

Microeconometrics Using Stata

Revised Edition

A. COLIN CAMERON
Department of Economics
University of California
Davis, CA

PRAVIN K. TRIVEDI
Department of Economics
Indiana University
Bloomington, IN



A Stata Press Publication
StataCorp LP
College Station, Texas

Contents

List of tables	xxxv
List of figures	xxxvii
Preface to the Revised Edition	xxxix
Preface to the First Edition	xli
Stata basics	1
1.1 Interactive use	1
1.2 Documentation	2
1.2.1 Stata manuals	2
1.2.2 Additional Stata resources	3
1.2.3 The help command	3
1.2.4 The search, findit, and hsearch commands	4
1.3 Command syntax and operators	5
1.3.1 Basic command syntax	5
1.3.2 Example: The summarize command	6
1.3.3 Example: The regress command	7
1.3.4 Factor variables	9
1.3.5 Abbreviations, case sensitivity, and wildcards	11
1.3.6 Arithmetic, relational, and logical operators	12
1.3.7 Error messages	12
1.4 Do-files and log files	13
1.4.1 Writing a do-file	13
1.4.2 Running do-files	14
1.4.3 Log files	14
1.4.4 A three-step process	15

1.4.5	Comments and long lines.16
1.4.6	Different implementations of Stata.17
1.5	Scalars and matrices.17
1.5.1	Scalars.17
1.5.2	Matrices.18
1.6	Using results from Stata commands.18
1.6.1	Using results from the r-class command summarize.18
1.6.2	Using results from the e-class command regress.19
1.7	Global and local macros.21
1.7.1	Global macros.21
1.7.2	Local macros.22
1.7.3	Scalar or macro?.23
1.8	Looping commands.24
1.8.1	The foreach loop.25
1.8.2	The forvalues loop.26
1.8.3	The while loop.26
1.8.4	The continue command.27
1.9	Some useful commands.27
1.10	Template do-file.27
1.11	User-written commands.28
1.12	Stata resources.29
1.13	Exercises.29
2	Data management and graphics	31
2.1	Introduction.	31
2.2	Types of data	31
2.2.1	Text or ASCII data	32
2.2.2	Internal numeric data.	32
2.2.3	String data.	33
2.2.4	Formats for displaying numeric data.	33

2.3	Inputting data	34
2.3.1	General principles.	34
2.3.2	Inputting data already in Stata format	35
2.3.3	Inputting data from the keyboard.	36
2.3.4	Inputting nontext data	36
2.3.5	Inputting text data from a spreadsheet	37
2.3.6	Inputting text data in free format.	38
2.3.7	Inputting text data in fixed format	38
2.3.8	Dictionary files.	39
2.3.9	Common pitfalls.	39
2.4	Data management	40
2.4.1	PSID example.	40
2.4.2	Naming and labeling variables	43
2.4.3	Viewing data	44
2.4.4	Using original documentation.	45
2.4.5	Missing values.	45
2.4.6	Imputing missing data.	47
2.4.7	Transforming data (generate, replace, egen, recode).	48
	The generate and replace commands.	48
	The egen command	49
	The recode command	49
	The by prefix.	49
	Indicator variables.	50
	Set of indicator variables.	50
	Interactions.	51
	Demeaning	52
2.4.8	Saving data	52
2.4.9	Selecting the sample.	53
2.5	Manipulating datasets.	54
2.5.1	Ordering observations and variables.	55

2.5.2	Preserving and restoring a dataset	55
2.5.3	Wide and long forms for a dataset	55
2.5.4	Merging datasets.	56
2.5.5	Appending datasets.	58
2.6	Graphical display of data	58
2.6.1	Stata graph commands.	59
	Example graph commands.	59
	Saving and exporting graphs.	60
	Learning how to use graph commands.	61
2.6.2	Box-and-whisker plot	61
2.6.3	Histogram	63
2.6.4	Kernel density plot	63
2.6.5	Twoway scatterplots and fitted lines.	66
2.6.6	Lowess, kernel, local linear, and nearest-neighbor regression	67
2.6.7	Multiple scatterplots.	69
2.7	Stata resources.	70
2.8	Exercises.	70
	Linear regression basics	73
3.1	Introduction	73
3.2	Data and data summary.	73
3.2.1	Data description	73
3.2.2	Variable description.	74
3.2.3	Summary statistics.	75
3.2.4	More-detailed summary statistics.	76
3.2.5	Tables for data	77
3.2.6	Statistical tests.	80
3.2.7	Data plots.	80
3.3	Regression in levels and logs.	81
3.3.1	Basic regression theory.	81
3.3.2	OLS regression and matrix algebra	82

