# Brief Contents

## PART ONE  Introduction and Review

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Economic Questions and Data</td>
<td>47</td>
</tr>
<tr>
<td>2</td>
<td>Review of Probability</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>Review of Statistics</td>
<td>111</td>
</tr>
</tbody>
</table>

## PART TWO  Fundamentals of Regression Analysis

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Linear Regression with One Regressor</td>
<td>155</td>
</tr>
<tr>
<td>5</td>
<td>Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals</td>
<td>192</td>
</tr>
<tr>
<td>6</td>
<td>Linear Regression with Multiple Regressors</td>
<td>228</td>
</tr>
<tr>
<td>7</td>
<td>Hypothesis Tests and Confidence Intervals in Multiple Regression</td>
<td>263</td>
</tr>
<tr>
<td>8</td>
<td>Nonlinear Regression Functions</td>
<td>302</td>
</tr>
<tr>
<td>9</td>
<td>Assessing Studies Based on Multiple Regression</td>
<td>361</td>
</tr>
</tbody>
</table>

## PART THREE  Further Topics in Regression Analysis

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Regression with Panel Data</td>
<td>396</td>
</tr>
<tr>
<td>11</td>
<td>Regression with a Binary Dependent Variable</td>
<td>431</td>
</tr>
<tr>
<td>12</td>
<td>Instrumental Variables Regression</td>
<td>470</td>
</tr>
<tr>
<td>13</td>
<td>Experiments and Quasi-Experiments</td>
<td>521</td>
</tr>
</tbody>
</table>

## PART FOUR  Regression Analysis of Economic Time Series Data

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Introduction to Time Series Regression and Forecasting</td>
<td>568</td>
</tr>
<tr>
<td>15</td>
<td>Estimation of Dynamic Causal Effects</td>
<td>635</td>
</tr>
<tr>
<td>16</td>
<td>Additional Topics in Time Series Regression</td>
<td>684</td>
</tr>
</tbody>
</table>

## PART FIVE  The Econometric Theory of Regression Analysis

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>The Theory of Linear Regression with One Regressor</td>
<td>722</td>
</tr>
<tr>
<td>18</td>
<td>The Theory of Multiple Regression</td>
<td>751</td>
</tr>
</tbody>
</table>
# Contents

**Preface** 31

## PART ONE  Introduction and Review

### CHAPTER 1  Economic Questions and Data  47

1.1  Economic Questions We Examine  47

- Question #1: Does Reducing Class Size Improve Elementary School Education?  48
- Question #2: Is There Racial Discrimination in the Market for Home Loans?  49
- Question #3: How Much Do Cigarette Taxes Reduce Smoking?  49
- Question #4: By How Much Will U.S. GDP Grow Next Year?  50

1.2  Causal Effects and Idealized Experiments  51

- Estimation of Causal Effects  52
- Forecasting and Causality  53

1.3  Data: Sources and Types  53

- Experimental Versus Observational Data  53
- Cross-Sectional Data  54
- Time Series Data  55
- Panel Data  57

### CHAPTER 2  Review of Probability  60

2.1  Random Variables and Probability Distributions  61

- Probabilities, the Sample Space, and Random Variables  61
- Probability Distribution of a Discrete Random Variable  62
- Probability Distribution of a Continuous Random Variable  65

2.2  Expected Values, Mean, and Variance  65

- The Expected Value of a Random Variable  65
- The Standard Deviation and Variance  67
- Mean and Variance of a Linear Function of a Random Variable  68
- Other Measures of the Shape of a Distribution  69

2.3  Two Random Variables  72

- Joint and Marginal Distributions  72
Conditional Distributions  73
Independence  77
Covariance and Correlation  77
The Mean and Variance of Sums of Random Variables  78

2.4  The Normal, Chi-Squared, Student t, and F Distributions  82
The Normal Distribution  82
The Chi-Squared Distribution  87
The Student t Distribution  87
The F Distribution  88

2.5  Random Sampling and the Distribution of the Sample Average  89
Random Sampling  89
The Sampling Distribution of the Sample Average  90

2.6  Large-Sample Approximations to Sampling Distributions  93
The Law of Large Numbers and Consistency  94
The Central Limit Theorem  96

APPENDIX 2.1 Derivation of Results in Key Concept 2.3  109

CHAPTER 3  Review of Statistics  111

3.1  Estimation of the Population Mean  112
Estimators and Their Properties  112
Properties of  114
The Importance of Random Sampling  116

3.2  Hypothesis Tests Concerning the Population Mean  117
Null and Alternative Hypotheses  117
The p-Value  118
Calculating the p-Value When  Is Known  119
The Sample Variance, Sample Standard Deviation, and Standard Error  120
Calculating the p-Value When  Is Unknown  122
The t-Statistic  122
Hypothesis Testing with a Prespecified Significance Level  123
One-Sided Alternatives  125

3.3  Confidence Intervals for the Population Mean  126

3.4  Comparing Means from Different Populations  128
Hypothesis Tests for the Difference Between Two Means  128
Confidence Intervals for the Difference Between Two Population Means  130
3.5 Differences-of-Means Estimation of Causal Effects Using Experimental Data 130
   - The Causal Effect as a Difference of Conditional Expectations 131
   - Estimation of the Causal Effect Using Differences of Means 131

