

Preface xvii
 About the Author xxvii

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* 11
- 1.4 Causality and the Notion of Ceteris Paribus in Econometric Analysis 12
- Summary 16
- Key Terms 17
- Problems 17
- Computer Exercises 17

PART 1

Regression Analysis with Cross-Sectional Data 19

CHAPTER 2 The Simple Regression Model 20

- 2.1 Definition of the Simple Regression Model 20
- 2.2 Deriving the Ordinary Least Squares Estimates 25
 - A Note on Terminology* 32

- 2.3 Properties of OLS on Any Sample of Data 33
 - Fitted Values and Residuals* 33
 - Algebraic Properties of OLS Statistics* 34
 - Goodness-of-Fit* 36
- 2.4 Units of Measurement and Functional Form 37
 - The Effects of Changing Units of Measurement on OLS Statistics* 38
 - Incorporating Nonlinearities in Simple Regression* 39
 - The Meaning of "Linear" Regression* 42
- 2.5 Expected Values and Variances of the OLS Estimators 43
 - Unbiasedness of OLS* 43
 - Variances of the OLS Estimators* 48
 - Estimating the Error Variance* 52
- 2.6 Regression through the Origin and Regression on a Constant 55
- Summary 56
- Key Terms 57
- Problems 58
- Computer Exercises 61

CHAPTER 3 Multiple Regression Analysis: Estimation 64

- 3.1 Motivation for Multiple Regression 65
 - The Model with Two Independent Variables* 65
 - The Model with k Independent Variables* 67
- 3.2 Mechanics and Interpretation of Ordinary Least Squares 68
 - Obtaining the OLS Estimates* 68
 - Interpreting the OLS Regression Equation* 70
 - On the Meaning of "Holding Other Factors Fixed" in Multiple Regression* 72
 - Changing More Than One Independent Variable Simultaneously* 73

	<i>OLS Fitted Values and Residuals</i>	73
	<i>A “Partialling Out” Interpretation of Multiple Regression</i>	74
	<i>Comparison of Simple and Multiple Regression Estimates</i>	74
	<i>Goodness-of-Fit</i>	76
	<i>Regression through the Origin</i>	77
3.3	The Expected Value of the OLS Estimators <i>Including Irrelevant Variables in a Regression Model</i>	79 84
	<i>Omitted Variable Bias: The Simple Case</i>	84
	<i>Omitted Variable Bias: More General Cases</i>	87
3.4	The Variance of the OLS Estimators <i>The Components of the OLS Variances: Multicollinearity</i>	89 90
	<i>Variances in Misspecified Models</i>	94
	<i>Estimating σ^2: Standard Errors of the OLS Estimators</i>	95
3.5	Efficiency of OLS: The Gauss-Markov Theorem	97
3.6	Some Comments on the Language of Multiple Regression Analysis	99
	Summary	100
	Key Terms	101
	Problems	102
	Computer Exercises	106

CHAPTER 4 Multiple Regression Analysis: Inference 110

4.1	Sampling Distributions of the OLS Estimators	110
4.2	Testing Hypotheses about a Single Population Parameter: The <i>t</i> Test <i>Testing against One-Sided Alternatives</i> <i>Two-Sided Alternatives</i> <i>Testing Other Hypotheses about β_j</i> <i>Computing <i>p</i>-Values for <i>t</i> Tests</i> <i>A Reminder on the Language of Classical Hypothesis Testing</i> <i>Economic, or Practical, versus Statistical Significance</i>	113 115 120 122 125 127 127
4.3	Confidence Intervals	130
4.4	Testing Hypotheses about a Single Linear Combination of the Parameters	132

4.5	Testing Multiple Linear Restrictions: <i>The <i>F</i> Test</i> <i>Testing Exclusion Restrictions</i> <i>Relationship between <i>F</i> and <i>t</i> Statistics</i> <i>The <i>R</i>-Squared Form of the <i>F</i> Statistic</i> <i>Computing <i>p</i>-Values for <i>F</i> Tests</i> <i>The <i>F</i> Statistic for Overall Significance of a Regression</i> <i>Testing General Linear Restrictions</i>	135 135 141 142 143 144 145
4.6	Reporting Regression Results	146
	Summary	149
	Key Terms	151
	Problems	151
	Computer Exercises	156

CHAPTER 5 Multiple Regression Analysis: OLS Asymptotics 160

5.1	Consistency <i>Deriving the Inconsistency in OLS</i>	161 164
5.2	Asymptotic Normality and Large Sample Inference <i>Other Large Sample Tests: The Lagrange Multiplier Statistic</i>	165 170
5.3	Asymptotic Efficiency of OLS	173
	Summary	174
	Key Terms	175
	Problems	175
	Computer Exercises	175

