Contents

CHAPTER 1

The Nature of Econometrics and Economic Data 1

1.1 What Is Econometrics? 1
1.2 Steps in Empirical Economic Analysis 2
1.3 The Structure of Economic Data 5
Cross-Sectional Data 5
Time Series Data 8
Pooled Cross Sections 9
Panel or Longitudinal Data 10
A Comment on Data Structures 12
1.4 Causality and the Notion of Ceteris Paribus in Econometric Analysis 12
Summary 17
Key Terms 17
Problems 17
Computer Exercises 18

PART 1

Regression Analysis with Cross-Sectional Data 21

CHAPTER 2

The Simple Regression Model 22

2.1 Definition of the Simple Regression Model 22
2.2 Deriving the Ordinary Least Squares Estimates 27
A Note on Terminology 35
2.3 Properties of OLS on Any Sample of Data 36
Fitted Values and Residuals 36
Algebraic Properties of OLS Statistics 37
Goodness-of-Fit 40
2.4 Units of Measurement and Functional Form 41
The Effects of Changing Units of Measurement on OLS Statistics 41
Incorporating Nonlinearities in Simple Regression 43
The Meaning of "Linear" Regression 46
2.5 Expected Values and Variances of the OLS Estimators 46
Unbiasedness of OLS 47
Variances of the OLS Estimators 52
Estimating the Error Variance 56
2.6 Regression through the Origin 58
Summary 59
Key Terms 60
Problems 61
Computer Exercises 64
Appendix 2A 66

CHAPTER 3

Multiple Regression Analysis: Estimation 68

3.1 Motivation for Multiple Regression 68
The Model with Two Independent Variables 68
The Model with k Independent Variables 71
3.2 Mechanics and Interpretation of Ordinary Least Squares 73
Obtaining the OLS Estimates 73
Interpreting the OLS Regression Equation 74
On the Meaning of "Holding Other Factors Fixed" in Multiple Regression 77
Changing More Than One Independent Variable Simultaneously 77
OLS Fitted Values and Residuals 77
A "Partialling Out" Interpretation of Multiple Regression 78
Comparison of Simple and Multiple Regression Estimates 79
Goodness-of-Fit 80
Regression through the Origin 83
3.3 The Expected Value of the OLS Estimators 84
Including Irrelevant Variables in a Regression Model 89
Omitted Variable Bias: The Simple Case 89
Omitted Variable Bias: More General Cases 93
3.4 The Variance of the OLS Estimators 94
The Components of the OLS Variances: Multicollinearity 95

Variances in Misspecified Models 99
Estimating σ^2: Standard Errors of the OLS Estimators 101
3.5 Efficiency of OLS: The Gauss-Markov Theorem 102
Summary 104
Key Terms 105
Problems 105
Computer Exercises 110
Appendix 3A 113

CHAPTER 4
Multiple Regression Analysis: Inference 117

4.1 Sampling Distributions of the OLS Estimators 117
4.2 Testing Hypotheses about a Single Population Parameter: The *t* Test 120
Testing against One-Sided Alternatives 123
Two-Sided Alternatives 128
Testing Other Hypotheses about β_j 130
Computing p-Values for t Tests 133
A Reminder on the Language of Classical Hypothesis Testing 135
Economic, or Practical, versus Statistical Significance 135
4.3 Confidence Intervals 138
4.4 Testing Hypotheses about a Single Linear Combination of the Parameters 140
4.5 Testing Multiple Linear Restrictions: The *F* Test 143
Testing Exclusion Restrictions 143
Relationship between F and t Statistics 149
The R-Squared Form of the F Statistic 150
Computing p-Values for F Tests 151
The F Statistic for Overall Significance of a Regression 152
Testing General Linear Restrictions 153
4.6 Reporting Regression Results 154
Summary 156
Key Terms 158
Problems 159
Computer Exercises 163

