Contents

INTRODUCTION 1

Why study econometrics? 1
Aim of this text 2
Mathematics and statistics prerequisites for studying econometrics 2
Additional resources 3
Econometrics software 4

REVIEW: RANDOM VARIABLES, SAMPLING, AND ESTIMATION 5

R.1 **Introduction** 5
R.2 **Discrete random variables and expectations** 6
Discrete random variables 6
Expected values of random variables 8
Expected values of functions of discrete random variables 9
Expected value rules 10
Population variance of a discrete random variable 11
Fixed and random components of a random variable 12
R.3 **Continuous random variables** 14
Probability density 14
R.4 **Population covariance, covariance and variance rules, and correlation** 17
Covariance 17
Independence of random variables 19
Covariance rules 19
Variance rules 21
Correlation 22
R.5 **Samples, the double structure of a variable, and estimators** 23
Sampling 23
Estimators 24
R.6 **Unbiasedness and efficiency** 27
Unbiasedness 27
Efficiency 28
Conflicts between unbiasedness and minimum variance 30

R.7 **Estimators of variance, covariance, and correlation** 33
R.8 **The normal distribution** 34
R.9 **Hypothesis testing** 36
Formulation of a null hypothesis and development of its implications 37
Compatibility, freakiness, and the significance level 38
R.10 **Type II error and the power of a test** 42
R.11 ***t* tests** 47
The reject/fail-to-reject terminology 50
R.12 **Confidence intervals** 51
R.13 **One-sided tests** 56
H_0: $\mu = \mu_0$, H_1: $\mu = \mu_1$ 57
One percent significance test 60
Generalizing from H_0: $\mu = \mu_0$, H_1: $\mu = \mu_1$ to H_0: $\mu = \mu_0$, H_1: $\mu > \mu_0$ 62
H_0: $\mu = \mu_0$, H_1: $\mu < \mu_0$ 63
One-sided *t* tests 63
Special case: H_0: $\mu = 0$ 63
Anomalous results 64
Justification of the use of a one-sided test 65
R.14 **Probability limits and consistency** 66
Probability limits 67
Consistency 68
Why is consistency of interest? 70
Simulations 72
R.15 **Convergence in distribution and central limit theorems** 74
Key terms 79
Appendix R.1: **Unbiased estimators of the population covariance and variance** 79
Appendix R.2: **Density functions of transformed random variables** 81

1. SIMPLE REGRESSION ANALYSIS 83

1.1 **The simple linear model** 83
1.2 **Least squares regression with one explanatory variable** 85
1.3 **Derivation of the regression coefficients** 88
Least squares regression with one explanatory variable: the general case 90
Two decompositions of the dependent variable 92
Regression model without an intercept 94
1.4 **Interpretation of a regression equation** 95
Changes in the units of measurement 97
1.5 **Two important results relating to OLS regressions** 103
The mean value of the residuals is zero 103

The sample correlation between the observations on X and the residuals is zero 103

1.6 **Goodness of fit: R^2** 104
Example of how R^2 is calculated 107
Alternative interpretation of R^2 107

Key terms 108

2. PROPERTIES OF THE REGRESSION COEFFICIENTS AND HYPOTHESIS TESTING 110

2.1 **Types of data and regression model** 110

2.2 **Assumptions for regression models with nonstochastic regressors** 111

2.3 **The random components and unbiasedness of the OLS regression coefficients** 114
The random components of the OLS regression coefficients 114
The unbiasedness of the OLS regression coefficients 118
Normal distribution of the regression coefficients 120

2.4 **A Monte Carlo experiment** 121

2.5 **Precision of the regression coefficients** 125
Variances of the regression coefficients 125
Standard errors of the regression coefficients 129
The Gauss–Markov theorem 132

2.6 **Testing hypotheses relating to the regression coefficients** 134
0.1 percent tests 139
p values 139
One-sided tests 140
Confidence intervals 142

2.7 **The F test of goodness of fit** 145
Relationship between the F test and the t test on the slope coefficient in simple regression analysis 147

