

Table of contents

[View table of contents >>](#)

List of tables

List of figures

[Preface to the Revised Edition \(PDF\)](#)

[Preface to the First Edition \(PDF\)](#)

1 Stata basics

1.1 Interactive use

1.2 Documentation

1.2.1 Stata manuals

1.2.2 Additional Stata resources

1.2.3 The help command

1.2.4 The search, findit, and hsearch commands

1.3 Command syntax and operators

1.3.1 Basic command syntax

1.3.2 Example: The summarize command

1.3.3 Example: The regress command

1.3.4 Abbreviations, case sensitivity, and wildcards

1.3.5 Arithmetic, relational, and logical operators

1.3.6 Error messages

1.4 Do-files and log files

1.4.1 Writing a do-file

1.4.2 Running do-files

1.4.3 Log files

1.4.4 A three-step process

1.4.5 Comments and long lines

1.4.6 Different implementations of Stata

1.5 Scalars and matrices

1.5.1 Scalars

1.5.2 Matrices

1.6 Using results from Stata commands

1.6.1 Using results from the r-class command summarize

1.6.2 Using results from the e-class command regress

1.7 Global and local macros

1.7.1 Global macros

1.7.2 Local macros

1.7.3 Scalar or macro?

1.8 Looping commands

1.8.1 The foreach loop

1.8.2 The forvalues loop

1.8.3 The while loop

1.8.4 The continue command

- 1.9 Some useful commands
- 1.10 Template do-file
- 1.11 User-written commands
- 1.12 Stata resources
- 1.13 Exercises

2 Data management and graphics

- 2.1 Introduction
- 2.2 Types of data
 - 2.2.1 Text or ASCII data
 - 2.2.2 Internal numeric data
 - 2.2.3 String data
 - 2.2.4 Formats for displaying numeric data
- 2.3 Inputting data
 - 2.3.1 General principles
 - 2.3.2 Inputting data already in Stata format
 - 2.3.3 Inputting data from the keyboard
 - 2.3.4 Inputting nontext data
 - 2.3.5 Inputting text data from a spreadsheet
 - 2.3.6 Inputting text data in free format
 - 2.3.7 Inputting text data in fixed format
 - 2.3.8 Dictionary files
 - 2.3.9 Common pitfalls
- 2.4 Data management
 - 2.4.1 PSID example
 - 2.4.2 Naming and labeling variables
 - 2.4.3 Viewing data
 - 2.4.4 Using original documentation
 - 2.4.5 Missing values
 - 2.4.6 Imputing missing data
 - 2.4.7 Transforming data (generate, replace, egen, recode)
 - The generate and replace commands
 - The egen command
 - The recode command
 - The by prefix
 - Indicator variables
 - Set of indicator variables
 - Interactions
 - Demeaning
 - 2.4.8 Saving data
 - 2.4.9 Selecting the sample
- 2.5 Manipulating datasets
 - 2.5.1 Ordering observations and variables
 - 2.5.2 Preserving and restoring a dataset
 - 2.5.3 Wide and long forms for a dataset

- 2.5.4 Merging datasets
- 2.5.5 Appending datasets
- 2.6 Graphical display of data
 - 2.6.1 Stata graph commands
 - Example graph commands
 - Saving and exporting graphs
 - Learning how to use graph commands
 - 2.6.2 Box-and-whisker plot
 - 2.6.3 Histogram
 - 2.6.4 Kernel density plot
 - 2.6.5 Twoway scatterplots and fitted lines
 - 2.6.6 Lowess, kernel, local linear, and nearest-neighbor regression
 - 2.6.7 Multiple scatterplots
- 2.7 Stata resources
- 2.8 Exercises

3 Linear regression basics

- 3.1 Introduction
- 3.2 Data and data summary
 - 3.2.1 Data description
 - 3.2.2 Variable description
 - 3.2.3 Summary statistics
 - 3.2.4 More-detailed summary statistics
 - 3.2.5 Tables for data
 - 3.2.6 Statistical tests
 - 3.2.7 Data plots
- 3.3 Regression in levels and logs
 - 3.3.1 Basic regression theory
 - 3.3.2 OLS regression and matrix algebra
 - 3.3.3 Properties of the OLS estimator
 - 3.3.4 Heteroskedasticity-robust standard errors
 - 3.3.5 Cluster-robust standard errors
 - 3.3.6 Regression in logs
- 3.4 Basic regression analysis
 - 3.4.1 Correlations
 - 3.4.2 The regress command
 - 3.4.3 Hypothesis tests
 - 3.4.4 Tables of output from several regressions
 - 3.4.5 Even better tables of regression output
 - 3.4.6 Factor variables for categorical variables and interactions
- 3.5 Specification analysis
 - 3.5.1 Specification tests and model diagnostics
 - 3.5.2 Residual diagnostic plots
 - 3.5.3 Influential observations
 - 3.5.4 Specification tests