3.3.3	Properties of the OLS estimator.	83
3.3.4	Heteroskedasticity-robust standard errors.	84
3.3.5	Cluster—robust standard errors.	84
3.3.6	Regression in logs.	85
3.4	Basic regression analysis.	86
3.4.1	Correlations.	86
3.4.2	The regress command.	87
3.4.3	Hypothesis tests.	88
3.4.4	Tables of output from several regressions.	89
3.4.5	Even better tables of regression output.	90
3.4.6	Factor variables for categorical variables and interactions	92
3.5	Specification analysis.	94
3.5.1	Specification tests and model diagnostics.	94
3.5.2	Residual diagnostic plots.	95
3.5.3	Influential observations.	96
3.5.4	Specification tests.	97
	Test of omitted variables.	98
	Test of the Box-Cox model.	98
	Test of the functional form of the conditional mean.	99
	Heteroskedasticity test.	100
	Omnibus test.	102
3.5.5	Tests have power in more than one direction.	102
3.6	Prediction.	104
3.6.1	In-sample prediction.	104
3.6.2	MEs and elasticities.	106
3.6.3	Prediction in logs: The retransformation problem.	108
3.6.4	Prediction exercise.	109
3.7	Sampling weights.	111
3.7.1	Weights.	111
3.7.2	Weighted mean.	112

3.7.3	Weighted regression113
3.7.4	Weighted prediction and MEs.114
3.8	OLS using Mata115
3.9	Stata resources.117
3.10	Exercises.117
	Simulation	119
4.1	Introduction119
4.2	Pseudorandom-number generators: Introduction.120
4.2.1	Uniform random-number generation.120
4.2.2	Draws from normal.122
4.2.3	Draws from t, chi-squared, F, gamma, and beta123
4.2.4	Draws from binomial, Poisson, and negative binomial124
	Independent (but not identically distributed) draws from binomial.124
	Independent (but not identically distributed) draws from Poisson.125
	Histograms and density plots.126
4.3	Distribution of the sample mean.127
4.3.1	Stata program128
4.3.2	The simulate command129
4.3.3	Central limit theorem simulation.129
4.3.4	The postfile command.130
4.3.5	Alternative central limit theorem simulation.131
4.4	Pseudorandom-number generators: Further details.131
4.4.1	Inverse-probability transformation.132
4.4.2	Direct transformation.133
4.4.3	Other methods.133
4.4.4	Draws from truncated normal134
4.4.5	Draws from multivariate normal135
	Direct draws from multivariate normal.135
	Transformation using Cholesky decomposition.136

4.4.6	Draws using Markov chain Monte Carlo method136
4.5	Computing integrals.138
4.5.1	Quadrature.139
4.5.2	Monte Carlo integration.139
4.5.3	Monte Carlo integration using different S.140
4.6	Simulation for regression: Introduction.141
4.6.1	Simulation example: OLS with $\sqrt{\sigma^2}$ errors.141
4.6.2	Interpreting simulation output.144
	Unbiasedness of estimator.144
	Standard errors.144
	t statistic.144
	Test size.145
	Number of simulations.146
4.6.3	Variations.146
	Different sample size and number of simulations.146
	Test power.146
	Different error distributions.147
4.6.4	Estimator inconsistency.147
4.6.5	Simulation with endogenous regressors.148
4.7	Stata resources.150
4.8	Exercises.150
5	GLS regression	153
5.1	Introduction.153
5.2	GLS and FGLS regression.153
5.2.1	GLS for heteroskedastic errors.153
5.2.2	GLS and FGLS.154
5.2.3	Weighted least squares and robust standard errors.155
5.2.4	Leading examples.155
5.3	Modeling heteroskedastic data.156
5.3.1	Simulated dataset.156

5.3.2	OLS estimation	157
5.3.3	Detecting heteroskedasticity.	158
5.3.4	FGLS estimation.	160
5.3.5	WLS estimation.	162
5.4	System of linear regressions.	162
5.4.1	SUR model.	162
5.4.2	The sureg command	163
5.4.3	Application to two categories of expenditures.	164
5.4.4	Robust standard errors.	166
5.4.5	Testing cross-equation constraints.	167
5.4.6	Imposing cross-equation constraints.	168
5.5	Survey data: Weighting, clustering, and stratification.	169
5.5.1	Survey design.	170
5.5.2	Survey mean estimation	173
5.5.3	Survey linear regression.	173
5.6	Stata resources.	175
5.7	Exercises.	175
6	Linear instrumental-variables regression	177
6.1	Introduction.	177
6.2	IV estimation	177
6.2.1	Basic IV theory.	177
6.2.2	Model setup.	179
6.2.3	IV estimators: IV, 2SLS. and GMM.	180
6.2.4	Instrument validity and relevance.	181
6.2.5	Robust standard-error estimates.	182
6.3	IV example.	183
6.3.1	The ivregress command	183
6.3.2	Medical expenditures with one endogenous regressor	184
6.3.3	Available instruments.	185
6.3.4	IV estimation of an exactly identified model.	186