CHAPTER 5
Multiple Regression Analysis: OLS Asymptotics 167

5.1 Consistency 167
Deriving the Inconsistency in OLS 170
5.2 Asymptotic Normality and Large Sample Inference 172
Other Large Sample Tests: The Lagrange Multiplier Statistic 176
5.3 Asymptotic Efficiency of OLS 179
Summary 180
Key Terms 181
Problems 181
Computer Exercises 181
Appendix 5A 182

CHAPTER 6
Multiple Regression Analysis: Further Issues 184

6.1 Effects of Data Scaling on OLS Statistics 184
Beta Coefficients 187
6.2 More on Functional Form 189
More on Using Logarithmic Functional Forms 189
Models with Quadratics 192
Models with Interaction Terms 197
6.3 More on Goodness-of-Fit and Selection of Regressors 199
Adjusted R-Squared 200
Using Adjusted R-Squared to Choose between Nonnested Models 201
Controlling for Too Many Factors in Regression Analysis 203
Adding Regressors to Reduce the Error Variance 205
6.4 Prediction and Residual Analysis 206
Confidence Intervals for Predictions 206
Residual Analysis 209
Predicting y When log(y) Is the Dependent Variable 210
Summary 215
Key Terms 215
Problems 216
Computer Exercises 218
Appendix 6A 223

CHAPTER 7
Multiple Regression Analysis with Qualitative Information: Binary (or Dummy) Variables 225

7.1 Describing Qualitative Information 225

7.2 A Single Dummy Independent Variable 226
Interpreting Coefficients on Dummy Explanatory Variables When the Dependent Variable Is log(y) 231
7.3 Using Dummy Variables for Multiple Categories 233
Incorporating Ordinal Informationby Using Dummy Variables 235
7.4 Interactions Involving Dummy Variables 238
Interactions among Dummy Variables 238
Allowing for Different Slopes 239
Testing for Differences in Regression Functions across Groups 243
7.5 A Binary Dependent Variable: The Linear Probability Model 246
7.6 More on Policy Analysis and Program Evaluation 251
Summary 254
Key Terms 255
Problems 255
Computer Exercises 258

CHAPTER 8

Heteroskedasticity 264

8.1 Consequences of Heteroskedasticity for OLS 264
8.2 Heteroskedasticity-Robust Inference after OLS Estimation 265
Computing Heteroskedasticity-Robust LM Tests 269
8.3 Testing for Heteroskedasticity 271
The White Test for Heteroskedasticity 274
8.4 Weighted Least Squares Estimation 276
The Heteroskedasticity Is Known up to a Multiplicative Constant 277
The Heteroskedasticity Function Must Be Estimated: Feasible GLS 282
What If the Assumed Heteroskedasticity Function Is Wrong? 287
Prediction and Prediction Intervals with Heteroskedasticity 289
8.5 The Linear Probability Model Revisited 290
Summary 293
Key Terms 294
Problems 294
Computer Exercises 296

CHAPTER 9

More on Specification and Data Issues 300

9.1 Functional Form Misspecification 300
RESET as a General Test for Functional Form Misspecification 303
Tests against Nonnested Alternatives 305
9.2 Using Proxy Variables for Unobserved Explanatory Variables 306
Using Lagged Dependent Variables as Proxy Variables 310
A Different Slant on Multiple Regression 312
9.3 Models with Random Slopes 313
9.4 Properties of OLS under Measurement Error 315
Measurement Error in the Dependent Variable 316
Measurement Error in an Explanatory Variable 318
9.5 Missing Data, Nonrandom Samples, and Outlying Observations 322
Missing Data 322
Nonrandom Samples 323
Outliers and Influential Observations 325
9.6 Least Absolute Deviations Estimation 330
Summary 331
Key Terms 332
Problems 332
Computer Exercises 334

PART 2

Regression Analysis with Time Series Data 339

CHAPTER 10

Basic Regression Analysis with Time Series Data 340

10.1 The Nature of Time Series Data 340
10.2 Examples of Time Series Regression Models 342
Static Models 342
Finite Distributed Lag Models 342
A Convention about the Time Index 345
10.3 Finite Sample Properties of OLS under Classical Assumptions 345
Unbiasedness of OLS 345
The Variances of the OLS Estimators and the Gauss-Markov Theorem 349
Inference under the Classical Linear Model Assumptions 351
10.4 Functional Form, Dummy Variables, and Index Numbers 353

