if

Introductory Econometrics

A Modern Approach

JEFFREY M. WOOLDRIDGE

Michigan State University

THOMSON

SOUTH-WESTERN

C

Australia • Canada • Mexico • Singapore • Spain • United Kingdom • United States

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 6

Time Series Data 8

Pooled Cross Sections 10

Panel or Longitudinal Data 10

A Comment on Data Structures 13

1.4 Causality and the Notion of Ceteris Paribus in Econometric Analysis 13

Summary 18

Key Terms 19

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 36

2.3 Mechanics of OLS 36

Fitted Values and Residuals 36

Algebraic Properties of OLS Statistics 38

Goodness-of-Fit 40

2.4 Units of Measurement and Functional Form 41

The Effects of Changing Units of Measurement on OLS Statistics 42

Incorporating Nonlinearities in Simple Regression 43

The Meaning of "Linear" Regression 46

2.5 Expected Values and Variances of the OLS Estimators 47

Unbiasedness of OLS 47

Variances of the OLS Estimators 53

Estimating the Error Variance 57

2.6 Regression Through the Origin 59

Summary 60
Key Terms 61
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 7.4
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 Bids: More General Cases 93
3.4- The Variance of the OLS Estimators 95
The Components of the OLS Variances: Multicollinearity 96
Variances in Misspecified Models 100
Estimating a^2: Standard Errors of the OLS Estimators 101
3.5 Efficiency of OLS: 'The Gauss-Markov Theorem 103
Summary 104
Key Terms 105
Problems 106
Computer Exercises 110
Appendix 3A 111

Chapter 4 Multiple Regression Analysis: Inference 116
4.1 Sampling[1] Distributions of the OLS Estimators 116
4.2 Testing Hypotheses About a Single Population Parameter: The t Test 119
Testing Against One-Sided Alternatives 121
Two-Sided A Itemalives 126
Testing Other Hypotheses About fij 128
Computing p- *Values for* t *Tests 131*
A Reminder on the Language of Classical Hypothesis Testing 134
Economic, or Practical, versus Statistical Significance 134
4.3 Confidence Intervals 137
4.4 Testing Hypotheses About a Single Linear Combination of the Parameters 139
4.5 Testing Multiple Linear Restrictions: The *F* Test 142

Testing Exclusion Restrictions 142
Relationship Between F *and* t *Statistics 148*
The K-Squared Form of the F *Statistic 149*
Computings-Valuesfor'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 157
Key Terms 157
Problems 158
Computer Exercises 163

Chapter 5 Multiple Regression Analysis: OLS Asymptotics 166
5.1 Consistency 166
Deriving the Inconsistency in OLS 169
5.2 Asymptotic Normality and Large Sample Inference 171
Other Large Sample Tests: The Lagrange Multiplier Statistic 175
5.3 Asymptotic Efficiency of OLS 177
Summary 179
Key Terms 179
Problems 180
Computer Exercises 180
Appendix 5 A 181

Chapter 6 Multiple Regression Analysis: Further Issues 182
6.1 Effects of Data Scaling, on OLS Statistics 182
Beta Coefficients 185
6.2 More on Functional Form 187
More on Using Logarithmic Functional Forms 187
Models with Quadratics 189
Models with Interaction Terms 194
6.3 More on Goodness-of-Fit and Selection of Regressors 196
Adjusted R-Squared 197
Using Adjusted K-Squared to Choose.Between Nonnested Models 198
Controlling for Too Many Factors in Regression Analysis 200
Adding Regressors to Reduce the Error Variance 202
6.4 Prediction and Residual Analysis 202
Confidence Intervalsfor Predictions 203
Residual Analysis 206
Predicting y *when logiy) Is the Dependent Variable 207*
Summary 210
Key Terms. 211
Problems 211
Computer Exercises 213

Chapter 7 • Multiple Regression Analysis with Qualitative Information: Binary (or Dummy) Variables 218
7.1 Describing Qualitative Information 218

7.2 A Single Dummy Independent Variable 220
Interpreting Coefficients on Dummy Explanatory Variables when the Dependent Variable Is log(y) 225
7.3 Using Dummy Variables for Multiple Categories 227
Incorporating Ordinal Information by Using Dummy Variables 228
7.4 Interactions Involving Dummy Variables 232
Interactions Among Dummy Variables 232
Allowing for Different Slopes 233
Testing for Differences in Regression Functions Across Groups .237
7.5 A Biliary Dependent Variable: The Linear Probability Model 240
7.6 More on Policy Analysis and Program Evaluation 245
Summary 248
Key Terms 249
Problems 249
Computer Exercises 252

Chapter 8 Heteroskedasticity 257
8.1 Consequences of Heteroskedasticity for OLS 257
8.2 Heteroskedasticity-Robust Inference After OLS Estimation 258
Computing Heteroskedasticity-Robust LM *Tests 262*
8.3 Testing for Heteroskedasticity 264
The White Test for Heteroskedasticity 268
8.4 Weighted Least Squares Estimation 270
The Heteroskedasticity Is Known up to a Multiplicative Constant 270
The Heteroskedasticity Function Must Be Estimated: Feasible GLS 276
8.5 The Linear Probability Model Revisited 280
Summary 283
Key Terms 283
Problems 284
Computer Exercises 285

