Contents

List of figures xix

Preface xxi

I General information 1

1 Introduction 7

1.1 What is this book about? . . . 7

1.2 Which models are considered? . . . 8

1.3 Whom is this book for? . . . 9

1.4 How is the book organized? . . . 9

1.5 The SPost software . . . 11

1.5.1 Updating Stata . . . 12

1.5.2 Installing SPost13 . . . 13

Uninstalling SPost9 . . . 14

Installing SPost13 using search . . . 14

Installing SPost13 using net install . . . 16

1.5.3 Uninstalling SPost13 . . . 17

1.6 Sample do-files and datasets . . . 17

1.6.1 Installing the spost13_do package . . . 17

1.6.2 Using spex to load data and run examples . . . 17

1.7 Getting help with SPost . . . 18

1.7.1 What if an SPost command does not work? . . . 18

1.7.2 Getting help from the authors . . . 19

What we need to help you . . . 20

1.8 Where can I learn more about the models? . . . 21

2 Introduction to Stata 23

2.1 The Stata interface 23
2.2 Abbreviations 27
2.3 Getting help 27
2.3.1 Online help 27
2.3.2 PDF manuals 28
2.3.3 Error messages 28
2.3.4 Asking for help 28
2.3.5 Other resources 29
2.4 The working directory 29
2.5 Stata file types 30
2.6 Saving output to log files 30
2.7 Using and saving datasets 32
2.7.1 Data in Stata format 32
2.7.2 Data in other formats 33
2.7.3 Entering data by hand 33
2.8 Size limitations on datasets 34
2.9 Do-files 34
2.9.1 Adding comments 35
2.9.2 Long lines 36
2.9.3 Stopping a do-file while it is running 37
2.9.4 Creating do-files 37
2.9.5 Recommended structure for do-files 38
2.10 Using Stata for serious data analysis 40
2.11 Syntax of Stata commands 41
2.11.1 Commands 43
2.11.2 Variable lists 43
2.11.3 if and in qualifiers 45
2.11.4 Options 46
2.12 Managing data 46
2.12.1 Looking at your data 46

2.12.2 Getting information about variables 47
2.12.3 Missing values 50
2.12.4 Selecting observations 51
2.12.5 Selecting variables 51
2.13 Creating new variables 52
2.13.1 The generate command 52
2.13.2 The replace command 54
2.13.3 The recode command 55
2.14 Labeling variables and values 56
2.14.1 Variable labels 56
2.14.2 Value labels 57
2.14.3 The notes command 59
2.15 Global and local macros 59
2.16 Loops using foreach and forvalues 61
2.17 Graphics 63
2.17.1 The graph command 65
2.18 A brief tutorial 73
2.19 A do-file template 79
2.20 Conclusion 81
3 Estimation, testing, and fit **83**
3.1 Estimation 84
3.1.1 Stata's output for ML estimation 84
3.1.2 ML and sample size 85
3.1.3 Problems in obtaining ML estimates 85
3.1.4 Syntax of estimation commands 86
3.1.5 Variable lists 87
Using factor-variable notation in the variable list 87
Specifying interaction and polynomials 89
More on factor-variable notation 90
3.1.6 Specifying the estimation sample 93

Missing data . . . 93
Information about missing values . . . 95
Postestimation commands and the estimation sample . . . 98
3.1.7 Weights and survey data . . . 99
Complex survey designs . . . 100
3.1.8 Options for regression models . . . 102
3.1.9 Robust standard errors . . . 103
3.1.10 Reading the estimation output . . . 105
3.1.11 Storing estimation results . . . 107
(Advanced) Saving estimates to a file . . . 108
3.1.12 Reformatting output with estimates table . . . 111
3.2 Testing . . . 114
3.2.1 One-tailed and two-tailed tests . . . 115
3.2.2 Wald and likelihood-ratio tests . . . 115
3.2.3 Wald tests with test and testparm . . . 116
3.2.4 LR tests with lrtest . . . 118
Avoiding invalid LR tests . . . 120
3.3 Measures of fit . . . 120
3.3.1 Syntax of fitstat . . . 120
3.3.2 Methods and formulas used by fitstat . . . 123
3.3.3 Example of fitstat . . . 129
3.4 estat postestimation commands . . . 130
3.5 Conclusion . . . 131
4 Methods of interpretation **133**
4.1 Comparing linear and nonlinear models . . . 133
4.2 Approaches to interpretation . . . 136
4.2.1 Method of interpretation based on predictions . . . 137
4.2.2 Method of interpretation using parameters . . . 138
4.2.3 Stata and SPost commands for interpretation . . . 138
4.3 Predictions for each observation . . . 138

