

# Contents

	List of figures	xvii
	List of tables	xix
	Preface	xxi
	Acknowledgments	xxiii
	Notation and typography	xxv
1	Why should you become a Stata programmer?	1
	Do-file programming . . . . .	2
	Ado-file programming . . . . .	2
	Mata programming for ado-files . . . . .	2
1.1	Plan of the book . . . . .	3
1.2	Installing the necessary software . . . . .	4
2	Some elementary concepts and tools	5
2.1	Introduction . . . . .	5
	2.1.1 What you should learn from this chapter . . . . .	5
2.2	Navigational and organizational issues . . . . .	5
	2.2.1 The current working directory and profile.do . . . . .	6
	2.2.2 Locating important directories: sysdir and adopath . . . . .	6
	2.2.3 Organization of do-files, ado-files, and data files . . . . .	7
2.3	Editing Stata do- and ado-files . . . . .	8
2.4	Data types . . . . .	9
	2.4.1 Storing data efficiently: The compress command . . . . .	11
	2.4.2 Date and time handling . . . . .	11
	2.4.3 Time-series operators . . . . .	13
	2.4.4 Factor variables and operators . . . . .	14

2.5	Handling errors: The capture command . . . . .	16
2.6	Protecting the data in memory: The preserve and restore commands	17
2.7	Getting your data into Stata . . . . .	18
2.7.1	Inputting and importing data . . . . .	18
	Handling text files . . . . .	19
	Free format versus fixed format . . . . .	20
	The import delimited command . . . . .	21
	Accessing data stored in spreadsheets . . . . .	23
	Fixed-format data files . . . . .	24
2.7.2	Importing data from other package formats . . . . .	29
2.8	Guidelines for Stata do-file programming style . . . . .	30
2.8.1	Basic guidelines for do-file writers . . . . .	31
2.8.2	Enhancing speed and efficiency . . . . .	33
2.9	How to seek help for Stata programming . . . . .	33
<b>3</b>	<b>Do-file programming: Functions, macros, scalars, and matrices</b>	<b>37</b>
3.1	Introduction . . . . .	37
3.1.1	What you should learn from this chapter . . . . .	37
3.2	Some general programming details . . . . .	38
3.2.1	The varlist . . . . .	39
3.2.2	The numlist . . . . .	39
3.2.3	The if exp and in range qualifiers . . . . .	39
3.2.4	Missing-data handling . . . . .	40
	Recoding missing values: The mvdecode and mvencode commands . . . . .	41
3.2.5	String-to-numeric conversion and vice versa . . . . .	42
	Numeric-to-string conversion . . . . .	43
	Working with quoted strings . . . . .	44
3.3	Functions for the generate command . . . . .	44
3.3.1	Using if exp with indicator variables . . . . .	47
3.3.2	The cond() function . . . . .	49

3.3.3	Recoding discrete and continuous variables . . . . .	49
3.4	Functions for the egen command . . . . .	51
	Official egen functions . . . . .	52
	egen functions from the user community . . . . .	53
3.5	Computation for by-groups . . . . .	54
3.5.1	Observation numbering: <code>_n</code> and <code>_N</code> . . . . .	55
3.6	Local macros . . . . .	57
3.7	Global macros . . . . .	60
3.8	Extended macro functions and macro list functions . . . . .	60
3.8.1	System parameters, settings, and constants: <code>creturn</code> . . . . .	62
3.9	Scalars . . . . .	62
3.10	Matrices . . . . .	64
<b>4</b>	<b>Cookbook: Do-file programming I</b>	<b>67</b>
4.1	Tabulating a logical condition across a set of variables . . . . .	67
4.2	Computing summary statistics over groups . . . . .	69
4.3	Computing the extreme values of a sequence . . . . .	70
4.4	Computing the length of spells . . . . .	71
4.5	Summarizing group characteristics over observations . . . . .	76
4.6	Using global macros to set up your environment . . . . .	78
4.7	List manipulation with extended macro functions . . . . .	79
4.8	Using <code>creturn</code> values to document your work . . . . .	81
<b>5</b>	<b>Do-file programming: Validation, results, and data management</b>	<b>83</b>
5.1	Introduction . . . . .	83
5.1.1	What you should learn from this chapter . . . . .	83
5.2	Data validation: The <code>assert</code> , <code>count</code> , and <code>duplicates</code> commands . . . . .	83
5.3	Reusing computed results: The <code>return</code> and <code>ereturn</code> commands . . . . .	90
5.3.1	The <code>ereturn list</code> command . . . . .	94
5.4	Storing, saving, and using estimated results . . . . .	97
5.4.1	Generating publication-quality tables from stored estimates . . . . .	102
5.5	Reorganizing datasets with the <code>reshape</code> command . . . . .	104