10.5 Trends and Seasonality 360
Characterizing Trending Time Series 360
Using Trending Variables in Regression Analysis 363
A Detrending Interpretation of Regressions with a Time Trend 365
Computing R-Squared when the Dependent Variable Is Trending 366
Seasonality 368
Summary 370
Key Terms 371
Problems 371
Computer Exercises 373

CHAPTER 11

Further Issues in Using OLS with Time Series Data 377

11.1 Stationary and Weakly Dependent Time Series 377
Stationary and Nonstationary Time Series 378
Weakly Dependent Time Series 379
11.2 Asymptotic Properties of OLS 381
11.3 Using Highly Persistent Time Series in Regression Analysis 388
Highly Persistent Time Series 388
Transformations on Highly Persistent Time Series 393
Deciding Whether a Time Series Is I(1) 394
11.4 Dynamically Complete Models and the Absence of Serial Correlation 396
11.5 The Homoskedasticity Assumption for Time Series Models 399
Summary 400
Key Terms 401
Problems 401
Computer Exercises 404

CHAPTER 12

Serial Correlation and Heteroskedasticity in Time Series Regressions 408

12.1 Properties of OLS with Serially Correlated Errors 408
Unbiasedness and Consistency 408
Efficiency and Inference 409
Goodness-of-Fit 410
Serial Correlation in the Presence of Lagged Dependent Variables 411
12.2 Testing for Serial Correlation 412
A t Test for AR(1) Serial Correlation with Strictly Exogenous Regressors 412
The Durbin-Watson Test under Classical Assumptions 415
Testing for AR(1) Serial Correlation without Strictly Exogenous Regressors 416
Testing for Higher Order Serial Correlation 417
12.3 Correcting for Serial Correlation with Strictly Exogenous Regressors 419
Obtaining the Best Linear Unbiased Estimator in the AR(1) Model 419
Feasible GLS Estimation with AR(1) Errors 421
Comparing OLS and FGLS 423
Correcting for Higher Order Serial Correlation 425
12.4 Differencing and Serial Correlation 426
12.5 Serial Correlation-Robust Inference after OLS 428
12.6 Heteroskedasticity in Time Series Regressions 432
Heteroskedasticity-Robust Statistics 432
Testing for Heteroskedasticity 432
Autoregressive Conditional Heteroskedasticity 433
Heteroskedasticity and Serial Correlation in Regression Models 435
Summary 437
Key Terms 437
Problems 438
Computer Exercises 438

PART 3

Advanced Topics 443

CHAPTER 13

Pooling Cross Sections across Time: Simple Panel Data Methods 444

13.1 Pooling Independent Cross Sections across Time 445
The Chow Test for Structural Change across Time 449
13.2 Policy Analysis with Pooled Cross Sections 450
13.3 Two-Period Panel Data Analysis 455
Organizing Panel Data 461

13.4 Policy Analysis with Two-Period Panel Data 462
13.5 Differencing with More Than Two Time Periods 465
Potential Pitfalls in First Differencing Panel Data 470
Summary 471
Key Terms 471
Problems 471
Computer Exercises 473
Appendix 13A 478

CHAPTER 14
Advanced Panel Data Methods 481

14.1 Fixed Effects Estimation 481
The Dummy Variable Regression 485
Fixed Effects or First Differencing? 487
Fixed Effects with Unbalanced Panels 488
14.2 Random Effects Models 489
Random Effects or Fixed Effects? 493
14.3 Applying Panel Data Methods to Other Data Structures 494
Summary 496
Key Terms 496
Problems 497
Computer Exercises 498
Appendix 14A 503

