

Microeconometrics Using Stata

Volume I: Cross-Sectional and Panel Regression Methods

Second Edition

A. COLIN CAMERON

*Department of Economics
University of California, Davis, CA*

and
*School of Economics
University of Sydney, Sydney, Australia*

PRAVIN K. TRIVEDI

*School of Economics
University of Queensland, Brisbane, Australia*

and
*Department of Economics
Indiana University, Bloomington, IN*



STATA Press

A Stata Press Publication
StataCorp LLC
College Station, Texas

136
Oth.
compa. 43

Contents

	List of tables	xiii
	List of figures	xv
	Preface to the Second Edition	xvii
	Preface to the First Edition	xix
1	Stata basics	1
	1.1 Interactive use	1
	1.2 Documentation	2
	1.3 Command syntax and operators	5
	1.4 Do-files and log files	14
	1.5 Scalars and matrices	19
	1.6 Using results from Stata commands	20
	1.7 Global and local macros	23
	1.8 Looping commands	26
	1.9 Mata and Python in Stata	29
	1.10 Some useful commands	29
	1.11 Template do-file	30
	1.12 Community-contributed commands	30
	1.13 Additional resources	31
	1.14 Exercises	31
2	Data management and graphics	33
	2.1 Introduction	33
	2.2 Types of data	33
	2.3 Inputting data	36
	2.4 Data management	43

2.5	Manipulating datasets	60
2.6	Graphical display of data	67
2.7	Additional resources	83
2.8	Exercises	83
3	Linear regression basics	85
3.1	Introduction	85
3.2	Data and data summary	85
3.3	Transformation of data before regression	94
3.4	Linear regression	96
3.5	Basic regression analysis	102
3.6	Specification analysis	123
3.7	Specification tests	132
3.8	Sampling weights	140
3.9	OLS using Mata	145
3.10	Additional resources	147
3.11	Exercises	147
4	Linear regression extensions	149
4.1	Introduction	149
4.2	In-sample prediction	149
4.3	Out-of-sample prediction	157
4.4	Predictive margins	161
4.5	Marginal effects	175
4.6	Regression decomposition analysis	186
4.7	Shapley decomposition of relative regressor importance	193
4.8	Difference-in-differences estimators	195
4.9	Additional resources	204
4.10	Exercises	204
5	Simulation	207
5.1	Introduction	207
5.2	Pseudorandom-number generators	208

5.3	Distribution of the sample mean	214
5.4	Pseudorandom-number generators: Further details	220
5.5	Computing integrals	227
5.6	Simulation for regression: Introduction	232
5.7	Additional resources	242
5.8	Exercises	242
6	Linear regression with correlated errors	245
6.1	Introduction	245
6.2	Generalized least-squares and FGLS regression	246
6.3	Modeling heteroskedastic data	250
6.4	OLS for clustered data	256
6.5	FGLS estimators for clustered data	265
6.6	Fixed-effects estimator for clustered data	269
6.7	Linear mixed models for clustered data	277
6.8	Systems of linear regressions	286
6.9	Survey data: Weighting, clustering, and stratification	295
6.10	Additional resources	301
6.11	Exercises	302
7	Linear instrumental-variables regression	305
7.1	Introduction	305
7.2	Simultaneous equations model	306
7.3	Instrumental-variables regression	310
7.4	Instrumental-variables example	316
7.5	Weak instruments	330
7.6	Diagnostics and tests for weak instruments	339
7.7	Inference with weak instruments	353
7.8	Finite sample inference with weak instruments	362
7.9	Other estimators	363
7.10	Three-stage least-squares systems estimation	367

7.11	Additional resources	368
7.12	Exercises	369
8	Linear panel-data models: Basics	373
8.1	Introduction	373
8.2	Panel-data methods overview	373
8.3	Summary of panel data	379
8.4	Pooled or population-averaged estimators	394
8.5	Fixed-effects or within estimator	397
8.6	Between estimator	401
8.7	Random-effects estimator	402
8.8	Comparison of estimators	406
8.9	First-difference estimator	412
8.10	Panel-data management	414
8.11	Additional resources	418
8.12	Exercises	419
9	Linear panel-data models: Extensions	421
9.1	Introduction	421
9.2	Panel instrumental-variables estimation	421
9.3	Hausman–Taylor estimator	425
9.4	Arellano–Bond estimator	428
9.5	Long panels	445
9.6	Additional resources	456
9.7	Exercises	456
10	Introduction to nonlinear regression	459
10.1	Introduction	459
10.2	Binary outcome models	459
10.3	Probit model	462
10.4	MEs and coefficient interpretation	466
10.5	Logit model	472
10.6	Nonlinear least squares	474

10.7	Other nonlinear estimators	476
10.8	Additional resources	477
10.9	Exercises	477
11	Tests of hypotheses and model specification	479
11.1	Introduction	479
11.2	Critical values and p-values	480
11.3	Wald tests and confidence intervals	485
11.4	Likelihood-ratio tests	498
11.5	Lagrange multiplier test (or score test)	502
11.6	Multiple testing	505
11.7	Test size and power	512
11.8	The power onemean command for multiple regression	519
11.9	Specification tests	529
11.10	Permutation tests and randomization tests	532
11.11	Additional resources	534
11.12	Exercises	534
12	Bootstrap methods	537
12.1	Introduction	537
12.2	Bootstrap methods	537
12.3	Bootstrap pairs using the vce(bootstrap) option	539
12.4	Bootstrap pairs using the bootstrap command	547
12.5	Percentile-t bootstraps with asymptotic refinement	555
12.6	Wild bootstrap with asymptotic refinement	560
12.7	Bootstrap pairs using bsample and simulate	569
12.8	Alternative resampling schemes	570
12.9	The jackknife	575
12.10	Additional resources	576
12.11	Exercises	577

13	Nonlinear regression methods	579
13.1	Introduction	579
13.2	Nonlinear example: Doctor visits	580
13.3	Nonlinear regression methods	582
13.4	Different estimates of the VCE	597
13.5	Prediction	604
13.6	Predictive margins	609
13.7	Marginal effects	612
13.8	Model diagnostics	629
13.9	Clustered data	632
13.10	Additional resources	640
13.11	Exercises	640
14	Flexible regression: Finite mixtures and nonparametric	643
14.1	Introduction	643
14.2	Models based on finite mixtures	644
14.3	FMM example: Earnings of doctors	650
14.4	Global polynomials	665
14.5	Regression splines	668
14.6	Nonparametric regression	675
14.7	Partially parametric regression	680
14.8	Additional resources	681
14.9	Exercises	681
15	Quantile regression	683
15.1	Introduction	683
15.2	Conditional quantile regression	684
15.3	CQR for medical expenditures data	688
15.4	CQR for generated heteroskedastic data	699
15.5	Quantile treatment effects for a binary treatment	703
15.6	Additional resources	706
15.7	Exercises	707

A	Programming in Stata	709
A.1	Stata matrix commands	709
A.2	Programs	716
A.3	Program debugging	722
A.4	Additional resources	725
B	Mata	727
B.1	How to run Mata	727
B.2	Mata matrix commands	729
B.3	Programming in Mata	738
B.4	Additional resources	740
C	Optimization in Mata	741
C.1	Mata moptimize() function	741
C.2	Mata optimize() function	751
C.3	Additional resources	754
	Glossary of abbreviations	755
	References	761
	Author index	777
	Subject index	783