- Test of omitted variables
- Test of the Box–Cox model
- Test of the functional form of the conditional mean
- Heteroskedasticity test
- Omnibus test
- 3.5.5 Tests have power in more than one direction
- 3.6 Prediction
 - 3.6.1 In-sample prediction
 - 3.6.2 MEs and elasticities
 - 3.6.3 Prediction in logs: The retransformation problem
 - 3.6.4 Prediction exercise
- 3.7 Sampling weights
 - 3.7.1 Weights
 - 3.7.2 Weighted mean
 - 3.7.3 Weighted regression
 - 3.7.4 Weighted prediction and MEs
- 3.8 OLS using Mata
- 3.9 Stata resources
- 3.10 Exercises

4 Simulation

- 4.1 Introduction
- 4.2 Pseudorandom-number generators: Introduction
 - 4.2.1 Uniform random-number generation
 - 4.2.2 Draws from normal
 - 4.2.3 Draws from t, chi-squared, F, gamma, and beta
 - 4.2.4 Draws from binomial, Poisson, and negative binomial
- Independent (but not identically distributed) draws from binomial
- Independent (but not identically distributed) draws from Poisson
- Histograms and density plots
- 4.3 Distribution of the sample mean
 - 4.3.1 Stata program
 - 4.3.2 The simulate command
 - 4.3.3 Central limit theorem simulation
 - 4.3.4 The postfile command
 - 4.3.5 Alternative central limit theorem simulation
- 4.4 Pseudorandom-number generators: Further details
 - 4.4.1 Inverse-probability transformation
 - 4.4.2 Direct transformation
 - 4.4.3 Other methods
 - 4.4.4 Draws from truncated normal
 - 4.4.5 Draws from multivariate normal
- Direct draws from multivariate normal
- Transformation using Cholesky decomposition
- 4.4.6 Draws using Markov chain Monte Carlo method
- 4.5 Computing integrals

- 4.5.1 Quadrature
- 4.5.2 Monte Carlo integration
- 4.5.3 Monte Carlo integration using different S
- 4.6 Simulation for regression: Introduction
 - 4.6.1 Simulation example: OLS with X^2 errors
 - 4.6.2 Interpreting simulation output
 - Unbiasedness of estimator
 - Standard errors
 - t statistic
 - Test size
 - Number of simulations
 - 4.6.3 Variations
 - Different sample size and number of simulations
 - Test power
 - Different error distributions
 - 4.6.4 Estimator inconsistency
 - 4.6.5 Simulation with endogenous regressors
- 4.7 Stata resources
- 4.8 Exercises

5 GLS regression

- 5.1 Introduction
- 5.2 GLS and FGLS regression
 - 5.2.1 GLS for heteroskedastic errors
 - 5.2.2 GLS and FGLS
 - 5.2.3 Weighted least squares and robust standard errors
 - 5.2.4 Leading examples
- 5.3 Modeling heteroskedastic data
 - 5.3.1 Simulated dataset
 - 5.3.2 OLS estimation
 - 5.3.3 Detecting heteroskedasticity
 - 5.3.4 FGLS estimation
 - 5.3.5 WLS estimation
- 5.4 System of linear regressions
 - 5.4.1 SUR model
 - 5.4.2 The sureg command
 - 5.4.3 Application to two categories of expenditures
 - 5.4.4 Robust standard errors
 - 5.4.5 Testing cross-equation constraints
 - 5.4.6 Imposing cross-equation constraints
- 5.5 Survey data: Weighting, clustering, and stratification
 - 5.5.1 Survey design
 - 5.5.2 Survey mean estimation
 - 5.5.3 Survey linear regression
- 5.6 Stata resources
- 5.7 Exercises