Chapter 9 More on Specification and Data Problems 289
9.1 Functional Form Misspecification 289
RESET as a General-Test for Functional Form Misspecification 292
Tests Against Nonnested Alternatives 294
9.2 Using Proxy Variables for Unobserved Explanatory Variables 295
Using Lagged Dependent Variables as Proxy Variables 300
9.3 Properties of OLS Under Measurement Error 302
Measurement Error in the Dependent Variable 302
Measurement Error in an Explanatory Variable 305
9.4 Missing Data, Nonrandom Samples, and Outlying Observations 309
Missing Data 309
Nonrandom Samples 310
Outliers and Influential Observations 312
Summary 317

Key Terms 317
Problems 318-
Computer Exercises . ' 320

PART 2

REGRESSION ANALYSIS WITH TIME SERIES DATA 323

Chapter 10 Basic Regression Analysis with Time Series Data 324
10.1 • The Nature of Time Series Data 324
10.2 Examples of Time Series Regression Models 326
Static Models 326
Finite Distributed Lag Models 326
A Convention about the Time Index 329.
10.3 Finite Sample Properties of OLS Under Classical Assumptions 329
Unbiasedness of OLS 329
The Variances of the OLS Estimators and the Gauss-Markov Theorem 333
Inference under the Classical Linear Model Assumptions 335
10.4 Functional Form, Dummy Variables, and Index Numbers 337
10.5 Trends and Seasonally 344
Characterizing Trending Time Series 344
Using Trending Variables in Regression Analysis 347
A Detrending Interpretation of Regressions with a Time Trend 350
Computing R-Squaredwhen the Dependent Variable Is Trending 351
Seasonality 353
Summary 355
Key Terms 355
Problems 356
Computer Exercises . 357

Chapter 11 Further Issues in Using OLS with Time Series Data 360
11.1 Stationary and Weakly Dependent Time Series 360
Stationary and Nonstationary Time Series 361
Weakly Dependent Time Series 362
11.2 Asymptotic Properties of OLS . 365
11.3 Using Highly Persistent Time Series in Regression Analysis 372
Highly Persistent Time Series 372
Transformations on Highly Persistent Time Series 377
Deciding Whether a Time Series Is 1(1) 378 .
11.4 Dynamically Complete Models and the Absence of Serial Correlation 380
11.5 The Homoskedasticity Assumption for Time Series Models 382
Summary 383
Key Terms ' 384
Problems 384
Computer Exercises 387

Chapter 12 Serial Correlation and Heteroskedasticity in Time Series Regressions 391

12.1 Properties of OLS with Serially Correlated Errors 391
Unbiasedness and Consistency 391
Efficiency and Inference 392
Goodness-of-Fit 393
Serial Correlation in the Presence of Lagged Dependent Variables 394

12.2 Testing for Serial Correlation 395
A t Test for AR(1) Serial Correlation with Strictly Exogenous Regressors 395
The Durbin-Watson Test under Classical Assumptions 397
Testing for AR(1) Serial Correlation without Strictly Exogenous Regressors 399
Testing for Higher Order Serial Correlation 400

12.3 Correcting for Serial Correlation with Strictly Exogenous Regressors 402
Obtaining the Best Linear Unbiased Estimator in the AR(1) Model 402
Feasible GLS Estimation with AR(1) Errors 404
Comparing OLS and FGLS 406
Correcting for Higher Order Serial Correlation 408

12.4 Differencing and Serial Correlation 409

12.5 Serial Correlation-Robust Inference After OLS 410

12.6 Heteroskedasticity in Time Series Regressions 414
Heteroskedasticity-Robust Statistics 414
Testing for Heteroskedasticity 414
Autoregressive Conditional Heteroskedasticity 416
Heteroskedasticity and Serial Correlation in Regression Models 418

Summary 419
Key Terms 420
Problems 420
Computer Exercises 421

PART 3

ADVANCED TOPICS 425

Chapter 13 Pooling Cross Sections Across Time. Simple Panel Data Methods 426

13.1 Pooling Independent Cross Sections Across Time 427
The Chow Test for Structural Change Across Time 431

13.2 Policy Analysis with Pooled Cross Sections 432

13.3 Two-Period Panel Data Analysis 438
Organizing Panel Data 444

13.4 Policy Analysis with Two-Period Panel Data 445

13.5 Differencing with More than Two Time Periods 447

Summary 452
Key Terms 453
Problems 453

- Computer Exercises 454
Appendix 13A 458

Chapter 14 Advanced Panel Data Methods 461
14.1 Fixed Effects Estimation 461
The Dummy Variable Regression 465
Fixed Effects or First Differencing ? 467
Fixed Effects with Unbalanced Panels 468
14.2 Random Effects Models 469
Random Effects or Fixed Effects? 473
14.3 Applying Panel Data Methods to Other Data Structures 473
Summary 475
Key Terms 476
Problems 476
Computer Exercises 478
Appendix 14A 481.