CHAPTER 15
Instrumental Variables Estimation and Two Stage Least Squares 506

15.1 Motivation: Omitted Variables in a Simple Regression Model 507
Statistical Inference with the IV Estimator 510
Properties of IV with a Poor Instrumental Variable 514
Computing R-Squared after IV Estimation 516
15.2 IV Estimation of the Multiple Regression Model 517
15.3 Two Stage Least Squares 521
A Single Endogenous Explanatory Variable 521
Multicollinearity and 2SLS 523
Multiple Endogenous Explanatory Variables 524
Testing Multiple Hypotheses after 2SLS Estimation 525
15.4 IV Solutions to Errors-in-Variables Problems 525
15.5 Testing for Endogeneity and Testing Overidentifying Restrictions 527
Testing for Endogeneity 527
Testing Overidentification Restrictions 529
15.6 2SLS with Heteroskedasticity 531
15.7 Applying 2SLS to Time Series Equations 531
15.8 Applying 2SLS to Pooled Cross Sections and Panel Data 534
Summary 536
Key Terms 536
Problems 536
Computer Exercises 539
Appendix 15A 543

CHAPTER 16
Simultaneous Equations Models 546

16.1 The Nature of Simultaneous Equations Models 546
16.2 Simultaneity Bias in OLS 550
16.3 Identifying and Estimating a Structural Equation 552
Identification in a Two-Equation System 552
Estimation by 2SLS 557
16.4 Systems with More Than Two Equations 559
Identification in Systems with Three or More Equations 559
Estimation 560
16.5 Simultaneous Equations Models with Time Series 560
16.6 Simultaneous Equations Models with Panel Data 564
Summary 566
Key Terms 567
Problems 567
Computer Exercises 570

CHAPTER 17
Limited Dependent Variable Models and Sample Selection Corrections 574

17.1 Logit and Probit Models for Binary Response 575
Specifying Logit and Probit Models 575
Maximum Likelihood Estimation of Logit and Probit Models 578
Testing Multiple Hypotheses 579

Interpreting the Logit and Probit Estimates 580
17.2 The Tobit Model for Corner Solution Responses 587
Interpreting the Tobit Estimates 589
Specification Issues in Tobit Models 594
17.3 The Poisson Regression Model 595
17.4 Censored and Truncated Regression Models 600
Censored Regression Models 601
Truncated Regression Models 604
17.5 Sample Selection Corrections 606
When Is OLS on the Selected Sample Consistent? 607
Incidental Truncation 608
Summary 612
Key Terms 613
Problems 614
Computer Exercises 615
Appendix 17A 620
Appendix 17B 621

CHAPTER 18

Advanced Time Series Topics 623

18.1 Infinite Distributed Lag Models 624
The Geometric (or Koyck) Distributed Lag 626
Rational Distributed Lag Models 628
18.2 Testing for Unit Roots 630
18.3 Spurious Regression 636
18.4 Cointegration and Error Correction Models 637
Cointegration 637
Error Correction Models 643
18.5 Forecasting 645
Types of Regression Models Used for Forecasting 646
One-Step-Ahead Forecasting 647
Comparing One-Step-Ahead Forecasts 651
Multiple-Step-Ahead Forecasts 652
Forecasting Trending, Seasonal, and Integrated Processes 655
Summary 660
Key Terms 661
Problems 661
Computer Exercises 663

CHAPTER 19

Carrying Out an Empirical Project 668

19.1 Posing a Question 668
19.2 Literature Review 670
19.3 Data Collection 671
Deciding on the Appropriate Data Set 671
Entering and Storing Your Data 672
Inspecting, Cleaning, and Summarizing Your Data 673
19.4 Econometric Analysis 675
19.5 Writing an Empirical Paper 678
Introduction 678
Conceptual (or Theoretical) Framework 679
Econometric Models and Estimation Methods 679
The Data 682
Results 682
Conclusions 683
Style Hints 684
Summary 687
Key Terms 687
Sample Empirical Projects 687
List of Journals 692
Data Sources 693

APPENDIX A

Basic Mathematical Tools 695

A.1 The Summation Operator and Descriptive Statistics 695
A.2 Properties of Linear Functions 697
A.3 Proportions and Percentages 699
A.4 Some Special Functions and Their Properties 702
Quadratic Functions 702
The Natural Logarithm 704
The Exponential Function 708
A.5 Differential Calculus 709
Summary 711
Key Terms 711
Problems 711

