BRIEF CONTENTS

Chapter 1	The Nature of Econometrics and Economic Data	1
PART 1: Regression Analysis with Cross-Sectional Data		**17**
Chapter 2	The Simple Regression Model	18
Chapter 3	Multiple Regression Analysis: Estimation	56
Chapter 4	Multiple Regression Analysis: Inference	94
Chapter 5	Multiple Regression Analysis: OLS Asymptotics	135
Chapter 6	Multiple Regression Analysis: Further Issues	150
Chapter 7	Multiple Regression Analysis with Qualitative Information: Binary (or Dummy) Variables	182
Chapter 8	Heteroskedasticity	212
Chapter 9	More on Specification and Data Issues	241
PART 2: Regression Analysis with Time Series Data		**273**
Chapter 10	Basic Regression Analysis with Time Series Data	274
Chapter 11	Further Issues in Using OLS with Time Series Data	305
Chapter 12	Serial Correlation and Heteroskedasticity in Time Series Regressions	330
PART 3: Advanced Topics		**359**
Chapter 13	Pooling Cross Sections Across Time: Simple Panel Data Methods	360
Chapter 14	Advanced Panel Data Methods	387
Chapter 15	Instrumental Variables Estimation and Two Stage Least Squares	405
Chapter 16	Simultaneous Equations Models	437
Chapter 17	Limited Dependent Variable Models and Sample Selection Corrections	459
Chapter 18	Advanced Time Series Topics	498
Chapter 19	Carrying Out an Empirical Project	535
References		**561**
Glossary		**567**
Index		**585**

CONTENTS

Preface xi

About the Author xviii

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

PART 1

Regression Analysis with Cross-Sectional Data 17

CHAPTER 2 The Simple Regression Model 18

2.1 Definition of the Simple Regression Model 18

2.2 Deriving the Ordinary Least Squares Estimates 23

A Note on Terminology 30

2.3 Properties of OLS on Any Sample of Data 31

Fitted Values and Residuals 31

Algebraic Properties of OLS Statistics 32

Goodness-of-Fit 34

2.4 Units of Measurement and Functional Form 35

The Effects of Changing Units of Measurement on OLS Statistics 36

Incorporating Nonlinearities in Simple Regression 37

The Meaning of "Linear" Regression 40

2.5 Expected Values and Variances of the OLS Estimators 41

Unbiasedness of OLS 41

Variances of the OLS Estimators 46

Estimating the Error Variance 50

2.6 Regression through the Origin and Regression on a Constant 53

Summary 54

CHAPTER 3 Multiple Regression Analysis: Estimation 56

3.1 Motivation for Multiple Regression 57

The Model with Two Independent Variables 57

The Model with k *Independent Variables 59*

3.2 Mechanics and Interpretation of Ordinary Least Squares 60

Obtaining the OLS Estimates 60

Interpreting the OLS Regression Equation 62

On the Meaning of "Holding Other Factors Fixed" in Multiple Regression 64

Changing More Than One Independent Variable Simultaneously 65

OLS Fitted Values and Residuals 65

A "Partialling Out" Interpretation of Multiple Regression 66

Comparison of Simple and Multiple Regression Estimates 66

Goodness-of-Fit 68

Regression through the Origin 69

3.3 The Expected Value of the OLS Estimators 71
Including Irrelevant Variables in a Regression Model 76
Omitted Variable Bias: The Simple Case 76
Omitted Variable Bias: More General Cases 79
3.4 The Variance of the OLS Estimators 81
The Components of the OLS Variances: Multicollinearity 82
Variances in Misspecified Models 86
Estimating σ^2: Standard Errors of the OLS Estimators 87
3.5 Efficiency of OLS: The Gauss-Markov Theorem 89
3.6 Some Comments on the Language of Multiple Regression Analysis 91
Summary 92

CHAPTER 4 Multiple Regression Analysis: Inference 94

4.1 Sampling Distributions of the OLS Estimators 94
4.2 Testing Hypotheses about a Single Population Parameter: The *t* Test 97
Testing against One-Sided Alternatives 99
Two-Sided Alternatives 104
Testing Other Hypotheses about β_j 106
Computing p-Values for t Tests 109
A Reminder on the Language of Classical Hypothesis Testing 111
Economic, or Practical, versus Statistical Significance 111
4.3 Confidence Intervals 114
4.4 Testing Hypotheses about a Single Linear Combination of the Parameters 116
4.5 Testing Multiple Linear Restrictions: The *F* Test 119
Testing Exclusion Restrictions 119
Relationship between F and t Statistics 125
The R-Squared Form of the F Statistic 126
Computing p-Values for F Tests 127
The F Statistic for Overall Significance of a Regression 128
Testing General Linear Restrictions 129
4.6 Reporting Regression Results 130
Summary 133

