



C. Ciccarelli, G. Vecchi

Empirical Econometrics with Stata Vol. I

Revised December 1, 2004

Preliminary draft. Comments, questions and suggestions are welcome. Contact Giovanni Vecchi, Dipartimento di Economia e Istituzioni, Università di Roma “Tor Vergata”. E-mail: giovanni.vecchi@uniroma2.it

This manuscript can be printed for individual use, but may not be printed or circulated for commercial purposes. Not to be quoted without the author's permission.



Contents

Introduction	xi
I MASTERING STATA	1
1 First Steps in Stata	1
1.1 Getting In and Out	2
1.2 Stata and the Operating System	6
1.3 Help and Learning Tools	10
1.4 Stata is a Language, it has a Syntax	15
1.5 Can You Answer your Boss's Questions?	24
2 Dataset Management	28
2.1 Loading Datasets	29
2.2 Exporting Data	39
2.3 A First Look to the Data	41
2.4 Shaping Datasets	47
2.5 Merging and Appending	51
2.6 Sample Session	54
3 Graphs	55
3.1 Introduction	55
3.2 Stata's Philosophy	56
3.3 Obtaining Graphs Fairly Quickly	57

3.4	How close can you get to my graphs?	63
4	Data Analysis	69
4.1	Handling Variables	70
4.2	Tables	85
4.3	Basic Descriptives	90
4.4	How Close Can You Get to my Tables?	103
5	Getting Started with Programming in Stata	107
5.1	A Few Ingredients and a Recipe for Stata Programming . . .	108
5.2	Do-files Vs. Ado-files	110
5.3	Macros and Scalars	119
5.4	Let's Play Consultants	124
6	Control Flow Commands in Stata	131
6.1	Gold-Plated Rules for Stata Users	132
6.2	Branching and Looping	135
6.3	Temporary Objects in Stata	141
6.4	Pythagoras and Probability Tables: Do-It-Yourself Session . .	144
6.5	Computer Exercises	147
7	Advanced Programming	148
7.1	What Do Stata Commands Hide?	149
7.2	Digging Deeper into Macros	154
7.3	Parsing the Syntax	162
7.4	Matrices, Vectors, and the Like	165
7.5	Posting	169
7.6	The Future Is Getting Near: Stata Plugins	169
7.7	Computer Exercises	169
II	BASIC TOPICS	171
8	The Linear Regression Model	172
8.1	The LRM: A Review	173
8.2	Running Regressions in Stata	189
8.3	After Estimation: What Everyone Should Know	201
8.4	Computer Exercises	206
9	How to Interpret Regression Results	209
9.1	The Level–Level Model	210
9.2	The Log–Level Model	213
9.3	The Log–Log Model	219
9.4	The Level–Log Model	219
9.5	Dummy Variables	219
10	Regression Diagnostics and Hypothesis Testing	229

10.1	From the Top Floor to the Basement	229
10.2	Outliers, Influential Observations and the Like	231
10.3	On the many faces of the OLS residuals	235
10.4	Influence and Leverage	240
10.5	Hypothesis Testing	242
10.6	Inference in the LRM in the Presence of Heteroskedasticity	261
10.7	Testing Heteroskedasticity	264
10.8	Is my regression good enough? - Sample session	267
10.9	Appendix	268
10.10	Data Appendix	269
10.11	Proof of Theorem 10.1 (OLS conditional normality)	269
10.12	Derivation of F-statistic	270
10.13	Further Readings	270
10.14	Computer exercises	271
11	Instrumental Variables	272
11.1	Measurement errors	272
11.2	Instrumental Variables and Two Stage Least Squares	272
11.3	Computer exercises	272
12	Models for Categorical Outcome Data	274
12.1	Binary Data	276
12.2	Ordinal Data	286
12.3	Nominal Data	286
12.4	Count Data Models	287
III	ADVANCED TOPICS	295
13	Models for Limited Dependent Data	296
13.1	Corner Solution Models	297
13.2	Censoring, Truncation and the LRM: A Motivation	311
13.3	Censored Regression Models	315
13.4	Truncation	316
13.5	Sample Selection Models	321
13.6	Computer exercises	331
14	Bootstrap and Monte Carlo	332
14.1	A 15-Minute Refresher Course on Resampling Techniques	332
14.2	Bootstrapping, How To	335
14.3	Monte Carlo Simulations	345
14.4	About seeds and Reproducibility	349
14.5	Computer Exercises	352
15	Nonparametric Methods	354
15.1	Density Estimation	355
15.2	Nonparametric Regression	368

15.3 Nonparametric Tests	382
15.4 Computer Exercises	382
16 Robust Econometrics	384
16.1 Introduction	384
16.2 M-Estimators	385
16.3 L-Estimators	388
16.4 The Generalized Linear Model	388
IV DATA SPECIALITIES	389
17 Complex Survey Data	390
17.1 Sampling	391
17.2 Weights: When to Use Them	396
17.3 Hands-on Survey Data	400
18 Time-Series Data	406
18.1 Time and Date Variables	406
18.2 Hands-on Time Series Datasets	411
18.3 ARIMA Processes, Estimation Of	412
18.4 Integration	417
18.5 ARCH and GARCH Models	417
19 Time-Series Data: Advanced Topics	421
19.1 Introduction	421
19.2 Distribute Lag Models	421
19.3 Cointegration	421
19.4 The Error Correction Model	421
19.5 Vector Autoregressions	421
20 Cross-Sectional Time-Series	423
20.1 Introduction	423
20.2 Descriptive Statistics for xt-data	425
20.3 Linear Regression Models with Individual Effects	429
20.4 Fixed vs Random Effects: the Hausman Test	437
20.5 Dynamic Linear Models	440
20.6 Nonlinear Models	440
20.7 Fixed Effects Models for Binary Data	446
20.8 Suggested Empirical Applications	446
20.9 Panel Data on the Web	447
Keys to Computer Exercises	448
20.10Chapter 3	448
20.11Chapter ??	450
20.12Chapter 14	456
References	457