6 Linear instrumental-variables regression

6.1 Introduction

6.2 IV estimation

6.2.1 Basic IV theory

6.2.2 Model setup

6.2.3 IV estimators: IV, 2SLS, and GMM

6.2.4 Instrument validity and relevance

6.2.5 Robust standard-error estimates

6.3 IV example

6.3.1 The ivregress command

6.3.2 Medical expenditures with one endogenous regressor

6.3.3 Available instruments

6.3.4 IV estimation of an exactly identified model

6.3.5 IV estimation of an overidentified model

6.3.6 Testing for regressor endogeneity

6.3.7 Tests of overidentifying restrictions

6.3.8 IV estimation with a binary endogenous regressor

6.4 Weak instruments

6.4.1 Finite-sample properties of IV estimators

6.4.2 Weak instruments

Diagnostics for weak instruments

Formal tests for weak instruments

6.4.3 The estat firststage command

6.4.4 Just-identified model

6.4.5 Overidentified model

6.4.6 More than one endogenous regressor

6.4.7 Sensitivity to choice of instruments

6.5 Better inference with weak instruments

6.5.1 Conditional tests and confidence intervals

6.5.2 LIML estimator

6.5.3 Jackknife IV estimator

6.5.4 Comparison of 2SLS, LIML, JIVE, and GMM

6.6 3SLS systems estimation

6.7 Stata resources

6.8 Exercises

7 Quantile regression

7.1 Introduction

7.2 QR

7.2.1 Conditional quantiles

7.2.2 Computation of QR estimates and standard errors

7.2.3 The qreg, bsqreg, and sqreg commands

7.3 QR for medical expenditures data

7.3.1 Data summary

7.3.2 QR estimates

7.3.3 Interpretation of conditional quantile coefficients

- 7.3.4 Retransformation
- 7.3.5 Comparison of estimates at different quantiles
- 7.3.6 Heteroskedasticity test
- 7.3.7 Hypothesis tests
- 7.3.8 Graphical display of coefficients over quantiles
- 7.4 QR for generated heteroskedastic data
 - 7.4.1 Simulated dataset
 - 7.4.2 QR estimates
- 7.5 QR for count data
 - 7.5.1 Quantile count regression
 - 7.5.2 The qcount command
 - 7.5.3 Summary of doctor visits data
 - 7.5.4 Results from QCR
- 7.6 Stata resources
- 7.7 Exercises

8 Linear panel-data models: Basics

- 8.1 Introduction
- 8.2 Panel-data methods overview
 - 8.2.1 Some basic considerations
 - 8.2.2 Some basic panel models
 - Individual-effects model
 - Fixed-effects model
 - Random-effects model
 - Pooled model or population-averaged model
 - Two-way-effects model
 - Mixed linear models
 - 8.2.3 Cluster-robust inference
 - 8.2.4 The xtreg command
 - 8.2.5 Stata linear panel-data commands
- 8.3 Panel-data summary
 - 8.3.1 Data description and summary statistics
 - 8.3.2 Panel-data organization
 - 8.3.3 Panel-data description
 - 8.3.4 Within and between variation
 - 8.3.5 Time-series plots for each individual
 - 8.3.6 Overall scatterplot
 - 8.3.7 Within scatterplot
 - 8.3.8 Pooled OLS regression with cluster-robust standard errors
 - 8.3.9 Time-series autocorrelations for panel data
 - 8.3.10 Error correlation in the RE model
- 8.4 Pooled or population-averaged estimators
 - 8.4.1 Pooled OLS estimator
 - 8.4.2 Pooled FGLS estimator or population-averaged estimator
 - 8.4.3 The xtreg, pa command
 - 8.4.4 Application of the xtreg, pa command

- 8.5 Within estimator
 - 8.5.1 Within estimator
 - 8.5.2 The xtreg, fe command
 - 8.5.3 Application of the xtreg, fe command
 - 8.5.4 Least-squares dummy-variables regression
- 8.6 Between estimator
 - 8.6.1 Between estimator
 - 8.6.2 Application of the xtreg, be command
- 8.7 RE estimator
 - 8.7.1 RE estimator
 - 8.7.2 The xtreg, re command
 - 8.7.3 Application of the xtreg, re command
- 8.8 Comparison of estimators
 - 8.8.1 Estimates of variance components
 - 8.8.2 Within and between R-squared
 - 8.8.3 Estimator comparison
 - 8.8.4 Fixed effects versus random effects
 - 8.8.5 Hausman test for fixed effects
 - The hausman command
 - Robust Hausman test
 - 8.8.6 Prediction
- 8.9 First-difference estimator
 - 8.9.1 First-difference estimator
 - 8.9.2 Strict and weak exogeneity
- 8.10 Long panels
 - 8.10.1 Long-panel dataset
 - 8.10.2 Pooled OLS and PFGLS
 - 8.10.3 The xtpcse and xtgls commands
 - 8.10.4 Application of the xtgls, xtpcse, and xtscs commands
 - 8.10.5 Separate regressions
 - 8.10.6 FE and RE models
 - 8.10.7 Unit roots and cointegration
- 8.11 Panel-data management
 - 8.11.1 Wide-form data
 - 8.11.2 Convert wide form to long form
 - 8.11.3 Convert long form to wide form
 - 8.11.4 An alternative to wide-form data
- 8.12 Stata resources
- 8.13 Exercises
- 9 Linear panel-data models: Extensions**
 - 9.1 Introduction
 - 9.2 Panel IV estimation
 - 9.2.1 Panel IV
 - 9.2.2 The xtivreg command

