David A. Hensher, John M. Rose and William H. Greene
Frontmatter
[More information](#)

xxiv Preface

material. This primer, however, is not about ensuring that all contributors to the literature on choice are acknowledged, but rather to ensure that the novice choice analyst is given a fair go in their first journey through this intriguing topic.

We dedicate this book to the beginners but we also acknowledge our research colleagues who have influenced our thinking as well as co-authored papers over many years. We especially recognize Dan McFadden (2000 Nobel Laureate in Economics), Ken Train, Chandra Bhat, Jordan Louviere, Andrew Daly, Moshe Ben-Akiva, and David Brownstone. Colleagues and doctoral students at the University of Sydney read earlier versions. In particular, we thank Sean Puckett, Kwang Kim and Louise Knowles and the January 2004 graduate class in Choice Analysis at The University of Sydney, who were guinea pigs for the first full use of the book in a teaching environment. Sean Puckett also contributed to the development of the glossary.