6.3.5	IV estimation of an overidentified model187
6.3.6	Testing for regressor endogeneity.188
6.3.7	Tests of overidentifying restrictions.191
6.3.8	IV estimation with a binary endogenous regressor.192
6.4	Weak instruments.194
6.4.1	Finite-sample properties of IV estimators194_
6.4.2	Weak instruments.195
	Diagnostics for weak instruments.195
	Formal tests for weak instruments.196
6.4.3	The estat firststage command197
6.4.4	Just-identified model197
6.4.5	Overidentified model199
6.4.6	More than one endogenous regressor.	200
6.4.7	Sensitivity to choice of instruments.	200
6.5	Better inference with weak instruments	202
6.5.1	Conditional tests and confidence intervals	202
6.5.2	LIML estimator.	204
6.5.3	Jackknife IV estimator.	204
6.5.4	Comparison of 2SLS, LIML, JIVE, and GMM.	205
6.6	3SLS systems estimation	206
6.7	Stata resources.	208
6.8	Exercises.	208
7	Quantile regression	211
7.1	Introduction	211
7.2	QR.	211
7.2.1	Conditional quantiles.	212
7.2.2	Computation of QR estimates and standard errors.	213
7.2.3	The qreg, bsqreg, and sqreg commands.	213
7.3	QR for medical expenditures data	214
7.3.1	Data summary.	214

7.3.2	QR estimates	215
7.3.3	Interpretation of conditional quantile coefficients	216
7.3.4	Retransformation	217
7.3.5	Comparison of estimates at different quantiles	218
7.3.6	Heteroskedasticity test	219
7.3.7	Hypothesis tests	220
7.3.8	Graphical display of coefficients over quantiles	221
7.4	QR for generated heteroskedastic data	222
7.4.1	Simulated dataset	222
7.4.2	QR estimates	225
7.5	QR for count data	226
7.5.1	Quantile count regression	227
7.5.2	The qcount command	228
7.5.3	Summary of doctor visits data	228
7.5.4	Results from QCR	230
7.6	Stata resources	232
7.7	Exercises	232
	Linear panel-data models: Basics	235
8.1	Introduction	235
8.2	Panel-data methods overview	235
8.2.1	Some basic considerations	236
8.2.2	Some basic panel models	237
	Individual-effects model	237
	Fixed-effects model	237
	Random-effects model	238
	Pooled model or population-averaged model	238
	Two-way-effects model	238
	Mixed linear models	239
8.2.3	Cluster-robust inference	239
8.2.4	The xtreg command	239

8.2.5	Stata linear panel-data commands.	240
8.3	Panel-data summary.	240
8.3.1	Data description and summary statistics.	240
8.3.2	Panel-data organization.	242
8.3.3	Panel-data description.	243
8.3.4	Within and between variation.	244
8.3.5	Time-series plots for each individual.	247
8.3.6	Overall scatterplot.	248
8.3.7	Within scatterplot.	249
8.3.8	Pooled OLS regression with cluster-robust standard errors.	250
8.3.9	Time-series autocorrelations for panel data.	251
8.3.10	Error correlation in the RE model.	253
8.4	Pooled or population-averaged estimators.	254
8.4.1	Pooled OLS estimator.	254
8.4.2	Pooled FGLS estimator or population-averaged estimator.	254
8.4.3	The xtreg, pa command.	255
8.4.4	Application of the xtreg, pa command.	256
8.5	Within estimator.	257
8.5.1	Within estimator.	257
8.5.2	The xtreg, fe command.	257
8.5.3	Application of the xtreg, fe command.	258
8.5.4	Least-squares dummy-variables regression.	259
8.6	Between estimator.	260
8.6.1	Between estimator.	260
8.6.2	Application of the xtreg, be command.	261
8.7	RE estimator.	261
8.7.1	RE estimator.	262
8.7.2	The xtreg, re command.	262
8.7.3	Application of the xtreg, re command.	263