- 9.2.3 Application of the xtivreg command
- 9.2.4 Panel IV extensions
- 9.3 Hausman–Taylor estimator
 - 9.3.1 Hausman–Taylor estimator
 - 9.3.2 The xthtaylor command
 - 9.3.3 Application of the xthtaylor command
- 9.4 Arellano–Bond estimator
 - 9.4.1 Dynamic model
 - 9.4.2 IV estimation in the FD model
 - 9.4.3 The xtabond command
 - 9.4.4 Arellano–Bond estimator: Pure time series
 - 9.4.5 Arellano–Bond estimator: Additional regressors
 - 9.4.6 Specification tests
 - 9.4.7 The xtdpdsys command
 - 9.4.8 The xtdpd command
- 9.5 Mixed linear models
 - 9.5.1 Mixed linear model
 - 9.5.2 The xtmixed command
 - 9.5.3 Random-intercept model
 - 9.5.4 Cluster–robust standard errors
 - 9.5.5 Random-slopes model
 - 9.5.6 Random-coefficients model
 - 9.5.7 Two-way random-effects model
- 9.6 Clustered data
 - 9.6.1 Clustered dataset
 - 9.6.2 Clustered data using nonpanel commands
 - 9.6.3 Clustered data using panel commands
 - 9.6.4 Hierarchical linear models
- 9.7 Stata resources
- 9.8 Exercises
- 10 Nonlinear regression methods**
 - 10.1 Introduction
 - 10.2 Nonlinear example: Doctor visits
 - 10.2.1 Data description
 - 10.2.2 Poisson model description
 - 10.3 Nonlinear regression methods
 - 10.3.1 MLE
 - 10.3.2 The poisson command
 - 10.3.3 Postestimation commands
 - 10.3.4 NLS
 - 10.3.5 The nl command
 - 10.3.6 GLM
 - 10.3.7 The glm command
 - 10.3.8 The gmm command
 - 10.3.9 Other estimators

- 10.4 Different estimates of the VCE
 - 10.4.1 General framework
 - 10.4.2 The vce() option
 - 10.4.3 Application of the vce() option
 - 10.4.4 Default estimate of the VCE
 - 10.4.5 Robust estimate of the VCE
 - 10.4.6 Cluster-robust estimate of the VCE
 - 10.4.7 Heteroskedasticity- and autocorrelation-consistent estimate of the VCE
 - 10.4.8 Bootstrap standard errors
 - 10.4.9 Statistical inference
- 10.5 Prediction
 - 10.5.1 The predict and predictnl commands
 - 10.5.2 Application of predict and predictnl
 - 10.5.3 Out-of-sample prediction
 - 10.5.4 Prediction at a specified value of one of the regressors
 - 10.5.5 Prediction at a specified value of all the regressors
 - 10.5.6 Prediction of other quantities
 - 10.5.7 The margins command for prediction
- 10.6 Marginal effects
 - 10.6.1 Calculus and finite-difference methods
 - 10.6.2 MEs estimates AME, MEM, and MER
 - 10.6.3 Elasticities and semielasticities
 - 10.6.4 Simple interpretations of coefficients in single-index models
 - 10.6.5 The margins command for marginal effects
 - 10.6.6 MEM: Marginal effect at mean
 - Comparison of calculus and finite-difference methods
 - 10.6.7 MER: Marginal effect at representative value
 - 10.6.8 AME: Average marginal effect
 - 10.6.9 Elasticities and semielasticities
 - 10.6.10 AME computed manually
 - 10.6.11 Polynomial regressors
 - 10.6.12 Interacted regressors
 - 10.6.13 Complex interactions and nonlinearities
- 10.7 Model diagnostics
 - 10.7.1 Goodness-of-fit measures
 - 10.7.2 Information criteria for model comparison
 - 10.7.3 Residuals
 - 10.7.4 Model-specification tests
- 10.8 Stata resources
- 10.9 Exercises
- 11 Nonlinear optimization methods**
 - 11.1 Introduction
 - 11.2 Newton-Raphson method