Key terms 148

Appendix 2.1 **The Gauss–Markov theorem** 149

3. MULTIPLE REGRESSION ANALYSIS 151

3.1 **Illustration: a model with two explanatory variables** 151

3.2 **Derivation and interpretation of the multiple regression coefficients** 153
The general model 155
Interpretation of the multiple regression coefficients 156

3.3 **Properties of the multiple regression coefficients** 159
Unbiasedness 160

Efficiency 161
Precision of the multiple regression coefficients 161
t tests and confidence intervals 164

3.4 **Multicollinearity** 165
Multicollinearity in models with more than two explanatory variables 168
Example of multicollinearity 169
What can you do about multicollinearity? 169

3.5 **Goodness of fit: R^2** 176
F tests 177
Further analysis of variance 180
Relationship between *F* statistic and *t* statistic 182
Adjusted' R^2 183

3.6 **Prediction** 185
Properties of least squares predictors 187

Key terms 190

4. NONLINEAR MODELS AND TRANSFORMATIONS OF VARIABLES 192

4.1 **Linearity and nonlinearity** 192

4.2 **Logarithmic transformations** 196
Logarithmic models 196
Semilogarithmic models 200
The disturbance term 202
Comparing linear and logarithmic specifications 205

4.3 **Models with quadratic and interactive variables** 209
Quadratic variables 209
Higher-order polynomials 211
Interactive explanatory variables 213
Ramsey's RESET test of functional misspecification 216

4.4 **Nonlinear regression** 218

Key terms 222

5. DUMMY VARIABLES 224

5.1 **Illustration of the use of a dummy variable** 224
Standard errors and hypothesis testing 227

5.2 **Extension to more than two categories and to multiple sets of dummy variables** 230
Joint explanatory power of a group of dummy variables 233
Change of reference category 234

The dummy variable trap 235
Multiple sets of dummy variables 237
5.3 **Slope dummy variables** 240
Joint explanatory power of the intercept and slope dummy variables 243
5.4 **The Chow test** 245
Relationship between the Chow test and the *F* test of the explanatory power of a set of dummy variables 247
Key terms 249

6. SPECIFICATION OF REGRESSION VARIABLES 250

6.1 **Model specification** 250
6.2 **The effect of omitting a variable that ought to be included** 251
The problem of bias 251
Invalidation of the statistical tests 254
R^2 in the presence of omitted variable bias 256
6.3 **The effect of including a variable that ought not to be included** 260
6.4 **Proxy variables** 263
Unintentional proxies 266
6.5 **Testing a linear restriction** 268
F test of a linear restriction 270
The reparameterization of a regression model 270
t test of a linear restriction 273
Multiple restrictions 274
Zero restrictions 275
6.6 **Getting the most out of your residuals** 278
Key terms 279

7. HETEROSCEDASTICITY 280

7.1 **Heteroscedasticity and its implications** 280
Possible causes of heteroscedasticity 283
7.2 **Detection of heteroscedasticity** 285
The Goldfeld–Quandt test 285
The White test 286
7.3 **Remedies for heteroscedasticity** 288
Nonlinear models 292
White's heteroscedasticity-consistent standard errors 294
How serious are the consequences of heteroscedasticity? 295
Key terms 297

8. STOCHASTIC REGRESSORS AND MEASUREMENT ERRORS 300

8.1 **Assumptions for models with stochastic regressors** 300

8.2 **Finite sample properties of the OLS regression estimators** 302
Precision and efficiency 303

8.3 **Asymptotic properties of the OLS regression estimators** 303
Consistency 304
Asymptotic normality of the regression coefficients 305

8.4 **The consequences of measurement errors** 306
Measurement errors in the explanatory variable(s) 306
Measurement errors in the dependent variable 309
Imperfect proxy variables 310

8.5 **Instrumental variables** 316
Asymptotic distribution of the IV estimator 318
Monte Carlo illustration 320
Multiple instruments 325
The Durbin–Wu–Hausman specification test 326

