

W 913 (4)

# Introductory Econometrics

A Modern Approach

## 4e

Jeffrey M. Wooldridge  
Michigan State University

H HOCHSCHULE  
m LIECHTENSTEIN  
Bibliothek

SOUTH-WESTERN  
CENGAGE Learning-

Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

# iSContents

## 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

- Uribiasedness 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 "Partiallying 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 irr. Misspecified Models* 99  
*Estimating  $\sigma^2$ : 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  $\beta_j$*  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 Information by 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 Internals 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 Seasonally	360
<i>Characterizing Trending Time Series</i>	360
<i>Using Trending Variables in. Regression Analysis</i>	363
<i>A Detrending Interpretation of Regressions with a Time Trend</i>	365
<i>Computing R-Squared when the Dependent Variable Is Trending</i>	366
<i>Seasonality</i>	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
<i>Stationary and Nonstationary Time Series</i>	378
<i>Weakly Dependent Time Series</i>	379
11.2 Asymptotic Properties of OLS	381
11.3 Using Highly Persistent Time Series in Regression Analysis	388
<i>Highly Persistent Time Series</i>	388
<i>Transformations of Highly Persistent Time Series</i>	393
<i>Deciding Whether a Time Series Is <math>I(1)</math></i>	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
<i>Unbiasedness and Consistency</i>	408
<i>Efficiency and Inference</i>	409
<i>Goodness-of-Fit</i>	410

<i>Serial Correlation in the Presence of Lagged Dependent Variables</i>	411
12.2 Testing for Serial Correlation	412
<i>A <math>t</math> Test for <math>AR(1)</math> Serial Correlation with Strictly Exogenous Regressors</i>	412
<i>The Durbin-Watson Test under Classical Assumptions</i>	415
<i>Testing for <math>AR(1)</math> Serial Correlation without Strictly Exogenous Regressors</i>	416
<i>Testing for Higher Order Serial Correlation</i>	417
12.3 Correcting for Serial Correlation with Strictly Exogenous Regressors	419
<i>Obtaining the Best Linear Unbiased Estimator in the <math>AR(1)</math> Model</i>	419
<i>Feasible GLS Estimation with <math>AR(1)</math> Errors</i>	421
<i>Comparing OLS and FGLS</i>	423
<i>Correcting for Higher Order Serial Correlation</i>	425
12.4 Differencing and Serial Correlation	426
<sup>1</sup> 12.5 Serial Correlation-Robust Inference after OLS	428
12.6 Heteroskedasticity in Time Series Regressions	432
<i>Heteroskedasticity-Robust Statistics</i>	432
<i>Testing for Heteroskedasticity</i>	432
<i>Autoregressive Conditional Heteroskedasticity</i>	433
<i>Heteroskedasticity and Serial Correlation in Regression Models</i>	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
<i>The Chow Test for Structural Change across Time</i>	449
13.2 Policy Analysis with Pooled Cross Sections	450
13.3 Two-Period Panel Data Analysis	455
<i>Organizing Panel Data</i>	461

13.4 Policy Analysis with Two-Period Panel Data	462	15.4 IV Solutions to Errors-in-Variables Problems	525
13.5 Differencing with More Than Two Time Periods	465	15.5 Testing for Eridogeneity and Testing Overidentifying Restrictions	527
<i>Potential Pitfalls in First Differencing Panel Data</i>	470	<i>Testing for Endogeneity</i>	527
Summary	471	<i>Testing Overidentification Restrictions</i>	529
Key Terms	471	15.6 2SLS with Heteroskedasticity	531
Problems	471	15.7 Applying 2SLS to Time Series Equations	531
Computer Exercises	473	15.8 Applying 2SLS to Pooled Cross Sections and Panel Data	534
Appendix 13A	478	Summary	536
<b>CHAPTER 14</b>		Key Terms	536
<b>Advanced Panel Data Methods</b>	<b>481</b>	Problems	536
14.1 Fixed Effects Estimation	481	Computer Exercises	539
<i>The Dummy Variable Regression</i>	485	Appendix 15A	543
<i>Fixed Effects or First Differencing?</i>	487	<b>CHAPTER 16</b>	
<i>Fixed Effects with Unbalanced Panels</i>	488	<b>Simultaneous Equations Models</b>	<b>546</b>
14.2 Random Effects Models	489	16.1 The Nature of Simultaneous Equations Models	546
<i>Random Effects or Fixed Effects?</i>	493	16.2 Simultaneity Bias in OLS	550
14.3 Applying Panel Data Methods to Other Data Structures	494	16.3 Identifying and Estimating a Structural Equation	552
Summary	496	<i>Identification in a Two-Equation System</i>	552
Key Terms	496	<i>Estimation by 2SLS</i>	557
Problems	497	16.4 Systems with More Than Two Equations	559
Computer Exercises	498	<i>j Identification in Systems with Three or More Equations</i>	559
Appendix 14A	503	<i>Estimation</i>	560
<b>CHAPTER 15</b>		16.5 Simultaneous Equations Models with Time Series	560
<b>Instrumental Variables Estimation and Two Stage Least Squares</b>	<b>506</b>	16.6 Simultaneous Equations Models with Panel Data	564
15.1 Motivation: Omitted Variables in a Simple Regression Model	507	Summary	566
<i>Statistical Inference with the IV Estimator</i>	510	Key Terms	567
<i>Properties of IV with a Poor Instrumental Variable</i>	514	Problems	567
<i>Computing R-Squared after IV Estimation</i>	516	Computer Exercises	570
15.2 IV Estimation of the Multiple Regression Model	517	<b>CHAPTER 17</b>	
15.3 Two Stage Least Squares	521	<b>Limited Dependent Variable Models and Sample Selection Corrections</b>	<b>574</b>
<i>A Single Endogenous Explanatory Variable</i>	521	17.1 Logit and Probit Models for Binary Response	575
<i>Multicollinearity and 2SLS</i>	523	<i>Specifying Logit and Probit Models</i>	575
<i>Multiple Endogenous Explanatory Variables</i>	524	<i>Maximum Likelihood Estimation of Logit and Probit Models</i>	578
<i>Testing Multiple Hypotheses after 2SLS Estimation</i>	525	<i>Testing Multiple Hypotheses</i>	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