- 11.2.1 NR method
- 11.2.2 NR method for Poisson
- 11.2.3 Poisson NR example using Mata
- Core Mata code for Poisson NR iterations
- Complete Stata and Mata code for Poisson NR iterations
- 11.3 Gradient methods
 - 11.3.1 Maximization options
 - 11.3.2 Gradient methods
 - 11.3.3 Messages during iterations
 - 11.3.4 Stopping criteria
 - 11.3.5 Multiple maximums
 - 11.3.6 Numerical derivatives
- 11.4 The ml command: lf method
 - 11.4.1 The ml command
 - 11.4.2 The lf method
 - 11.4.3 Poisson example: Single-index model
 - 11.4.4 Negative binomial example: Two-index model
 - 11.4.5 NLS example: Nonlikelihood model
- 11.5 Checking the program
 - 11.5.1 Program debugging using ml check and ml trace
 - 11.5.2 Getting the program to run
 - 11.5.3 Checking the data
 - 11.5.4 Multicollinearity and near collinearity
 - 11.5.5 Multiple optimums
 - 11.5.6 Checking parameter estimation
 - 11.5.7 Checking standard-error estimation
- 11.6 The ml command: d0, d1, d2, lf0, lf1, and lf2 methods
 - 11.6.1 Evaluator functions
 - 11.6.2 The d0 method
 - 11.6.3 The d1 method
 - 11.6.4 The lf1 method with the robust estimate of the VCE
 - 11.6.5 The d2 and lf2 methods
- 11.7 The Mata optimize() function
 - 11.7.1 Type d and gf evaluators
 - 11.7.2 Optimize functions
 - 11.7.3 Poisson example
- Evaluator program for Poisson MLE
- The optimize() function for Poisson MLE
- 11.8 Generalized method of moments
 - 11.8.1 Definition
 - 11.8.2 Nonlinear IV example
 - 11.8.3 GMM using the Mata optimize() function
- 11.9 Stata resources
- 11.10 Exercises
- 12 Testing methods**

- 12.1 Introduction
- 12.2 Critical values and p-values
 - 12.2.1 Standard normal compared with Student's t
 - 12.2.2 Chi-squared compared with F
 - 12.2.3 Plotting densities
 - 12.2.4 Computing p-values and critical values
 - 12.2.5 Which distributions does Stata use?
- 12.3 Wald tests and confidence intervals
 - 12.3.1 Wald test of linear hypotheses
 - 12.3.2 The test command
 - Test single coefficient
 - Test several hypotheses
 - Test of overall significance
 - Test calculated from retrieved coefficients and VCE
 - 12.3.3 One-sided Wald tests
 - 12.3.4 Wald test of nonlinear hypotheses (delta method)
 - 12.3.5 The testnl command
 - 12.3.6 Wald confidence intervals
 - 12.3.7 The lincom command
 - 12.3.8 The nlcom command (delta method)
 - 12.3.9 Asymmetric confidence intervals
- 12.4 Likelihood-ratio tests
 - 12.4.1 Likelihood-ratio tests
 - 12.4.2 The lrtest command
 - 12.4.3 Direct computation of LR tests
- 12.5 Lagrange multiplier test (or score test)
 - 12.5.1 LM tests
 - 12.5.2 The estat command
 - 12.5.3 LM test by auxiliary regression
- 12.6 Test size and power
 - 12.6.1 Simulation DGP: OLS with chi-squared errors
 - 12.6.2 Test size
 - 12.6.3 Test power
 - 12.6.4 Asymptotic test power
- 12.7 Specification tests
 - 12.7.1 Moment-based tests
 - 12.7.2 Information matrix test
 - 12.7.3 Chi-squared goodness-of-fit test
 - 12.7.4 Overidentifying restrictions test
 - 12.7.5 Hausman test
 - 12.7.6 Other tests
- 12.8 Stata resources
- 12.9 Exercises
- 13 Bootstrap methods**