4.4 Predictions at specified values 139
4.4.1 Why use the m* commands instead of margins? 140
4.4.2 Using margins for predictions 141
Predictions using interaction and polynomial terms 146
Making multiple predictions 146
Predictions for groups defined by levels of categorical variables 150
4.4.3 (Advanced) Nondefault predictions using margins 153
The predict() option 153
The expression() option 154
4.4.4 Tables of predictions using mtable 155
mtable with categorical and count outcomes 158
(Advanced) Combining and formatting tables using mtable . 160
4.5 Marginal effects: Changes in predictions 162
4.5.1 Marginal effects using margins 163
4.5.2 Marginal effects using mtable 164
4.5.3 Posting predictions and using mlincom 165
4.5.4 Marginal effects using mchange 166
4.6 Plotting predictions 171
4.6.1 Plotting predictions with marginsplot 171
4.6.2 Plotting predictions using mgen 173
4.7 Interpretation of parameters 178
4.7.1 The listcoef command 179
4.7.2 Standardized coefficients 180
4.7.3 Factor and percentage change coefficients 184
4.8 Next steps 184

II Models for specific kinds of outcomes 185

5 Models for binary outcomes: Estimation, testing, and fit 187

5.1 The statistical model 187
5.1.1 A latent-variable model 188

5.1.2 A nonlinear probability model . . . 192
5.2 Estimation using logit and probit commands . . . 192
5.2.1 Example of logit model . . . 194
5.2.2 Comparing logit and probit . . . 196
5.2.3 (Advanced) Observations predicted perfectly . . . 197
5.3 Hypothesis testing . . . 200
5.3.1 Testing individual coefficients . . . 200
5.3.2 Testing multiple coefficients . . . 203
5.3.3 Comparing LR and Wald tests . . . 205
5.4 Predicted probabilities, residuals, and influential observations . . . 206
5.4.1 Predicted probabilities using predict . . . 206
5.4.2 Residuals and influential observations using predict . . . 209
5.4.3 Least likely observations . . . 216
5.5 Measures of fit . . . 218
5.5.1 Information criteria . . . 219
5.5.2 Pseudo-R^2's . . . 221
5.5.3 (Advanced) Hosmer–Lemeshow statistic . . . 223
5.6 Other commands for binary outcomes . . . 225
5.7 Conclusion . . . 225
6 Models for binary outcomes: Interpretation 227
6.1 Interpretation using regression coefficients . . . 228
6.1.1 Interpretation using odds ratios . . . 228
6.1.2 (Advanced) Interpretation using y* . . . 235
6.2 Marginal effects: Changes in probabilities . . . 239
6.2.1 Linked variables . . . 241
6.2.2 Summary measures of change . . . 242
MEMs and MERs . . . 243
AMEs . . . 243
Standard errors of marginal effects . . . 244
6.2.3 Should you use the AME, the MEM, or the MER? . . . 244

6.2.4 Examples of marginal effects . . . 246
AMEs for continuous variables . . . 248
AMEs for factor variables . . . 251
Summary table of AMEs . . . 252
Marginal effects for subgroups . . . 254
MEMs and MERs . . . 255
Marginal effects with powers and interactions . . . 259
6.2.5 The distribution of marginal effects . . . 261
6.2.6 (Advanced) Algorithm for computing the distribution of effects . . . 265
6.3 Ideal types . . . 270
6.3.1 Using local means with ideal types . . . 273
6.3.2 Comparing ideal types with statistical tests . . . 274
6.3.3 (Advanced) Using macros to test differences between ideal types . . . 275
6.3.4 Marginal effects for ideal types . . . 278
6.4 Tables of predicted probabilities . . . 280
6.5 Second differences comparing marginal effects . . . 285
6.6 Graphing predicted probabilities . . . 286
6.6.1 Using marginsplot . . . 287
6.6.2 Using mgen with the graph command . . . 290
6.6.3 Graphing multiple predictions . . . 293
6.6.4 Overlapping confidence intervals . . . 297
6.6.5 Adding power terms and plotting predictions . . . 301
6.6.6 (Advanced) Graphs with local means . . . 303
6.7 Conclusion . . . 308
7 Models for ordinal outcomes 309
7.1 The statistical model . . . 310
7.1.1 A latent-variable model . . . 310
7.1.2 A nonlinear probability model . . . 314
7.2 Estimation using ologit and oprobit . . . 314

7.2.1 Example of ordinal logit model 315
7.2.2 Predicting perfectly 319
7.3 Hypothesis testing 320
7.3.1 Testing individual coefficients 321
7.3.2 Testing multiple coefficients 322
7.4 Measures of fit using fitstat 324
7.5 (Advanced) Converting to a different parameterization 325
7.6 The parallel regression assumption 326
7.6.1 Testing the parallel regression assumption using oparallel . . 329
7.6.2 Testing the parallel regression assumption using brant . . . 330
7.6.3 Caveat regarding the parallel regression assumption 331
7.7 Overview of interpretation 331
7.8 Interpreting transformed coefficients 332
7.8.1 Marginal change in y* 332
7.8.2 Odds ratios 335
7.9 Interpretations based on predicted probabilities 338
7.10 Predicted probabilities with predict 339
7.11 Marginal effects 341
7.11.1 Plotting marginal effects 344
7.11.2 Marginal effects for a quick overview 350
7.12 Predicted probabilities for ideal types 351
7.12.1 (Advanced) Testing differences between ideal types 354
7.13 Tables of predicted probabilities 355
7.14 Plotting predicted probabilities 359
7.15 Probability plots and marginal effects 364
7.16 Less common models for ordinal outcomes 370
7.16.1 The stereotype logistic model 370
7.16.2 The generalized ordered logit model 371
7.16.3 (Advanced) Predictions without using factor-variable notation 374