CHAPTER 5 Multiple Regression Analysis: OLS Asymptotics 135

5.1 Consistency 136
Deriving the Inconsistency in OLS 139
5.2 Asymptotic Normality and Large Sample Inference 140
Other Large Sample Tests: The Lagrange Multiplier Statistic 145
5.3 Asymptotic Efficiency of OLS 148
Summary 149

CHAPTER 6 Multiple Regression Analysis: Further Issues 150

6.1 Effects of Data Scaling on OLS Statistics 150
Beta Coefficients 153
6.2 More on Functional Form 155
More on Using Logarithmic Functional Forms 155
Models with Quadratics 158
Models with Interaction Terms 162
6.3 More on Goodness-of-Fit and Selection of Regressors 164
Adjusted R-Squared 166
Using Adjusted R-Squared to Choose between Nonnested Models 167
Controlling for Too Many Factors in Regression Analysis 169
Adding Regressors to Reduce the Error Variance 170
6.4 Prediction and Residual Analysis 171
Confidence Intervals for Predictions 171
Residual Analysis 175
Predicting y When log(y) Is the Dependent Variable 176
Summary 180

CHAPTER 7 Multiple Regression Analysis with Qualitative Information: Binary (or Dummy) Variables 182

7.1 Describing Qualitative Information 182
7.2 A Single Dummy Independent Variable 183
Interpreting Coefficients on Dummy Explanatory Variables When the Dependent Variable Is log(y) 188

7.3 Using Dummy Variables for Multiple Categories 190
Incorporating Ordinal Information by Using Dummy Variables 192
7.4 Interactions Involving Dummy Variables 195
Interactions among Dummy Variables 195
Allowing for Different Slopes 196
Testing for Differences in Regression Functions across Groups 199
7.5 A Binary Dependent Variable: The Linear Probability Model 202
7.6 More on Policy Analysis and Programme Evaluation 207
7.7 Interpreting Regression Results with Discrete Dependent Variables 209
Summary 211

CHAPTER 8 Heteroskedasticity 212

8.1 Consequences of Heteroskedasticity for OLS 212
8.2 Heteroskedasticity-Robust Inference after OLS Estimation 213
Computing Heteroskedasticity-Robust LM *Tests 218*
8.3 Testing for Heteroskedasticity 219
The White Test for Heteroskedasticity 223
8.4 Weighted Least Squares Estimation 224
The Heteroskedasticity Is Known up to a Multiplicative Constant 225
The Heteroskedasticity Function Must Be Estimated: Feasible GLS 230
What If the Assumed Heteroskedasticity Function Is Wrong? 234
Prediction and Prediction Intervals with Heteroskedasticity 236
8.5 The Linear Probability Model Revisited 238
Summary 240

CHAPTER 9 More on Specification and Data Issues 241

9.1 Functional Form Misspecification 242
RESET as a General Test for Functional Form Misspecification 244
Tests against Nonnested Alternatives 245
9.2 Using Proxy Variables for Unobserved Explanatory Variables 246
Using Lagged Dependent Variables as Proxy Variables 251
A Different Slant on Multiple Regression 252
9.3 Models with Random Slopes 253
9.4 Properties of OLS under Measurement Error 255
Measurement Error in the Dependent Variable 256
Measurement Error in an Explanatory Variable 258
9.5 Missing Data, Nonrandom Samples, and Outlying Observations 262
Missing Data 262
Nonrandom Samples 262
Outliers and Influential Observations 264
9.6 Least Absolute Deviations Estimation 269
Summary 272

PART 2

Regression Analysis with Time Series Data 273

CHAPTER 10 Basic Regression Analysis with Time Series Data 274

10.1 The Nature of Time Series Data 274
10.2 Examples of Time Series Regression Models 275
Static Models 276
Finite Distributed Lag Models 276
A Convention about the Time Index 279
10.3 Finite Sample Properties of OLS under Classical Assumptions 279
Unbiasedness of OLS 279
The Variances of the OLS Estimators and the Gauss-Markov Theorem 282
Inference under the Classical Linear Model Assumptions 285
10.4 Functional Form, Dummy Variables, and Index Numbers 286
10.5 Trends and Seasonality 293
Characterising Trending Time Series 293
Using Trending Variables in Regression Analysis 296

A Detrending Interpretation of Regressions with a Time Trend 298
Computing R-Squared when the Dependent Variable Is Trending 300
Seasonality 301
Summary 303

CHAPTER 11 Further Issues in Using OLS with Time Series Data 305

11.1 Stationary and Weakly Dependent Time Series 306
Stationary and Nonstationary Time Series 306
Weakly Dependent Time Series 307
11.2 Asymptotic Properties of OLS 309
11.3 Using Highly Persistent Time Series in Regression Analysis 316
Highly Persistent Time Series 316
Transformations on Highly Persistent Time Series 320
Deciding Whether a Time Series Is I(1) 321
11.4 Dynamically Complete Models and the Absence of Serial Correlation 324
11.5 The Homoskedasticity Assumption for Time Series Models 327
Summary 327