CHAPTER 6 Multiple Regression Analysis: Further Issues 178

6.1	Effects of Data Scaling on OLS Statistics <i>Beta Coefficients</i>	178 181
6.2	More on Functional Form <i>More on Using Logarithmic Functional Forms</i> <i>Models with Quadratics</i> <i>Models with Interaction Terms</i>	183 183 186 190
6.3	More on Goodness-of-Fit and Selection of Regressors <i>Adjusted <i>R</i>-Squared</i> <i>Using Adjusted <i>R</i>-Squared to Choose between Nonnested Models</i>	192 194 195

Controlling for Too Many Factors in Regression Analysis 197

Adding Regressors to Reduce the Error Variance 198

- 6.4 Prediction and Residual Analysis 199
Confidence Intervals for Predictions 199
Residual Analysis 203
Predicting y When $\log(y)$ Is the Dependent Variable 204

Summary 208

Key Terms 209

Problems 210

Computer Exercises 212

CHAPTER 7 Multiple Regression Analysis with Qualitative Information: Binary (or Dummy) Variables 217

7.1 Describing Qualitative Information 217

7.2 A Single Dummy Independent Variable 218

Interpreting Coefficients on Dummy Explanatory Variables When the Dependent Variable Is $\log(y)$ 223

7.3 Using Dummy Variables for Multiple Categories 225

Incorporating Ordinal Information by Using Dummy Variables 227

7.4 Interactions Involving Dummy Variables 230

Interactions among Dummy Variables 230

Allowing for Different Slopes 231

Testing for Differences in Regression Functions across Groups 235

7.5 A Binary Dependent Variable: The Linear Probability Model 238

7.6 More on Policy Analysis and Program Evaluation 243

7.7 Interpreting Regression Results with Discrete Dependent Variables 246

Summary 247

Key Terms 248

Problems 248

Computer Exercises 251

CHAPTER 8 Heteroskedasticity 258

8.1 Consequences of Heteroskedasticity for OLS 258

8.2 Heteroskedasticity-Robust Inference after OLS Estimation 259

Computing Heteroskedasticity-Robust LM Tests 264

8.3 Testing for Heteroskedasticity 265

The White Test for Heteroskedasticity 269

8.4 Weighted Least Squares Estimation 270

The Heteroskedasticity Is Known up to a Multiplicative Constant 271

The Heteroskedasticity Function Must Be Estimated: Feasible GLS 276

What If the Assumed Heteroskedasticity Function Is Wrong? 280

Prediction and Prediction Intervals with Heteroskedasticity 282

8.5 The Linear Probability Model Revisited 284

Summary 286

Key Terms 287

Problems 287

Computer Exercises 289

CHAPTER 9 More on Specification and Data Issues 293

9.1 Functional Form Misspecification 294

RESET as a General Test for Functional Form Misspecification 296

Tests against Nonnested Alternatives 297

9.2 Using Proxy Variables for Unobserved Explanatory Variables 298

Using Lagged Dependent Variables as Proxy Variables 303

A Different Slant on Multiple Regression 304

9.3 Models with Random Slopes 305

9.4 Properties of OLS under Measurement Error 307

Measurement Error in the Dependent Variable 308

Measurement Error in an Explanatory Variable 310

9.5 Missing Data, Nonrandom Samples, and Outlying Observations 314

Missing Data 314
Nonrandom Samples 314
Outliers and Influential Observations 316

- 9.6** Least Absolute Deviations Estimation 321
Summary 324
Key Terms 325
Problems 325
Computer Exercises 327

PART 2

Regression Analysis with Time Series Data 331

CHAPTER 10 Basic Regression Analysis with Time Series Data 332

- 10.1** The Nature of Time Series Data 332
- 10.2** Examples of Time Series Regression Models 333
Static Models 334
Finite Distributed Lag Models 334
A Convention about the Time Index 337
- 10.3** Finite Sample Properties of OLS under Classical Assumptions 337
Unbiasedness of OLS 337
The Variances of the OLS Estimators and the Gauss-Markov Theorem 340
Inference under the Classical Linear Model Assumptions 343
- 10.4** Functional Form, Dummy Variables, and Index Numbers 344
- 10.5** Trends and Seasonality 351
Characterizing Trending Time Series 351
Using Trending Variables in Regression Analysis 354
A Detrending Interpretation of Regressions with a Time Trend 356
Computing R-Squared when the Dependent Variable Is Trending 358
Seasonality 359
- Summary 361
Key Terms 362
Problems 363
Computer Exercises 364