5.6	Combining datasets . . . . .	109
5.7	Combining datasets with the append command . . . . .	111
5.8	Combining datasets with the merge command . . . . .	113
5.8.1	The one-to-one match-merge . . . . .	115
5.8.2	The dangers of many-to-many merges . . . . .	116
5.9	Other data management commands . . . . .	117
5.9.1	The fillin command . . . . .	117
5.9.2	The cross command . . . . .	117
5.9.3	The stack command . . . . .	118
5.9.4	The separate command . . . . .	119
5.9.5	The joinby command . . . . .	120
5.9.6	The xpose command . . . . .	121
<b>6</b>	<b>Cookbook: Do-file programming II</b>	<b>123</b>
6.1	Efficiently defining group characteristics and subsets . . . . .	123
6.1.1	Using a complicated criterion to select a subset of observations	124
6.2	Applying reshape repeatedly . . . . .	125
6.3	Handling time-series data effectively . . . . .	129
6.3.1	Working with a business-daily calendar . . . . .	132
6.4	reshape to perform rowwise computation . . . . .	133
6.5	Adding computed statistics to presentation-quality tables . . . . .	136
6.6	Presenting marginal effects rather than coefficients . . . . .	138
6.6.1	Graphing marginal effects with marginsplot . . . . .	140
6.7	Generating time-series data at a lower frequency . . . . .	141
6.8	Using suest and gsem to compare estimates from nonoverlapping samples . . . . .	146
6.9	Using reshape to produce forecasts from a VAR or VECM . . . . .	149
6.10	Working with IRF files . . . . .	152
<b>7</b>	<b>Do-file programming: Prefixes, loops, and lists</b>	<b>157</b>
7.1	Introduction . . . . .	157
7.1.1	What you should learn from this chapter . . . . .	157



7.2	Prefix commands . . . . .	157
7.2.1	The by prefix . . . . .	158
7.2.2	The statsby prefix . . . . .	160
7.2.3	The xi prefix and factor-variable notation . . . . .	161
7.2.4	The rolling prefix . . . . .	162
7.2.5	The simulate and permute prefixes . . . . .	164
7.2.6	The bootstrap and jackknife prefixes . . . . .	167
7.2.7	Other prefix commands . . . . .	169
7.3	The forvalues and foreach commands . . . . .	169
<b>8</b>	<b>Cookbook: Do-file programming III</b>	<b>177</b>
8.1	Handling parallel lists . . . . .	177
8.2	Calculating moving-window summary statistics . . . . .	178
8.2.1	Producing summary statistics with rolling and merge . . . . .	180
8.2.2	Calculating moving-window correlations . . . . .	181
8.3	Computing monthly statistics from daily data . . . . .	182
8.4	Requiring at least n observations per panel unit . . . . .	184
8.5	Counting the number of distinct values per individual . . . . .	185
8.6	Importing multiple spreadsheet pages . . . . .	186
<b>9</b>	<b>Do-file programming: Other topics</b>	<b>189</b>
9.1	Introduction . . . . .	189
9.1.1	What you should learn from this chapter . . . . .	189
9.2	Storing results in Stata matrices . . . . .	189
9.3	The post and postfile commands . . . . .	193
9.4	Output: The export delimited, outfile, and file commands . . . . .	196
9.5	Automating estimation output . . . . .	199
9.6	Automating graphics . . . . .	203
9.7	Characteristics . . . . .	207
<b>10</b>	<b>Cookbook: Do-file programming IV</b>	<b>211</b>
10.1	Computing firm-level correlations with multiple indices . . . . .	211
10.2	Computing marginal effects for graphical presentation . . . . .	214