CHAPTER 12 Serial Correlation and Heteroskedasticity in Time Series Regressions 330

12.1 Properties of OLS with Serially Correlated Errors 330
Unbiasedness and Consistency 330
Efficiency and Inference 331
Goodness-of-Fit 332
Serial Correlation in the Presence of Lagged Dependent Variables 333
12.2 Testing for Serial Correlation 334
A t Test for AR(1) Serial Correlation with Strictly Exogenous Regressors 334
The Durbin-Watson Test under Classical Assumptions 336
Testing for AR(1) Serial Correlation without Strictly Exogenous Regressors 338
Testing for Higher Order Serial Correlation 339
12.3 Correcting for Serial Correlation with Strictly Exogenous Regressors 341
Obtaining the Best Linear Unbiased Estimator in the AR(1) Model 341
Feasible GLS Estimation with AR(1) Errors 343
Comparing OLS and FGLS 345
Correcting for Higher Order Serial Correlation 346
12.4 Differencing and Serial Correlation 347
12.5 Serial Correlation-Robust Inference after OLS 349
12.6 Heteroskedasticity in Time Series Regressions 352
Heteroskedasticity-Robust Statistics 353
Testing for Heteroskedasticity 353
Autoregressive Conditional Heteroskedasticity 354
Heteroskedasticity and Serial Correlation in Regression Models 356
Summary 357

PART 3

Advanced Topics 359

CHAPTER 13 Pooling Cross Sections across Time: Simple Panel Data Methods 360

13.1 Pooling Independent Cross Sections across Time 361
The Chow Test for Structural Change across Time 365
13.2 Policy Analysis with Pooled Cross Sections 366
13.3 Two-Period Panel Data Analysis 371
Organising Panel Data 377
13.4 Policy Analysis with Two-Period Panel Data 377
13.5 Differencing with More Than Two Time Periods 380
Potential Pitfalls in First Differencing Panel Data 385
Summary 386

CHAPTER 14 Advanced Panel Data Methods 387

14.1 Fixed Effects Estimation 387
The Dummy Variable Regression 391
Fixed Effects or First Differencing? 392
Fixed Effects with Unbalanced Panels 394
14.2 Random Effects Models 395
Random Effects or Fixed Effects? 398
14.3 The Correlated Random Effects Approach 400
14.4 Applying Panel Data Methods to Other Data Structures 402
Summary 404

CHAPTER 15 Instrumental Variables Estimation and Two Stage Least Squares 405

15.1 Motivation: Omitted Variables in a Simple Regression Model 406
Statistical Inference with the IV Estimator 410
Properties of IV with a Poor Instrumental Variable 414
Computing R-*Squared after IV Estimation 416*
15.2 IV Estimation of the Multiple Regression Model 417
15.3 Two Stage Least Squares 421
A Single Endogenous Explanatory Variable 421
Multicollinearity and 2SLS 423
Multiple Endogenous Explanatory Variables 424
Testing Multiple Hypotheses after 2SLS Estimation 424
15.4 IV Solutions to Errors-in-Variables Problems 425
15.5 Testing for Endogeneity and Testing Overidentifying Restrictions 427
Testing for Endogeneity 427
Testing Overidentification Restrictions 428
15.6 2SLS with Heteroskedasticity 431
15.7 Applying 2SLS to Time Series Equations 431
15.8 Applying 2SLS to Pooled Cross Sections and Panel Data 433
Summary 435

CHAPTER 16 Simultaneous Equations Models 437

16.1 The Nature of Simultaneous Equations Models 438
16.2 Simultaneity Bias in OLS 441
16.3 Identifying and Estimating a Structural Equation 443
Identification in a Two-Equation System 443
Estimation by 2SLS 448
16.4 Systems with More Than Two Equations 450
Identification in Systems with Three or More Equations 450
Estimation 451
16.5 Simultaneous Equations Models with Time Series 451
16.6 Simultaneous Equations Models with Panel Data 455
Summary 457

CHAPTER 17 Limited Dependent Variable Models and Sample Selection Corrections 459

17.1 Logit and Probit Models for Binary Response 460
Specifying Logit and Probit Models 460
Maximum Likelihood Estimation of Logit and Probit Models 463
Testing Multiple Hypotheses 464
Interpreting the Logit and Probit Estimates 465
17.2 The Tobit Model for Corner Solution Responses 472
Interpreting the Tobit Estimates 474
Specification Issues in Tobit Models 479
17.3 The Poisson Regression Model 480
17.4 Censored and Truncated Regression Models 485