- 13.1 Introduction
- 13.2 Bootstrap methods
 - 13.2.1 Bootstrap estimate of standard error
 - 13.2.2 Bootstrap methods
 - 13.2.3 Asymptotic refinement
 - 13.2.4 Use the bootstrap with caution
- 13.3 Bootstrap pairs using the `vce(bootstrap)` option
 - 13.3.1 Bootstrap-pairs method to estimate VCE
 - 13.3.2 The `vce(bootstrap)` option
 - 13.3.3 Bootstrap standard-errors example
 - 13.3.4 How many bootstraps?
 - 13.3.5 Clustered bootstraps
 - 13.3.6 Bootstrap confidence intervals
 - 13.3.7 The `postestimation estat bootstrap` command
 - 13.3.8 Bootstrap confidence-intervals example
 - 13.3.9 Bootstrap estimate of bias
- 13.4 Bootstrap pairs using the `bootstrap` command
 - 13.4.1 The `bootstrap` command
 - 13.4.2 Bootstrap parameter estimate from a Stata estimation command
 - 13.4.3 Bootstrap standard error from a Stata estimation command
 - 13.4.4 Bootstrap standard error from a user-written estimation command
 - 13.4.5 Bootstrap two-step estimator
 - 13.4.6 Bootstrap Hausman test
 - 13.4.7 Bootstrap standard error of the coefficient of variation
- 13.5 Bootstraps with asymptotic refinement
 - 13.5.1 Percentile-t method
 - 13.5.2 Percentile-t Wald test
 - 13.5.3 Percentile-t Wald confidence interval
- 13.6 Bootstrap pairs using `bsample` and `simulate`
 - 13.6.1 The `bsample` command
 - 13.6.2 The `bsample` command with `simulate`
 - 13.6.3 Bootstrap Monte Carlo exercise
- 13.7 Alternative resampling schemes
 - 13.7.1 Bootstrap pairs
 - 13.7.2 Parametric bootstrap
 - 13.7.3 Residual bootstrap
 - 13.7.4 Wild bootstrap
 - 13.7.5 Subsampling
- 13.8 The jackknife
 - 13.8.1 Jackknife method
 - 13.8.2 The `vce(jackknife)` option and the `jackknife` command
- 13.9 Stata resources
- 13.10 Exercises
- 14 Binary outcome models**

- 14.1 Introduction
- 14.2 Some parametric models
 - 14.2.1 Basic model
 - 14.2.2 Logit, probit, linear probability, and clog-log models
- 14.3 Estimation
 - 14.3.1 Latent-variable interpretation and identification
 - 14.3.2 ML estimation
 - 14.3.3 The logit and probit commands
 - 14.3.4 Robust estimate of the VCE
 - 14.3.5 OLS estimation of LPM
- 14.4 Example
 - 14.4.1 Data description
 - 14.4.2 Logit regression
 - 14.4.3 Comparison of binary models and parameter estimates
- 14.5 Hypothesis and specification tests
 - 14.5.1 Wald tests
 - 14.5.2 Likelihood-ratio tests
 - 14.5.3 Additional model-specification tests
 - Lagrange multiplier test of generalized logit
 - Heteroskedastic probit regression
 - 14.5.4 Model comparison
- 14.6 Goodness of fit and prediction
 - 14.6.1 Pseudo- R^2 measure
 - 14.6.2 Comparing predicted probabilities with sample frequencies
 - 14.6.3 Comparing predicted outcomes with actual outcomes
 - 14.6.4 The predict command for fitted probabilities
 - 14.6.5 The prvalue command for fitted probabilities
- 14.7 Marginal effects
 - 14.7.1 Marginal effect at a representative value (MER)
 - 14.7.2 Marginal effect at the mean (MEM)
 - 14.7.3 Average marginal effect (AME)
 - 14.7.4 The prchange command
- 14.8 Endogenous regressors
 - 14.8.1 Example
 - 14.8.2 Model assumptions
 - 14.8.3 Structural-model approach
 - The ivprobit command
 - Maximum likelihood estimates
 - Two-step sequential estimates
 - 14.8.4 IVs approach
- 14.9 Grouped data
 - 14.9.1 Estimation with aggregate data
 - 14.9.2 Grouped-data application
- 14.10 Stata resources
- 14.11 Exercises