8.8	Comparison of estimators	264
8.8.1	Estimates of variance components	264
8.8.2	Within and between R-squared	264
8.8.3	Estimator comparison	265
8.8.4	Fixed effects versus random effects	266
8.8.5	Hausman test for fixed effects	266
	The hausman command	267
	Robust Hausman test	267
8.8.6	Prediction	268
8.9	First-difference estimator	269
8.9.1	First-difference estimator	270
8.9.2	Strict and weak exogeneity	271
8.10	Long panels	271
8.10.1	Long-panel dataset	271
8.10.2	Pooled OLS and PGLS	273
8.10.3	The xtpcse and xtglm commands	273
8.10.4	Application of the xtglm, xtpcse, and xtsc commands	274
8.10.5	Separate regressions	276
8.10.6	FE and RE models	277
8.10.7	Unit roots and cointegration	278
8.11	Panel-data management	280
8.11.1	Wide-form data	280
8.11.2	Convert wide form to long form	280
8.11.3	Convert long form to wide form	281
8.11.4	An alternative wide-form data	282
8.12	Stata resources	284
8.13	Exercises	284
	Linear panel-data models: Extensions	287
9.1	Introduction	287
9.2	Panel IV estimation	287

9.2.1	Panel IV.	287
9.2.2	The xtivreg command.	288
9.2.3	Application of the xtivreg command.	288
9.2.4	Panel IV extensions.	290
9.3	Hausman-Taylor estimator.	290
9.3.1	Hausman-Taylor estimator.	290
9.3.2	The xthtaylor command.	291
9.3.3	Application of the xthtaylor command.	291
9.4	Arellano-Bond estimator.	293
9.4.1	Dynamic model	293
9.4.2	IV estimation in the FD model	294
9.4.3	The xtabond command.	295
9.4.4	Arellano-Bond estimator: Pure time series.	296
9.4.5	Arellano-Bond estimator: Additional regressors.	298
9.4.6	Specification tests.	300
9.4.7	The xtdpdsys command.	301
9.4.8	The xtdpd command.	303
9.5	Mixed linear models.	305
9.5.1	Mixed linear model.	305
9.5.2	The xtmixed command.	306
9.5.3	Random-intercept model.	306
9.5.4	Cluster-robust standard errors.	307
9.5.5	Random-slopes model.	308
9.5.6	Random-coefficients model.	310
9.5.7	Two-way random-effects model.	311
9.6	Clustered data	312
9.6.1	Clustered dataset.	312
9.6.2	Clustered data using nonpanel commands.	313
9.6.3	Clustered data using panel commands.	314
9.6.4	Hierarchical linear models.	316

9.7	Stata resources	317
9.8	Exercises	318
10	Nonlinear regression methods	319
10.1	Introduction	319
10.2	Nonlinear example: Doctor visits	320
10.2.1	Data description	320
10.2.2	Poisson model description	321
10.3	Nonlinear regression methods	322
10.3.1	MLE	322
10.3.2	The poisson command	323
10.3.3	Postestimation commands	324
10.3.4	NLS	325
10.3.5	The nl command	325
10.3.6	GLM	327
10.3.7	The glm command	327
10.3.8	The grmn command	328
10.3.9	Other estimators	330
10.4	Different estimates of the VCE	330
10.4.1	General framework	330
10.4.2	The vce() option	331
10.4.3	Application of the vce() option	332
10.4.4	Default estimate of the VCE	333
10.4.5	Robust estimate of the VCE	334
10.4.6	Cluster-robust estimate of the VCE	335
10.4.7	Heteroskedasticity- and autocorrelation-consistent estimate of the VCE	335
10.4.8	Bootstrap standard errors	336
10.4.9	Statistical inference	336
10.5	Prediction	336
10.5.1	The predict and predictnl commands	337

10.5.2	Application of predict and predictnl	337
10.5.3	Out-of-sample prediction	338
10.5.4	Prediction at a specified value of one of the regressors	339
10.5.5	Prediction at a specified value of all the regressors.	340
10.5.6	Prediction of other quantities.	341
10.5.7	The margins command for prediction	341
10.6	Marginal effects	343
10.6.1	Calculus and finite-difference methods.	343
10.6.2	MEs estimates AME, MEM, and MER.	344
10.6.3	Elasticities and semielasticities.	344
10.6.4	Simple interpretations of coefficients in single-index models	345
10.6.5	The margins command for marginal effects.	346
10.6.6	MEM: Marginal effect at mean	347
	Comparison of calculus and finite-difference methods	348
10.6.7	MER: Marginal effect at representative value.	348
10.6.8	AME: Average marginal effect	349
10.6.9	Elasticities and semielasticities.	351
10.6.10	AME computed manually.	352
10.6.11	Polynomial regressors	354
10.6.12	Interacted regressors	355
10.6.13	Complex interactions and nonlinearities	356
10.7	Model diagnostics	357
10.7.1	Goodness-of-fit measures.	357
10.7.2	Information criteria for model comparison.	359
10.7.3	Residuals.	359
10.7.4	Model-specification tests	361
10.8	Stata resources	361
10.9	Exercises.	361
11	Nonlinear optimization methods	363
11.1	Introduction.	363