10.3	Automating the production of L <sup>A</sup> T <sub>E</sub> X tables . . . . .	216
10.4	Extracting data from graph files' sersets . . . . .	220
10.5	Constructing continuous price and returns series . . . . .	225
<b>11</b>	<b>Ado-file programming</b>	<b>231</b>
11.1	Introduction . . . . .	231
11.1.1	What you should learn from this chapter . . . . .	232
11.2	The structure of a Stata program . . . . .	232
11.3	The program statement . . . . .	233
11.4	The syntax and return statements . . . . .	234
11.5	Implementing program options . . . . .	237
11.6	Including a subset of observations . . . . .	238
11.7	Generalizing the command to handle multiple variables . . . . .	240
11.8	Making commands byable . . . . .	242
	Program properties . . . . .	243
11.9	Documenting your program . . . . .	244
11.10	egen function programs . . . . .	246
11.11	Writing an e-class program . . . . .	248
11.11.1	Defining subprograms . . . . .	250
11.12	Certifying your program . . . . .	250
11.13	Programs for ml, nl, and nlsur . . . . .	252
	Maximum likelihood estimation of distributions' parameters	255
11.13.1	Writing an ml-based command . . . . .	260
11.13.2	Programs for the nl and nlsur commands . . . . .	263
11.14	Programs for gmm . . . . .	265
11.15	Programs for the simulate, bootstrap, and jackknife prefixes . . . . .	270
11.16	Guidelines for Stata ado-file programming style . . . . .	272
11.16.1	Presentation . . . . .	273
11.16.2	Helpful Stata features . . . . .	274
11.16.3	Respect for datasets . . . . .	274
11.16.4	Speed and efficiency . . . . .	275

11.16.5	Reminders . . . . .	275
11.16.6	Style in the large : . . . . .	276
11.16.7	Use the best tools . . . . .	276
<b>12</b>	<b>Cookbook: Ado-file programming</b>	<b>277</b>
12.1	Retrieving results from rolling . . . . .	277
12.2	Generalization of egen function pct9010() to support all pairs of quantiles . . . . .	280
12.3	Constructing a certification script . . . . .	282
12.4	Using the ml command to estimate means and variances . . . . .	287
12.4.1	Applying equality constraints in ml estimation . . . . .	289
12.5	Applying inequality constraints in ml estimation . . . . .	291
12.6	Generating a dataset containing the longest spell . . . . .	294
12.7	Using suest on a fixed-effects model . . . . .	297
<b>13</b>	<b>Mata functions for do-file and ado-file programming</b>	<b>301</b>
13.1	Mata: First principles . . . . .	301
13.1.1	What you should learn from this chapter . . . . .	302
13.2	Mata fundamentals . . . . .	302
13.2.1	Operators . . . . .	303
13.2.2	Relational and logical operators . . . . .	304
13.2.3	Subscripts . . . . .	305
13.2.4	Populating matrix elements . . . . .	305
13.2.5	Mata loop commands . . . . .	307
13.2.6	Conditional statements . . . . .	308
13.3	Mata's st_ interface functions . . . . .	309
13.3.1	Data access . . . . .	309
13.3.2	Access to locals, globals, scalars, and matrices . . . . .	311
13.3.3	Access to Stata variables' attributes . . . . .	312
13.4	Calling Mata with a single command line . . . . .	312
13.5	Components of a Mata function . . . . .	316
13.5.1	Arguments . . . . .	316

13.5.2	Variables . . . . .	317
13.5.3	Stored results . . . . .	317
13.6	Calling Mata functions . . . . .	318
13.7	Example: <code>st_</code> interface function usage . . . . .	320
13.8	Example: Matrix operations . . . . .	322
13.8.1	Extending the command . . . . .	327
13.9	Mata-based likelihood function evaluators . . . . .	329
13.10	Creating arrays of temporary objects with pointers . . . . .	331
13.11	Structures . . . . .	334
13.12	Additional Mata features . . . . .	337
13.12.1	Macros in Mata functions . . . . .	337
13.12.2	Associative arrays in Mata functions . . . . .	338
13.12.3	Compiling Mata functions . . . . .	340
13.12.4	Building and maintaining an object library . . . . .	341
13.12.5	A useful collection of Mata routines . . . . .	342
<b>14</b>	<b>Cookbook: Mata function programming</b>	<b>343</b>
14.1	Reversing the rows or columns of a Stata matrix . . . . .	343
14.2	Shuffling the elements of a string variable . . . . .	346
14.3	Firm-level correlations with multiple indices with Mata . . . . .	348
14.4	Passing a function to a Mata function . . . . .	353
14.5	Using subviews in Mata . . . . .	356
14.6	Storing and retrieving country-level data with Mata structures . . . . .	358
14.7	Locating nearest neighbors with Mata . . . . .	363
14.8	Using a permutation vector to reorder results . . . . .	368
14.9	Producing $\text{\LaTeX}$ tables from <code>svy</code> results . . . . .	370
14.10	Computing marginal effects for quantile regression . . . . .	375
14.11	Computing the seemingly unrelated regression estimator . . . . .	379
14.12	A GMM-CUE estimator using Mata's <code>optimize()</code> functions . . . . .	384



References	397
Author index	403
Subject index	405