APPENDIX B

Fundamentals of Probability 714

B.1 Random Variables and Their Probability Distributions 714
Discrete Random Variables 715
Continuous Random Variables 717
B.2 Joint Distributions, Conditional Distributions, and Independence 719
Joint Distributions and Independence 719
Conditional Distributions 721
B.3 Features of Probability Distributions 722

A Measure of Central Tendency: The Expected Value 722
Properties of Expected Values 724
Another Measure of Central Tendency: The Median 725
Measures of Variability: Variance and Standard Deviation 726
Variance 726
Standard Deviation 728
Standardizing a Random Variable 728
Skewness and Kurtosis 729
B.4 Features of Joint and Conditional Distributions 729
Measures of Association: Covariance and Correlation 729
Covariance 729
Correlation Coefficient 731
Variance of Sums of Random Variables 732
Conditional Expectation 733
Properties of Conditional Expectation 734
Conditional Variance 736
B.5 The Normal and Related Distributions 737
The Normal Distribution 737
The Standard Normal Distribution 738
Additional Properties of the Normal Distribution 740
The Chi-Square Distribution 741
The t Distribution 741
The F Distribution 743
Summary 744
Key Terms 744
Problems 745

APPENDIX C
Fundamentals of Mathematical Statistics 747

C.1 Populations, Parameters, and Random Sampling 747
Sampling 748
C.2 Finite Sample Properties of Estimators 748
Estimators and Estimates 749
Unbiasedness 750
The Sampling Variance of Estimators 752
Efficiency 754
C.3 Asymptotic or Larger Sample Properties of Estimators 755
Consistency 755
Asymptotic Normality 758
C.4 General Approaches to Parameter Estimation 760
Method of Moments 760
Maximum Likelihood 761
Least Squares 762
C.5 Interval Estimation and Confidence Intervals 762
The Nature of Interval Estimation 762
Confidence Intervals for the Mean from a Normally Distributed Population 764
A Simple Rule of Thumb for a 95% Confidence Interval 768
Asymptotic Confidence Intervals for Nonnormal Populations 768
C.6 Hypothesis Testing 770
Fundamentals of Hypothesis Testing 770
Testing Hypotheses about the Mean in a Normal Population 772
Asymptotic Tests for Nonnormal Populations 774
Computing and Using p-Values 776
The Relationship between Confidence Intervals and Hypothesis Testing 779
Practical versus Statistical Significance 780
C.7 Remarks on Notation 781
Summary 782
Key Terms 782
Problems 783

APPENDIX D
Summary of Matrix Algebra 788

D.1 Basic Definitions 788
D.2 Matrix Operations 789
Matrix Addition 789
Scalar Multiplication 790
Matrix Multiplication 790
Transpose 791
Partitioned Matrix Multiplication 792
Trace 792
Inverse 792
D.3 Linear Independence and Rank of a Matrix 793
D.4 Quadratic Forms and Positive Definite Matrices 793
D.5 Idempotent Matrices 794
D.6 Differentiation of Linear and Quadratic Forms 795
D.7 Moments and Distributions of Random Vectors 795
Expected Value 795
Variance-Covariance Matrix 795
Multivariate Normal Distribution 796

Chi-Square Distribution 796
t Distribution 797
F Distribution 797
Summary 797
Key Terms 797
Problems 798

APPENDIX E
The Linear Regression Model in Matrix Form 799

E.1 The Model and Ordinary Least Squares Estimation 799
E.2 Finite Sample Properties of OLS 801
E.3 Statistical Inference 805
E.4 Some Asymptotic Analysis 807
Wald Statistics for Testing Multiple Hypotheses 809
Summary 810
Key Terms 811
Problems 811

APPENDIX F
Answers to Chapter Questions 813

APPENDIX G
Statistical Tables 823

References 830
Glossary 835
Index 849