CHAPTER 11 Further Issues in Using OLS with Time Series Data 368

- 11.1** Stationary and Weakly Dependent Time Series 369
Stationary and Nonstationary Time Series 369
Weakly Dependent Time Series 370
- 11.2** Asymptotic Properties of OLS 372
- 11.3** Using Highly Persistent Time Series in Regression Analysis 379
Highly Persistent Time Series 379
Transformations on Highly Persistent Time Series 383
Deciding Whether a Time Series Is $I(1)$ 384
- 11.4** Dynamically Complete Models and the Absence of Serial Correlation 387
- 11.5** The Homoskedasticity Assumption for Time Series Models 390
- Summary 390
Key Terms 392
Problems 392
Computer Exercises 394

CHAPTER 12 Serial Correlation and Heteroskedasticity in Time Series Regressions 398

- 12.1** Properties of OLS with Serially Correlated Errors 398
Unbiasedness and Consistency 398
Efficiency and Inference 399
Goodness-of-Fit 400
Serial Correlation in the Presence of Lagged Dependent Variables 401
- 12.2** Testing for Serial Correlation 402
A t Test for $AR(1)$ Serial Correlation with Strictly Exogenous Regressors 402
The Durbin-Watson Test under Classical Assumptions 404
Testing for $AR(1)$ Serial Correlation without Strictly Exogenous Regressors 406
Testing for Higher Order Serial Correlation 407
- 12.3** Correcting for Serial Correlation with Strictly Exogenous Regressors 409
Obtaining the Best Linear Unbiased Estimator in the $AR(1)$ Model 409

Feasible GLS Estimation with AR(1) Errors 411

Comparing OLS and FGLS 413

*Correcting for Higher Order Serial
Correlation* 414

12.4 Differencing and Serial Correlation 415

12.5 Serial Correlation-Robust Inference
after OLS 417

12.6 Heteroskedasticity in Time Series
Regressions 420
Heteroskedasticity-Robust Statistics 421

Testing for Heteroskedasticity 421

Autoregressive Conditional

Heteroskedasticity 422

*Heteroskedasticity and Serial Correlation in
Regression Models* 424

Summary 425

Key Terms 426

Problems 426

Computer Exercises 427

PART 3

Advanced Topics 431

CHAPTER 13 Pooling Cross Sections across Time: Simple Panel Data Methods 432

13.1 Pooling Independent Cross Sections across
Time 433

*The Chow Test for Structural Change across
Time* 437

13.2 Policy Analysis with Pooled Cross Sections 438

13.3 Two-Period Panel Data Analysis 443
Organizing Panel Data 449

13.4 Policy Analysis with Two-Period Panel Data 449

13.5 Differencing with More Than Two Time
Periods 452
*Potential Pitfalls in First Differencing Panel
Data* 457

Summary 458

Key Terms 458

Problems 458

Computer Exercises 460

CHAPTER 14 Advanced Panel Data Methods 466

14.1 Fixed Effects Estimation 466

The Dummy Variable Regression 470

Fixed Effects or First Differencing? 471

Fixed Effects with Unbalanced Panels 473

14.2 Random Effects Models 474

Random Effects or Fixed Effects? 477

14.3 The Correlated Random Effects
Approach 479

14.4 Applying Panel Data Methods to Other Data
Structures 481

Summary 483

Key Terms 484

Problems 484

Computer Exercises 485

CHAPTER 15 Instrumental Variables Estimation and Two Stage Least Squares 490

15.1 Motivation: Omitted Variables in a Simple
Regression Model 491

Statistical Inference with the IV

Estimator 495

*Properties of IV with a Poor Instrumental
Variable* 499

Computing R-Squared after

IV Estimation 501

15.2 IV Estimation of the Multiple Regression
Model 502

15.3 Two Stage Least Squares 506

*A Single Endogenous Explanatory
Variable* 506

Multicollinearity and 2SLS 508

*Multiple Endogenous Explanatory
Variables* 509

*Testing Multiple Hypotheses after 2SLS
Estimation* 510

15.4 IV Solutions to Errors-in-Variables
Problems 510

15.5 Testing for Endogeneity and Testing
Overidentifying Restrictions 512

<i>Testing for Endogeneity</i>	512
<i>Testing Overidentification Restrictions</i>	513
15.6	2SLS with Heteroskedasticity 516
15.7	Applying 2SLS to Time Series Equations 516
15.8	Applying 2SLS to Pooled Cross Sections and Panel Data 518
	Summary 520
	Key Terms 521
	Problems 521
	Computer Exercises 524