15 Multinomial models

15.1 Introduction

15.2 Multinomial models overview

15.2.1 Probabilities and MEs

15.2.2 Maximum likelihood estimation

15.2.3 Case-specific and alternative-specific regressors

15.2.4 Additive random-utility model

15.2.5 Stata multinomial model commands

15.3 Multinomial example: Choice of fishing mode

15.3.1 Data description

15.3.2 Case-specific regressors

15.3.3 Alternative-specific regressors

15.4 Multinomial logit model

15.4.1 The mlogit command

15.4.2 Application of the mlogit command

15.4.3 Coefficient interpretation

15.4.4 Predicted probabilities

15.4.5 MEs

15.5 Conditional logit model

15.5.1 Creating long-form data from wide-form data

15.5.2 The asclogit command

15.5.3 The clogit command

15.5.4 Application of the asclogit command

15.5.5 Relationship to multinomial logit model

15.5.6 Coefficient interpretation

15.5.7 Predicted probabilities

15.5.8 MEs

15.6 Nested logit model

15.6.1 Relaxing the independence of irrelevant alternatives assumption

15.6.2 NL model

15.6.3 The nlogit command

15.6.4 Model estimates

15.6.5 Predicted probabilities

15.6.6 MEs

15.6.7 Comparison of logit models

15.7 Multinomial probit model

15.7.1 MNP

15.7.2 The mprobit command

15.7.3 Maximum simulated likelihood

15.7.4 The asmprobit command

15.7.5 Application of the asmprobit command

15.7.6 Predicted probabilities and MEs

15.8 Random-parameters logit

15.8.1 Random-parameters logit

15.8.2 The mixlogit command

- 15.8.3 Data preparation for mixlogit
- 15.8.4 Application of the mixlogit command
- 15.9 Ordered outcome models
 - 15.9.1 Data summary
 - 15.9.2 Ordered outcomes
 - 15.9.3 Application of the ologit command
 - 15.9.4 Predicted probabilities
 - 15.9.5 MEs
 - 15.9.6 Other ordered models
- 15.10 Multivariate outcomes
 - 15.10.1 Bivariate probit
 - 15.10.2 Nonlinear SUR
- 15.11 Stata resources
- 15.12 Exercises

16 Tobit and selection models

- 16.1 Introduction
 - 16.2 Tobit model
 - 16.2.1 Regression with censored data
 - 16.2.2 Tobit model setup
 - 16.2.3 Unknown censoring point
 - 16.2.4 Tobit estimation
 - 16.2.5 ML estimation in Stata
 - 16.3 Tobit model example
 - 16.3.1 Data summary
 - 16.3.2 Tobit analysis
 - 16.3.3 Prediction after tobit
 - 16.3.4 Marginal effects

Left-truncated, left-censored, and right-truncated examples

Left-censored case computed directly

Marginal impact on probabilities
 - 16.3.5 The ivtobit command
 - 16.3.6 Additional commands for censored regression
- 16.4 Tobit for lognormal data
 - 16.4.1 Data example
 - 16.4.2 Setting the censoring point for data in logs
 - 16.4.3 Results
 - 16.4.4 Two-limit tobit
 - 16.4.5 Model diagnostics
 - 16.4.6 Tests of normality and homoskedasticity

Generalized residuals and scores

Test of normality

Test of homoskedasticity
- 16.4.7 Next step?
- 16.5 Two-part model in logs

- 16.5.1 Model structure
- 16.5.2 Part 1 specification
- 16.5.3 Part 2 of the two-part model
- 16.6 Selection model
 - 16.6.1 Model structure and assumptions
 - 16.6.2 ML estimation of the sample-selection model
 - 16.6.3 Estimation without exclusion restrictions
 - 16.6.4 Two-step estimation
 - 16.6.5 Estimation with exclusion restrictions
- 16.7 Prediction from models with outcome in logs
 - 16.7.1 Predictions from tobit
 - 16.7.2 Predictions from two-part model
 - 16.7.3 Predictions from selection model
- 16.8 Stata resources
- 16.9 Exercises

17 Count-data models

- 17.1 Introduction
- 17.2 Features of count data
 - 17.2.1 Generated Poisson data
 - 17.2.2 Overdispersion and negative binomial data
 - 17.2.3 Modeling strategies
 - 17.2.4 Estimation methods
- 17.3 Empirical example 1
 - 17.3.1 Data summary
 - 17.3.2 Poisson model
 - Poisson model results
 - Robust estimate of VCE for Poisson MLE
 - Test of overdispersion
 - Coefficient interpretation and marginal effects
 - 17.3.3 NB2 model
 - NB2 model results
 - Fitted probabilities for Poisson and NB2 models
 - The countfit command
 - The prvalue command
 - Discussion
 - Generalized NB model
 - 17.3.4 Nonlinear least-squares estimation
 - 17.3.5 Hurdle model
 - Variants of the hurdle model
 - Application of the hurdle model
 - 17.3.6 Finite-mixture models
 - FMM specification
 - Simulated FMM sample with comparisons
 - ML estimation of the FMM
 - The fmm command