11.2	Newton-Raphson method	363
11.2.1	NR method	363
11.2.2	NR method for Poisson.	364
11.2.3	Poisson NR example using Mata	365
	Core Mata code for Poisson NR iterations.	365
	Complete Stata and Mata code for Poisson NR iterations	365
11.3	Gradient methods.	367
11.3.1	Maximization options.	367
11.3.2	Gradient methods.	368
11.3.3	Messages during iterations.	369
11.3.4	Stopping criteria	369
11.3.5	Multiple maximums.	369
11.3.6	Numerical derivatives.	370
11.4	The ml command: If method.	371
11.4.1	The ml command.	372
11.4.2	The If method	372
11.4.3	Poisson example: Single-index model.	373
11.4.4	Negative binomial example: Two-index model	375
11.4.5	NLS example: Nonlikelihood model.	376
11.5	Checking the program	376
11.5.1	Program debugging using ml check and ml trace.	377
11.5.2	Getting the program to run.	378
11.5.3	Checking the data	379
11.5.4	Multicollinearity and near collinearity.	379
11.5.5	Multiple optimums.	380
11.5.6	Checking parameter estimation.	381
11.5.7	Checking standard-error estimation.	382
11.6	The ml command: dO, d1, d2, lfO, lf1, and lf2 methods.	383
11.6.1	Evaluator functions.	383
11.6.2	The dO method.	385

11.6.3	The dl method.	386
11.6.4	The lfl method with the robust estimate of the VCE	387
11.6.5	The d2 and lf2 methods.	388
11.7	The Mata optimize() function.	389
11.7.1	Type d and gf evaluators.	389
11.7.2	Optimize functions.	390
11.7.3	Poisson example.	390
	Evaluator program for Poisson MLE.	390
	The optimize() function for Poisson MLE.	391
11.8	Generalized method of moments.	392
11.8.1	Definition.	393
11.8.2	Nonlinear IV example.	393
11.8.3	GMM using the Mata optimize() function.	394
11.9	Stata resources.	396
11.10	Exercises.	396
12	Testing methods	399
12.1	Introduction.	399
12.2	Critical values and p-values.	399
12.2.1	Standard normal compared with Student's t.	400
12.2.2	Chi-squared compared with F.	400
12.2.3	Plotting densities.	400
12.2.4	Computing p-values and critical values.	402
12.2.5	Which distributions does Stata use?.	403
12.3	Wald tests and confidence intervals.	403
12.3.1	Wald test of linear hypotheses.	403
12.3.2	The test command.	405
	Test single coefficient.	406
	Test several hypotheses.	406
	Test of overall significance.	407
	Test calculated from retrieved coefficients and VCE.	407

12.3.3	One-sided Wald tests.	408
12.3.4	Wald test of nonlinear hypotheses (delta method).	409
12.3.5	The testnl command.	409
12.3.6	Wald confidence intervals.	410
12.3.7	The lincom command.	410
12.3.8	The nlcom command (delta method).	411
12.3.9	Asymmetric confidence intervals.	412
12.4	Likelihood-ratio tests.	413
12.4.1	Likelihood-ratio tests.	413
12.4.2	The lrtest command.	415
12.4.3	Direct computation of LR tests.	415
12.5	Lagrange multiplier test (or score test).	416
12.5.1	LM tests.	416
12.5.2	The estat command.	417
12.5.3	LM test by auxiliary regression.	417
12.6	Test size and power.	419
12.6.1	Simulation DGP: OLS with chi-squared errors.	419
12.6.2	Test size.	420
12.6.3	Test power.	421
12.6.4	Asymptotic test power.	424
12.7	Specification tests.	425
12.7.1	Moment-based tests.	425
12.7.2	Information matrix test.	425
12.7.3	Chi-squared goodness-of-fit test.	426
12.7.4	Overidentifying restrictions test.	426
12.7.5	Hausman test.	426
12.7.6	Other tests.	427
12.8	Stata resources.	427
12.9	Exercises.	427

13	Bootstrap methods	429
13.1	Introduction	429
13.2	Bootstrap methods	429
13.2.1	Bootstrap estimate of standard error	429
13.2.2	Bootstrap methods	430
13.2.3	Asymptotic refinement	430
13.2.4	Use the bootstrap with caution	430
13.3	Bootstrap pairs using the <code>vce(bootstrap)</code> option	431
13.3.1	Bootstrap-pairs method to estimate VCE	431
13.3.2	The <code>vce(bootstrap)</code> option	432
13.3.3	Bootstrap standard-errors example	432
13.3.4	How many bootstraps?	433
13.3.5	Clustered bootstraps	434
13.3.6	Bootstrap confidence intervals	435
13.3.7	The <code>postestimation estat bootstrap</code> command	436
13.3.8	Bootstrap confidence-intervals example	437
13.3.9	Bootstrap estimate of bias	437
13.4	Bootstrap pairs using the <code>bootstrap</code> command	438
13.4.1	The <code>bootstrap</code> command	438
13.4.2	Bootstrap parameter estimate from a Stata estimation command	439
13.4.3	Bootstrap standard error from a Stata estimation command	440
13.4.4	Bootstrap standard error from a user-written estimation command	440
13.4.5	Bootstrap two-step estimator	441
13.4.6	Bootstrap Hausman test	443
13.4.7	Bootstrap standard error of the coefficient of variation	444
13.5	Bootstraps with asymptotic refinement	445
13.5.1	Percentile-t method	445
13.5.2	Percentile-t Wald test	446
13.5.3	Percentile-t Wald confidence interval	447