7.16.4 The sequential logit model 378
7.17 Conclusion 382
8 Models for nominal outcomes 385
8.1 The multinomial logit model 386
8.1.1 Formal statement of the model 390
8.2 Estimation using the mlogit command 390
Weights and complex samples 391
Options 391
8.2.1 Example of MNLM 392
8.2.2 Selecting different base outcomes 395
8.2.3 Predicting perfectly 397
8.3 Hypothesis testing 398
8.3.1 mlogtest for tests of the MNLM 398
8.3.2 Testing the effects of the independent variables 399
8.3.3 Tests for combining alternatives 403
8.4 Independence of irrelevant alternatives 407
8.4.1 Hausman–McFadden test of IIA 408
8.4.2 Small–Hsiao test of IIA 409
8.5 Measures of fit 411
8.6 Overview of interpretation 411
8.7 Predicted probabilities with predict 412
8.8 Marginal effects 415
8.8.1 (Advanced) The distribution of marginal effects 420
8.9 Tables of predicted probabilities 423
8.9.1 (Advanced) Testing second differences 425
8.9.2 (Advanced) Predictions using local means and subsamples . 428
8.10 Graphing predicted probabilities 432
8.11 Odds ratios 435
8.11.1 Listing odds ratios with listcoef 435
8.11.2 Plotting odds ratios 436

8.12 (Advanced) Additional models for nominal outcomes 444
8.12.1 Stereotype logistic regression 445
8.12.2 Conditional logit model 454
8.12.3 Multinomial probit model with IIA 465
8.12.4 Alternative-specific multinomial probit 469
8.12.5 Rank-ordered logit model 475
8.13 Conclusion 479
9 Models for count outcomes **481**
9.1 The Poisson distribution 481
9.1.1 Fitting the Poisson distribution with the poisson command 483
9.1.2 Comparing observed and predicted counts with mgen 484
9.2 The Poisson regression model 487
9.2.1 Estimation using poisson 488
Example of the PRM 489
9.2.2 Factor and percentage changes in E(y | x) 490
Example of factor and percentage change 492
9.2.3 Marginal effects on E(y | x) 493
Examples of marginal effects 495
9.2.4 Interpretation using predicted probabilities 496
Predicted probabilities using mtable and mchange 496
Treating a count independent variable as a factor variable . 498
Predicted probabilities using mgen 500
9.2.5 Comparing observed and predicted counts to evaluate model specification 501
9.2.6 (Advanced) Exposure time 504
9.3 The negative binomial regression model 507
9.3.1 Estimation using nbreg 509
NB1 and NB2 variance functions 509
9.3.2 Example of NBRM 510
9.3.3 Testing for overdispersion 511

9.3.4 Comparing the PRM and NBRM using estimates table . . . 511
9.3.5 Robust standard errors . . . 512
9.3.6 Interpretation using E(y | x) . . . 514
9.3.7 Interpretation using predicted probabilities . . . 516
9.4 Models for truncated counts . . . 518
9.4.1 Estimation using tpoisson and tnbreg . . . 521
Example of zero-truncated model . . . 521
9.4.2 Interpretation using E(y | x) . . . 523
9.4.3 Predictions in the estimation sample . . . 524
9.4.4 Interpretation using predicted rates and probabilities . . . 525
9.5 (Advanced) The hurdle regression model . . . 527
9.5.1 Fitting the hurdle model . . . 528
9.5.2 Predictions in the sample . . . 531
9.5.3 Predictions at user-specified values . . . 533
9.5.4 Warning regarding sample specification . . . 534
9.6 Zero-inflated count models . . . 535
9.6.1 Estimation using zinb and zip . . . 538
9.6.2 Example of zero-inflated models . . . 539
9.6.3 Interpretation of coefficients . . . 540
9.6.4 Interpretation of predicted probabilities . . . 541
Predicted probabilities with mtable . . . 542
Plotting predicted probabilities with mgen . . . 543
9.7 Comparisons among count models . . . 544
9.7.1 Comparing mean probabilities . . . 545
9.7.2 Tests to compare count models . . . 547
9.7.3 Using countfit to compare count models . . . 551
9.8 Conclusion . . . 558
References **561**
Author index **569**
Subject index **573**