Application: Poisson finite-mixture model

Interpretation

Comparing marginal effects

Application: NB finite-mixture model

Model selection

Cautionary note

17.4 Empirical example 2

17.4.1 Zero-inflated data

17.4.2 Models for zero-inflated data

17.4.3 Results for the NB2 model

The prcounts command

17.4.4 Results for ZINB

17.4.5 Model comparison

The countfit command

Model comparison using countfit

17.5 Models with endogenous regressors

17.5.1 Structural-model approach

Model and assumptions

Two-step estimation

Application

17.5.2 Nonlinear IV method

17.6 Stata resources

17.7 Exercises

18 Nonlinear panel models

18.1 Introduction

18.2 Nonlinear panel-data overview

18.2.1 Some basic nonlinear panel models

FE models

RE models

Pooled models or population-averaged models

Comparison of models

18.2.2 Dynamic models

18.2.3 Stata nonlinear panel commands

18.3 Nonlinear panel-data example

18.3.1 Data description and summary statistics

18.3.2 Panel-data organization

18.3.3 Within and between variation

18.3.4 FE or RE model for these data?

18.4 Binary outcome models

18.4.1 Panel summary of the dependent variable

18.4.2 Pooled logit estimator

18.4.3 The xtlogit command

18.4.4 The xtgee command

18.4.5 PA logit estimator

18.4.6 RE logit estimator

- 18.4.7 FE logit estimator
- 18.4.8 Panel logit estimator comparison
- 18.4.9 Prediction and marginal effects
- 18.4.10 Mixed-effects logit estimator
- 18.5 Tobit model
 - 18.5.1 Panel summary of the dependent variable
 - 18.5.2 RE tobit model
 - 18.5.3 Generalized tobit models
 - 18.5.4 Parametric nonlinear panel models
- 18.6 Count-data models
 - 18.6.1 The xtpoisson command
 - 18.6.2 Panel summary of the dependent variable
 - 18.6.3 Pooled Poisson estimator
 - 18.6.4 PA Poisson estimator
 - 18.6.5 RE Poisson estimators
 - 18.6.6 FE Poisson estimator
 - 18.6.7 Panel Poisson estimators comparison
 - 18.6.8 Negative binomial estimators
- 18.7 Stata resources
- 18.8 Exercises

A Programming in Stata

- A.1 Stata matrix commands
 - A.1.1 Stata matrix overview
 - A.1.2 Stata matrix input and output
 - Matrix input by hand
 - Matrix input from Stata estimation results
 - A.1.3 Stata matrix subscripts and combining matrices
 - A.1.4 Matrix operators
 - A.1.5 Matrix functions
 - A.1.6 Matrix accumulation commands
 - A.1.7 OLS using Stata matrix commands
- A.2 Programs
 - A.2.1 Simple programs (no arguments or access to results)
 - A.2.2 Modifying a program
 - A.2.3 Programs with positional arguments
 - A.2.4 Temporary variables
 - A.2.5 Programs with named positional arguments
 - A.2.6 Storing and retrieving program results
 - A.2.7 Programs with arguments using standard Stata syntax
 - A.2.8 Ado-files
- A.3 Program debugging
 - A.3.1 Some simple tips
 - A.3.2 Error messages and return code
 - A.3.3 Trace

B Mata

B.1 How to run Mata

B.1.1 Mata commands in Mata

B.1.2 Mata commands in Stata

B.1.3 Stata commands in Mata

B.1.4 Interactive versus batch use

B.1.5 Mata help

B.2 Mata matrix commands

B.2.1 Mata matrix input

Matrix input by hand

Identity matrices, unit vectors, and matrices of constants

Matrix input from Stata data

Matrix input from Stata matrix

Stata interface functions

B.2.2 Mata matrix operators

Element-by-element operators

B.2.3 Mata functions

Scalar and matrix functions

Matrix inversion

B.2.4 Mata cross products

B.2.5 Mata matrix subscripts and combining matrices

B.2.6 Transferring Mata data and matrices to Stata

Creating Stata matrices from Mata matrices

Creating Stata data from a Mata vector

B.3 Programming in Mata

B.3.1 Declarations

B.3.2 Mata program

B.3.3 Mata program with results output to Stata

B.3.4 Stata program that calls a Mata program

B.3.5 Using Mata in ado-files

Glossary of abbreviations

References