## Contents

<i>A Measure of Central Tendency: The Expected Value</i>	722
<i>Properties of Expected Values</i>	724
<i>Another Measure of Central Tendency: The Median</i>	725
<i>Measures of Variability: Variance and Standard Deviation</i>	726
<i>Variance</i>	726
<i>Standard Deviation</i>	728
<i>Standardizing a Random Variable</i>	728
<i>Skewness and Kurtosis</i>	729
B.4 Features of Joint and Conditional Distributions	729
<i>Measures of Association: Covariance and Correlation</i>	729
<i>Covariance</i>	729
<i>Correlation Coefficient</i>	731
<i>Variance of Sums of Random Variables</i>	732
<i>Conditional Expectation</i>	733
<i>Properties of Conditional Expectation</i>	734
<i>Conditional Variance</i>	736
B.5 The Normal and Related Distributions	737
<i>The Normal Distribution</i>	737
<i>The Standard Normal Distribution</i>	738
<i>Additional Properties of the Normal Distribution</i>	740
<i>The Chi-Square Distribution</i>	741
<i>The t Distribution</i>	741
<i>The F Distribution</i>	743
Summary	744
Key Terms	744
Problems	745
<b>APPENDIX C</b>	
<b>Fundamentals of Mathematical Statistics 747</b>	
C.1 Populations, Parameters, and Random Sampling	747
<i>Sampling</i>	748
C.2 Finite Sample Properties of Estimators	748
<i>Estimators and Estimates</i>	749
<i>Unbiasedness</i>	750
<i>The Sampling Variance of Estimators</i>	752
<i>Efficiency</i>	754
C.3 Asymptotic or Larger Sample Properties of Estimators	755
<i>Consistency</i>	755
<i>Asymptotic Normality</i>	758
C.4 General Approaches to Parameter Estimation	760
<i>Method of Moments</i>	760
<i>Maximum Likelihood</i>	761
<i>Least Squares</i>	762
C.5 Interval Estimation and Confidence Intervals	762
<i>The Nature of Interval Estimation</i>	762
<i>Confidence Intervals for the Mean from a Normally Distributed Population</i>	764
<i>A Simple Rule of Thumb for a 95% Confidence Interval</i>	768
<i>Asymptotic Confidence Intervals for Nonnormal Populations</i>	768
C.6 Hypothesis Testing	770
<i>Fundamentals of Hypothesis Testing</i>	770
<i>Testing Hypotheses about the Mean in a Normal Population</i>	772
<i>Asymptotic Tests for Nonnormal Populations</i>	774
<i>Computing and Using p-Values</i>	776
<i>The Relationship between Confidence Intervals and Hypothesis Testing</i>	779
<i>Practical versus Statistical Significance</i>	780
C.1 Remarks on Notation	781
Summary	782
Key Terms	782
Problems	783
<b>APPENDIX D</b>	
<b>Summary of Matrix Algebra 788</b>	
D.1 Basic Definitions	788
D.2 Matrix Operations	789
<i>Matrix Addition</i>	789
<i>Scalar Multiplication</i>	790
<i>Matrix Multiplication</i>	790
<i>Transpose</i>	791
<i>Partitioned Matrix Multiplication</i>	792
<i>Trace</i>	792
<i>Inverse</i>	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
<i>Expected Value</i>	795
<i>Variance-Covariance Matrix</i>	795
<i>Multivariate Normal Distribution</i>	796



*Ch'i-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