13.6	Bootstrap pairs using <code>bsample</code> and <code>simulate</code>	448
13.6.1	The <code>bsample</code> command.	448
13.6.2	The <code>bsample</code> command with <code>simulate</code>	448
13.6.3	Bootstrap Monte Carlo exercise.	450
13.7	Alternative resampling schemes.	450
13.7.1	Bootstrap pairs.	451
13.7.2	Parametric bootstrap.	451
13.7.3	Residual bootstrap.	453
13.7.4	Wild bootstrap.	454
13.7.5	Subsampling.	455
13.8	The <code>jackknife</code>	455
13.8.1	Jackknife method.	455
13.8.2	The <code>vce(jackknife)</code> option and the <code>jackknife</code> command . . .	456
13.9	Stata resources.	456
13.10	Exercises.	456
14	Binary outcome models	459
14.1	Introduction.	459
14.2	Some parametric models.	459
14.2.1	Basic model.	459
14.2.2	Logit, probit, linear probability, and clog-log models	460
14.3	Estimation.	460
14.3.1	Latent-variable interpretation and identification.	461
14.3.2	ML estimation.	461
14.3.3	The <code>logit</code> and <code>probit</code> commands.	462
14.3.4	Robust estimate of the VCE.	462
14.3.5	OLS estimation of LPM.	462
14.4	Example.	463
14.4.1	Data description.	463
14.4.2	Logit regression.	464
14.4.3	Comparison of binary models and parameter estimates . . .	465

- 14.5 Hypothesis and specification tests. 466
 - 14.5.1 Wald tests. 467
 - 14.5.2 Likelihood-ratio tests. 467
 - 14.5.3 Additional model-specification tests. 468
 - Lagrange multiplier test of generalized logit 468
 - Heteroskedastic probit regression. 469
 - 14.5.4 Model comparison. 470
- 14.6 Goodness of fit and prediction. 471
 - 14.6.1 Pseudo- R^2 measure. 471
 - 14.6.2 Comparing predicted probabilities with sample frequencies . 471
 - 14.6.3 Comparing predicted outcomes with actual outcomes 473
 - 14.6.4 The predict command for fitted probabilities. 474
 - 14.6.5 The prvalue command for fitted probabilities. 475
- 14.7 Marginal effects. 476
 - 14.7.1 Marginal effect at a representative value (MER). 476
 - 14.7.2 Marginal effect at the mean (MEM). 477
 - 14.7.3 Average marginal effect (AME). 478
 - 14.7.4 The prchange command. 479
- 14.8 Endogenous regressors. 479
 - 14.8.1 Example. 480
 - 14.8.2 Model assumptions. 481
 - 14.8.3 Structural-model approach. 481
 - The ivprobit command 482
 - Maximum likelihood estimates. 482
 - Two-step sequential estimates. 483
 - 14.8.4 IVs approach 485
- 14.9 Grouped data 486
 - 14.9.1 Estimation with aggregate data 487
 - 14.9.2 Grouped-data application. 487
- 14.10 Stata resources. 489

14.11	Exercises	489
15	Multinomial models	491
15.1	Introduction	491
15.2	Multinomial models overview.	491
15.2.1	Probabilities and MEs.	491
15.2.2	Maximum likelihood estimation.	492
15.2.3	Case-specific and alternative-specific regressors.	493
15.2.4	Additive random-utility model	493
15.2.5	Stata multinomial model commands.	494
15.3	Multinomial example: Choice of fishing mode.	494
15.3.1	Data description	494
15.3.2	Case-specific regressors.	497
15.3.3	Alternative-specific regressors.	497
15.4	Multinomial logit model.	498
15.4.1	The mlogit command	498
15.4.2	Application of the mlogit command.	499
15.4.3	Coefficient interpretation.	500
15.4.4	Predicted probabilities.	501
15.4.5	MEs.	502
15.5	Conditional logit model.	503
15.5.1	Creating long-form data from wide-form data	503
15.5.2	The asclogit command.	505
15.5.3	The clogit command.	506
15.5.4	Application of the asclogit command.	506
15.5.5	Relationship to multinomial logit model.	507
15.5.6	Coefficient interpretation.	507
15.5.7	Predicted probabilities.	508
15.5.8	MEs.	509