CHAPTER 16 Simultaneous Equations Models 530

16.1	The Nature of Simultaneous Equations Models 531
16.2	Simultaneity Bias in OLS 534
16.3	Identifying and Estimating a Structural Equation 536
	<i>Identification in a Two-Equation System</i> 536
	<i>Estimation by 2SLS</i> 541
16.4	Systems with More Than Two Equations 543
	<i>Identification in Systems with Three or More Equations</i> 543
	<i>Estimation</i> 544
16.5	Simultaneous Equations Models with Time Series 544
16.6	Simultaneous Equations Models with Panel Data 548
	Summary 550
	Key Terms 551
	Problems 551
	Computer Exercises 554

CHAPTER 17 Limited Dependent Variable Models and Sample Selection Corrections 559

17.1	Logit and Probit Models for Binary Response 560
	<i>Specifying Logit and Probit Models</i> 560

	<i>Maximum Likelihood Estimation of Logit and Probit Models</i> 563
	<i>Testing Multiple Hypotheses</i> 564
	<i>Interpreting the Logit and Probit Estimates</i> 565
17.2	The Tobit Model for Corner Solution Responses 572
	<i>Interpreting the Tobit Estimates</i> 574
	<i>Specification Issues in Tobit Models</i> 579
17.3	The Poisson Regression Model 580
17.4	Censored and Truncated Regression Models 585
	<i>Censored Regression Models</i> 585
	<i>Truncated Regression Models</i> 589
17.5	Sample Selection Corrections 591
	<i>When Is OLS on the Selected Sample Consistent?</i> 591
	<i>Incidental Truncation</i> 593
	Summary 597
	Key Terms 598
	Problems 598
	Computer Exercises 600

CHAPTER 18 Advanced Time Series Topics 606

18.1	Infinite Distributed Lag Models 607
	<i>The Geometric (or Koyck) Distributed Lag</i> 609
	<i>Rational Distributed Lag Models</i> 611
18.2	Testing for Unit Roots 613
18.3	Spurious Regression 618
18.4	Cointegration and Error Correction Models 620
	<i>Cointegration</i> 620
	<i>Error Correction Models</i> 625
18.5	Forecasting 626
	<i>Types of Regression Models Used for Forecasting</i> 628
	<i>One-Step-Ahead Forecasting</i> 629
	<i>Comparing One-Step-Ahead Forecasts</i> 632
	<i>Multiple-Step-Ahead Forecasts</i> 634
	<i>Forecasting Trending, Seasonal, and Integrated Processes</i> 636

Summary	641
Key Terms	643
Problems	643
Computer Exercises	645

CHAPTER 19 Carrying Out an Empirical Project 650

19.1 Posing a Question	650
19.2 Literature Review	652
19.3 Data Collection	653
<i>Deciding on the Appropriate Data Set</i>	653
<i>Entering and Storing Your Data</i>	654
<i>Inspecting, Cleaning, and Summarizing Your Data</i>	656
19.4 Econometric Analysis	657
19.5 Writing an Empirical Paper	660
<i>Introduction</i>	660
<i>Conceptual (or Theoretical) Framework</i>	661
<i>Econometric Models and Estimation Methods</i>	661
<i>The Data</i>	664
<i>Results</i>	664
<i>Conclusions</i>	665
<i>Style Hints</i>	666
Summary	668
Key Terms	668
Sample Empirical Projects	668
List of Journals	674
Data Sources	675

CHAPTER APPENDICES

Appendix 2A	678
Appendix 3A	680
Appendix 5A	684
Appendix 6A	685
Appendix 13A	686
Appendix 14A	689
Appendix 15A	692
Appendix 17A	694
Appendix 17B	695

APPENDIX A Basic Mathematical Tools 697

A.1 The Summation Operator and Descriptive Statistics	697
A.2 Properties of Linear Functions	699
A.3 Proportions and Percentages	701
A.4 Some Special Functions and Their Properties	704
<i>Quadratic Functions</i>	704
<i>The Natural Logarithm</i>	706
<i>The Exponential Function</i>	710
A.5 Differential Calculus	711
Summary	713
Key Terms	713
Problems	713