Chapter 15 Instrumental Variables Estimation and Two Stage Least Squares 484
15.1 Motivation: Omitted Variables in a Simple Regression Model 485
Statistical Inference, with the IV Estimator 488
Properties of IV with a Poor Instrumental Variable 493
Computing R-Squared after IV Estimation 494
15.2 IV Estimation of the Multiple Regression Model 495
.15.-3 Two Stage Least Squares 499
A *Single Endogenous.Explanatory Variable 499*
MuMcollinearity and 2SLS' 502
Multiple Endogenous Explanatory Variables 502
Testing Multiple Hypotheses after 2SLS Estimation 503
15.4 IV Solutions to Errors-in-Variables Problems 503
15.5 Testing for Endogeneity and Testing Overidentifying Restrictions 505
Testing for Endogeneity 506
Testing Over-identification Restrictions 507
15.6 2SLS with Heteroskedasticity 509
15.7 -Applying .2SLS to Time Series Equations 509
• 15.8 Applying 2SLS to Pooled Cross Sections and Panel Data 511
Summary 514
Key Terms 514
Problems 514
Computer Exercises 518
Appendix 15A 522

Chapter 16 Simultaneous Equations Models 525
16.1 The,Nature of Simultaneous Equations Models 525
16.2 Simultaneity Bias in OLS 530
16.3 Identifying and Estimating a Structural Equation 531
Identification in a Two-Equation System 532
Estimation by 2SLS 536
16.4 Systems with More than Two Equations 538
Identification in Systems with Three or More Equations 538
Estimation 539

Variance 709
Standard Deviation 711
Standardizing a Random Variable 711
B.4 Features of Joint and Conditional Distributions 712
Measures of Association: Covariance and Correlation 712
Covariance 712
Correlation Coefficient 713
Variance of Sums of Random Variables 714
Conditional Expectation 716
Properties of Conditional Expectation 718
Conditional Variance 720
B.5 The Normal and Related Distributions 720
The Normal Distribution 720
The Standard Normal Distribution 722
Additional Properties of the Normal Distribution 724
The Chi-Square Distribution 725
The t *Distribution 725*
The F *Distribution 726*
Summary 727
Key Terms 728
Problems 729

Appendix C Fundamentals of Mathematical Statistics 731
C.I Populations, Parameters, and Random Sampling 731
Sampling 732
C.2 Finite Sample Properties of Estimators 732
Estimators and Estimates 733
Unbiasedness 734
The Sampling Variance of Estimators 736
Efficiency 739
C.3 Asymptotic or Large Sample Properties of Estimators 740
Consistency 740
Asymptotic Normality 743
C.4 General Approaches to Parameter Estimation 745
Method of Moments 745
Maximum Likelihood 746
Least Squares 747
C.5 Interval Estimation and Confidence Intervals 748
The Nature of Interval Estimation 748.
Confidence Intervals for the Mean from a Normally Distributed Population 750
A Simple Ride of Thumb for a 95% Confidence Interval 754
Asymptotic Confidence Intervals for Non normal Populations 755
C.6 Hypothesis Testing 756
Fundamentals of Hypothesis Testing- 756
Testing Hypotheses About the Mean in a Normal Population 758
Asymptotic Tests for Nonnormal Populations 762
Computing and Using p- *Values 762*

The Relationship Between Confidence Intervals and Hypothesis Testing 766
Practical Versus Statistical Significance- 767
C.7 Remarks on Notation 768
Summary , 769
Key Terms 769
Problems 770

Appendix D Summary of Matrix Algebra **776**
D.I Basic Definitions 776
D.2 Matrix Operations 777
Matrix Addition' 777
Scalar Multiplication 778
Matrix Multiplication 778
Transpose 779
Partitioned Matrix Multiplication 780
Trace 780
Inverse 781
D.3 Linear Independence. Rank of a Matrix 781
D.4 Quadratic Forms and Positive Definite Matrices 782
D.5 Idempotent Matrices 782
D.6- Differentiation of Linear and Quadratic Forms 783
D.7 Moments and Distributions of Random Vectors 783
Expected Value 784
Variance-Covariance Matrix 784
Mullivariate Normal Distribution 784
Chi-Square Distribution 785
t *Distribution 785*
F *Distribution 785*
Summary 785
Key Terms 786
Problems 786

Appendix E The Linear Regression Model in Matrix Form **787**
E.I The Model and Ordinary Least Squares Estimation 787
E.2 Finite Sample Properties of OLS 790
E-.3 Statistical Inference 793
E.4 Some Asymptotic Analysis 795
Wald Statistics for Testing Multiple Hypotheses 798
Summary 799
Key Terms 799
Problems 799

Appendix F Answers to Chapter Questions 802

Appendix G Statistical Tables **815**