15.6	Nested logit model	511
15.6.1	Relaxing the independence of irrelevant alternatives assumption.	511
15.6.2	NL model	511
15.6.3	The nlogit command	512
15.6.4	Model estimates.	513
15.6.5	Predicted probabilities.	516
15.6.6	MEs.	516
15.6.7	Comparison of logit models.	517
15.7	Multinomial probit model.	517
15.7.1	MNP.	517
15.7.2	The mprobit command.	518
15.7.3	Maximum simulated likelihood.	519
15.7.4	The asmprobit command.	519
15.7.5	Application of the asmprobit command.	520
15.7.6	Predicted probabilities and MEs.	522
15.8	Random-parameters logit	522
15.8.1	Random-parameters logit	523
15.8.2	The mixlogit command	523
15.8.3	Data preparation for mixlogit	524
15.8.4	Application of the mixlogit command.	524
15.9	Ordered outcome models.	525
15.9.1	Data summary.	525
15.9.2	Ordered outcomes.	526
15.9.3	Application of the ologit command.	527
15.9.4	Predicted probabilities.	528
15.9.5	MEs.	528
15.9.6	Other ordered models.	529
15.10	Multivariate outcomes.	529
15.10.1	Bivariate probit.	529

15.10.2	Nonlinear SUR	532
15.11	Stata resources	532
15.12	Exercises	533
16	Tobit and selection models	535
16.1	Introduction	535
16.2	Tobit model	535
16.2.1	Regression with censored data	535
16.2.2	Tobit model setup	536
16.2.3	Unknown censoring point	537
16.2.4	Tobit estimation	537
16.2.5	ML estimation in Stata	538
16.3	Tobit model example	538
16.3.1	Data summary	538
16.3.2	Tobit analysis	539
16.3.3	Prediction after tobit	540
16.3.4	Marginal effects	541
Left-truncated, left-censored, and right-truncated examples	541	
Left-censored case computed directly	542	
Marginal impact on probabilities	543	
16.3.5	The ivtobit command	544
16.3.6	Additional commands for censored regression	545
16.4	Tobit for lognormal data	545
16.4.1	Data example	546
16.4.2	Setting the censoring point for data in logs	546
16.4.3	Results	547
16.4.4	Two-limit tobit	548
16.4.5	Model diagnostics	549
16.4.6	Tests of normality and homoskedasticity	550
Generalized residuals and scores	550	
Test of normality	551	

	Test of homoskedasticity.	552
16.4.7	Next step?.	553
16.5	Two-part model in logs.	553
16.5.1	Model structure.	553
16.5.2	Part 1 specification.	554
16.5.3	Part 2 of the two-part model.	555
16.6	Selection model.	556
16.6.1	Model structure and assumptions.	556
16.6.2	ML estimation of the sample-selection model.	558
16.6.3	Estimation without exclusion restrictions.	558
16.6.4	Two-step estimation.	560
16.6.5	Estimation with exclusion restrictions.	561
16.7	Prediction from models with outcome in logs.	562
16.7.1	Predictions from tobit.	563
16.7.2	Predictions from two-part model.	564
16.7.3	Predictions from selection model.	565
16.8	Stata resources.	565
16.9	Exercises.	566
17	Count-data models	567
17.1	Introduction.	567
17.2	Features of count data.	567
17.2.1	Generated Poisson data.	568
17.2.2	Overdispersion and negative binomial data.	569
17.2.3	Modeling strategies.	570
17.2.4	Estimation methods.	571
17.3	Empirical example 1.	571
17.3.1	Data summary.	571
17.3.2	Poisson model.	572
	Poisson model results.	573
	Robust estimate of VCE for Poisson MLE.	574

	Test of overdispersion	575
	Coefficient interpretation and marginal effects	576
17.3.3	NB2 model	577
	NB2 model results	577
	Fitted probabilities for Poisson and NB2 models	579
	The countfit command	579
	The prvalue command	581
	Discussion	581
	Generalized NB model	581
17.3.4	Nonlinear least-squares estimation	582
17.3.5	Hurdle model	583
	Variants of the hurdle model	585
	Application of the hurdle model	585
17.3.6	Finite-mixture models	589
	FMM specification	589
	Simulated FMM sample with comparisons	589
	ML estimation of the FMM	591
	The fmm command	592
	Application: Poisson finite-mixture model	592
	Interpretation	593
	Comparing marginal effects	594
	Application: NB finite-mixture model	596
	Model selection	598
	Cautionary note	599
17.4	Empirical example 2	599
17.4.1	Zero-inflated data	599
17.4.2	Models for zero-inflated data	600
17.4.3	Results for the NB2 model	601
	The prcounts command	602
17.4.4	Results for ZINB	603