Key terms 329

9. SIMULTANEOUS EQUATIONS ESTIMATION 331

9.1 **Simultaneous equations models: structural and reduced form equations** 331

9.2 **Simultaneous equations bias** 333
A Monte Carlo experiment 335

9.3 **Instrumental variables estimation** 338
Underidentification 341
Overidentification 342
Two-stage least squares 343
The order condition for identification 344
Unobserved heterogeneity 345
Durbin–Wu–Hausman test 346

Key terms 350

10 BINARY CHOICE AND LIMITED DEPENDENT VARIABLE MODELS, AND MAXIMUM LIKELIHOOD ESTIMATION 354

10.1 **The linear probability model** 354

10.2 **Logit analysis** 359
Generalization to more than one explanatory variable 361
Goodness of fit and statistical tests 362

10.3 **Probit analysis** 365

10.4 **Censored regressions: tobit analysis** 368

10.5 **Sample selection bias** 374
10.6 **An introduction to maximum likelihood estimation** 378
Generalization to a sample of n observations 382
Generalization to the case where σ is unknown 383
Application to the simple regression model 385
Goodness of fit and statistical tests 386
Key terms 387
Appendix 10.1 **Comparing linear and logarithmic specifications** 388

11. MODELS USING TIME SERIES DATA 391
11.1 **Assumptions for regressions with time series data** 391
11.2 **Static models** 393
11.3 **Models with lagged explanatory variables** 398
Estimating long-run effects 400
11.4 **Models with a lagged dependent variable** 401
The partial adjustment model 404
The error correction model 405
The adaptive expectations model 406
More general autoregressive models 409
11.5 **Assumption C.7 and the properties of estimators in autoregressive models** 411
Consistency 414
Limiting distributions 415
t tests in an autoregressive model 417
11.6 **Simultaneous equations models** 419
11.7 **Alternative dynamic representations of time series processes** 422
Time series analysis 423
Vector autoregressions 425
Key terms 427

12. AUTOCORRELATION 429
12.1 **Definition and consequences of autocorrelation** 429
Consequences of autocorrelation 431
Autocorrelation with a lagged dependent variable 433
12.2 **Detection of autocorrelation** 434
The Breusch–Godfrey test 435
The Durbin–Watson test 436
Durbin's h test 438
12.3 **Fitting a model subject to AR(1) autocorrelation** 440
Issues 441
Inference 442
The common factor test 445

12.4 **Apparent autocorrelation** 451
12.5 **Model specification: specific-to-general versus general-to-specific** 457
Comparison of alternative models 458
The general-to-specific approach to model specification 460
Key terms 461
Appendix 12.1: **Demonstration that the Durbin–Watson statistic approximates $2 - 2\rho$ in large samples** 462

13. INTRODUCTION TO NONSTATIONARY TIME SERIES 463

13.1 **Stationarity and nonstationarity** 463
Stationary time series 463
Nonstationary time series 469
Deterministic trend 472
Difference-stationarity and trend-stationarity 473
13.2 **Spurious regressions** 475
Spurious regressions with variables possessing deterministic trends 475
Spurious regressions with variables that are random walks 476
13.3 **Graphical techniques for detecting nonstationarity** 484
13.4 **Tests of nonstationarity** 489
Untrended process 491
Trended process 497
Augmented Dickey–Fuller tests 498
Other tests 500
Further complications 502
13.5 **Cointegration** 504
13.6 **Fitting models with nonstationary time series** 508
Detrending 509
Differencing 509
Error correction models 510
Key terms 513

14. INTRODUCTION TO PANEL DATA MODELS 514

14.1 **Introduction** 514
Example of the use of a panel data set to investigate dynamics 515
14.2 **Fixed effects regressions** 518
Within-groups fixed effects 518
First differences fixed effects 519
Least squares dummy variable fixed effects 520

14.3 **Random effects regressions** 522
Assessing the appropriateness of fixed effects and random effects estimation 525
Random effects or OLS? 526
A note on the random effects and fixed effects terminology 527
Key terms 527

APPENDIX A: Statistical tables 531
APPENDIX B: Data Sets 548
Bibliography 559
Author Index 563
Subject Index 565