

Applied Choice Analysis

A Primer

David A. Hensher, John M. Rose, and William H. Greene

CAMBRIDGE

Contents

<i>List of figures</i>	<i>page</i> 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

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

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

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

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