17.4.5	Model comparison604
	The countfit command604
	Model comparison using countfit604
17.5	Models with endogenous regressors.605
17.5.1	Structural-model approach606
	Model and assumptions.606 •
	Two-step estimation.607
	Application.607
17.5.2	Nonlinear IV method.610
17.6	Stata resources.611
17.7	Exercises.612
18	Nonlinear panel models	.615
18.1	Introduction.615
18.2	Nonlinear panel-data overview.615
18.2.1	Some basic nonlinear panel models.615
	FE models.616
	RE models.616
	Pooled models or population-averaged models.616
	Comparison of models.617
18.2.2	Dynamic models.617
18.2.3	Stata nonlinear panel commands.617
18.3	Nonlinear panel-data example.618
18.3.1	Data description and summary statistics.618
18.3.2	Panel-data organization.620
18.3.3	Within and between variation.620
18.3.4	FE or RE model for these data?.621
18.4	Binary outcome models.621
18.4.1	Panel summary of the dependent variable.621
18.4.2	Pooled logit estimator.622
18.4.3	The xtlogit command.623

18.4.4	The xtgee command	624
18.4.5	PA logit estimator.	624
18.4.6	RE logit estimator.	625
18.4.7	FE logit estimator.	627
18.4.8	Panel logit estimator comparison.	629
18.4.9	Prediction and marginal effects.	630
18.4.10	Mixed-effects logit estimator.	630
18.5	Tobit model.	631
18.5.1	Panel summary of the dependent variable.	631
18.5.2	RE tobit model.	631
18.5.3	Generalized tobit models.	632
18.5.4	Parametric nonlinear panel models.	633
18.6	Count-data models.	633
18.6.1	The xtpoisson command.	633
18.6.2	Panel summary of the dependent variable.	634
18.6.3	Pooled Poisson estimator.	634
18.6.4	PA Poisson estimator.	635
18.6.5	RE Poisson estimators.	636
18.6.6	FE Poisson estimator.	638
18.6.7	Panel Poisson estimators comparison.	640
18.6.8	Negative binomial estimators.	641
18.7	Stata resources.	642
18.8	Exercises.	643
A	Programming in Stata	645
A.I	Stata matrix commands.	645
A.1.1	Stata matrix overview.	645
A.1.2	Stata matrix input and output	645
	Matrix input by hand.	645
	Matrix input from Stata estimation results.	646
A.1.3	Stata matrix subscripts and combining matrices.	647

A.1.4	Matrix operators	648
A.1.5	Matrix functions	648
A.1.6	Matrix accumulation commands	649
A.1.7	OLS using Stata matrix commands	650
A.2	Programs	651
A.2.1	Simple programs (no arguments or access to results)	651 •
A.2.2	Modifying a program	652
A.2.3	Programs with positional arguments	652
A.2.4	Temporary variables	653
A.2.5	Programs with named positional arguments	653
A.2.6	Storing and retrieving program results	654
A.2.7	Programs with arguments using standard Stata syntax	655
A.2.8	Ado-files	656
A.3	Program debugging	657
A.3.1	Some simple tips	658
A.3.2	Error messages and return code	658
A.3.3	Trace	659
B	Mata	661
B.I	How to run Mata	661
B.I.1	Mata commands in Mata	661
B.I.2	Mata commands in Stata	662
B.I.3	Stata commands in Mata	662
B.I.4	Interactive versus batch use	662
B.I.5	Mata help	662
B.2	Mata matrix commands	663
B.2.1	Mata matrix input	663
	Matrix input by hand	663
	Identity matrices, unit vectors, and matrices of constants	664
	Matrix input from Stata data	665
	Matrix input from Stata matrix	665

	Stata interface functions.	666
B.2.2	Mata matrix operators.	666
	Element-by-element operators.	666
B.2.3	Mata functions.	667
	Scalar and matrix functions.	667
	Matrix inversion.	668
B.2.4	Mata cross products.	669
B.2.5	Mata matrix subscripts and combining matrices.	669
B.2.6	Transferring Mata data and matrices to Stata.	671
	Creating Stata matrices from Mata matrices.	671
	Creating Stata data from a Mata vector.	671
B.3	Programming in Mata.	672
B.3.1	Declarations.	672
B.3.2	Mata program.	672
B.3.3	Mata program with results output to Stata.	673
B.3.4	Stata program that calls a Mata program.	673
B.3.5	Using Mata in ado-files.	674
	Glossary of abbreviations	675
	References	679
	Author index	687
	Subject index	691