APPENDIX B Fundamentals of Probability 716

B.1 Random Variables and Their Probability Distributions	716
<i>Discrete Random Variables</i>	717
<i>Continuous Random Variables</i>	719
B.2 Joint Distributions, Conditional Distributions, and Independence	721
<i>Joint Distributions and Independence</i>	721
<i>Conditional Distributions</i>	723
B.3 Features of Probability Distributions	724
<i>A Measure of Central Tendency: The Expected Value</i>	724
<i>Properties of Expected Values</i>	725
<i>Another Measure of Central Tendency: The Median</i>	727
<i>Measures of Variability: Variance and Standard Deviation</i>	728
<i>Variance</i>	728
<i>Standard Deviation</i>	730
<i>Standardizing a Random Variable</i>	730
<i>Skewness and Kurtosis</i>	731
B.4 Features of Joint and Conditional Distributions	731

	<i>Measures of Association: Covariance and Correlation</i>	731
	<i>Covariance</i>	731
	<i>Correlation Coefficient</i>	733
	<i>Variance of Sums of Random Variables</i>	734
	<i>Conditional Expectation</i>	735
	<i>Properties of Conditional Expectation</i>	736
	<i>Conditional Variance</i>	738
B.5	The Normal and Related Distributions	739
	<i>The Normal Distribution</i>	739
	<i>The Standard Normal Distribution</i>	740
	<i>Additional Properties of the Normal Distribution</i>	742
	<i>The Chi-Square Distribution</i>	743
	<i>The t Distribution</i>	743
	<i>The F Distribution</i>	744
	Summary	746
	Key Terms	746
	Problems	746

APPENDIX C Fundamentals of Mathematical Statistics 749

C.1	Populations, Parameters, and Random Sampling	749
	<i>Sampling</i>	750
C.2	Finite Sample Properties of Estimators	750
	<i>Estimators and Estimates</i>	751
	<i>Unbiasedness</i>	752
	<i>The Sampling Variance of Estimators</i>	754
	<i>Efficiency</i>	756
C.3	Asymptotic or Large Sample Properties of Estimators	757
	<i>Consistency</i>	757
	<i>Asymptotic Normality</i>	760
C.4	General Approaches to Parameter Estimation	762
	<i>Method of Moments</i>	762
	<i>Maximum Likelihood</i>	763
	<i>Least Squares</i>	764
C.5	Interval Estimation and Confidence Intervals	764

	<i>The Nature of Interval Estimation</i>	764
	<i>Confidence Intervals for the Mean from a Normally Distributed Population</i>	766
	<i>A Simple Rule of Thumb for a 95% Confidence Interval</i>	769
	<i>Asymptotic Confidence Intervals for Nonnormal Populations</i>	770
C.6	Hypothesis Testing	771
	<i>Fundamentals of Hypothesis Testing</i>	772
	<i>Testing Hypotheses about the Mean in a Normal Population</i>	774
	<i>Asymptotic Tests for Nonnormal Populations</i>	777
	<i>Computing and Using p-Values</i>	778
	<i>The Relationship between Confidence Intervals and Hypothesis Testing</i>	781
	<i>Practical versus Statistical Significance</i>	782
C.7	Remarks on Notation	783
	Summary	784
	Key Terms	784
	Problems	785

APPENDIX D Summary of Matrix Algebra 790

D.1	Basic Definitions	790
D.2	Matrix Operations	791
	<i>Matrix Addition</i>	791
	<i>Scalar Multiplication</i>	792
	<i>Matrix Multiplication</i>	792
	<i>Transpose</i>	793
	<i>Partitioned Matrix Multiplication</i>	794
	<i>Trace</i>	794
	<i>Inverse</i>	795
D.3	Linear Independence and Rank of a Matrix	795
D.4	Quadratic Forms and Positive Definite Matrices	796
D.5	Idempotent Matrices	796
D.6	Differentiation of Linear and Quadratic Forms	797
D.7	Moments and Distributions of Random Vectors	797

Expected Value 797
Variance-Covariance Matrix 797
Multivariate Normal Distribution 798
Chi-Square Distribution 798
t Distribution 799
F Distribution 799

Summary 799
Key Terms 799
Problems 800

APPENDIX E The Linear Regression Model in Matrix Form 801

E.1 The Model and Ordinary Least Squares Estimation 801
E.2 Finite Sample Properties of OLS 803
E.3 Statistical Inference 807

E.4 Some Asymptotic Analysis 809
Wald Statistics for Testing Multiple Hypotheses 812
Summary 813
Key Terms 813
Problems 813

APPENDIX F Answers to Chapter Questions 815

APPENDIX G Statistical Tables 825

References 832
Glossary 838
Index 856