3.6 Using the t-Statistic When the Sample Size Is Small 133
   - The t-Statistic and the Student t Distribution 133
   - Use of the Student t Distribution in Practice 135

3.7 Scatterplots, the Sample Covariance, and the Sample Correlation 137
   - Scatterplots 137
   - Sample Covariance and Correlation 138

APPENDIX 3.1 The U.S. Current Population Survey 152
APPENDIX 3.2 Two Proofs That Y Is the Least Squares Estimator of μY 153
APPENDIX 3.3 A Proof That the Sample Variance Is Consistent 154

PART TWO  Fundamentals of Regression Analysis

CHAPTER 4  Linear Regression with One Regressor 155

4.1 The Linear Regression Model 155

4.2 Estimating the Coefficients of the Linear Regression Model 160
   - The Ordinary Least Squares Estimator 162
   - OLS Estimates of the Relationship Between Test Scores and the Student–Teacher Ratio 164
   - Why Use the OLS Estimator? 165

4.3 Measures of Fit 167
   - The R² 167
   - The Standard Error of the Regression 168
   - Application to the Test Score Data 169

4.4 The Least Squares Assumptions 170
   - Assumption #1: The Conditional Distribution of ui Given Xi Has a Mean of Zero 170
   - Assumption #2: (X_i, Y_i), i = 1, ..., n, Are Independently and Identically Distributed 172
   - Assumption #3: Large Outliers Are Unlikely 173
   - Use of the Least Squares Assumptions 174
CHAPTER 6  Linear Regression with Multiple Regressors  228

6.1  Omitted Variable Bias  228
   Definition of Omitted Variable Bias  229
   A Formula for Omitted Variable Bias  231
   Addressing Omitted Variable Bias by Dividing the Data into Groups  233

6.2  The Multiple Regression Model  235
   The Population Regression Line  235
   The Population Multiple Regression Model  236

6.3  The OLS Estimator in Multiple Regression  238
   The OLS Estimator  239
   Application to Test Scores and the Student–Teacher Ratio  240

6.4  Measures of Fit in Multiple Regression  242
   The Standard Error of the Regression (SER)  242
   The $R^2$  242
   The “Adjusted $R^2_*$”  243
   Application to Test Scores  244

6.5  The Least Squares Assumptions in Multiple Regression  245
   Assumption #1: The Conditional Distribution of $u_i$ Given $X_{1i}, X_{2i}, \ldots, X_{ki}$ Has a Mean of Zero  245
   Assumption #2: $(X_{1i}, X_{2i}, \ldots, X_{ki}, Y_i), i = 1, \ldots, n$, Are i.i.d.  245
   Assumption #3: Large Outliers Are Unlikely  245
   Assumption #4: No Perfect Multicollinearity  246

6.6  The Distribution of the OLS Estimators in Multiple Regression  247

6.7  Multicollinearity  248
   Examples of Perfect Multicollinearity  249
   Imperfect Multicollinearity  251

6.8  Conclusion  252

APPENDIX 6.1 Derivation of Equation (6.1)  260
APPENDIX 6.2 Distribution of the OLS Estimators When There Are Two Regressors and Homoskedastic Errors  260
APPENDIX 6.3 The Frisch–Waugh Theorem  261
CHAPTER 7  Hypothesis Tests and Confidence Intervals in Multiple Regression  263

7.1  Hypothesis Tests and Confidence Intervals for a Single Coefficient  263
  Standard Errors for the OLS Estimators  263
  Hypothesis Tests for a Single Coefficient  264
  Confidence Intervals for a Single Coefficient  265
  Application to Test Scores and the Student–Teacher Ratio  266

7.2  Tests of Joint Hypotheses  268
  Testing Hypotheses on Two or More Coefficients  268
  The F-Statistic  270
  Application to Test Scores and the Student–Teacher Ratio  272
  The Homoskedasticity-Only F-Statistic  273

7.3  Testing Single Restrictions Involving Multiple Coefficients  275

7.4  Confidence Sets for Multiple Coefficients  277

7.5  Model Specification for Multiple Regression  278
  Omitted Variable Bias in Multiple Regression  279
  The Role of Control Variables in Multiple Regression  280
  Model Specification in Theory and in Practice  282
  Interpreting the $R^2$ and the Adjusted $R^2$ in Practice  283

7.6  Analysis of the Test Score Data Set  284

7.7  Conclusion  289

APPENDIX 7.1 The Bonferroni Test of a Joint Hypothesis  297
APPENDIX 7.2 Conditional Mean Independence  299

CHAPTER 8  Nonlinear Regression Functions  302

8.1  A General Strategy for Modeling Nonlinear Regression Functions  304
  Test Scores and District Income  304
  The Effect on $Y$ of a Change in $X$ in Nonlinear Specifications  307
  A General Approach to Modeling Nonlinearities Using Multiple Regression  312

8.2  Nonlinear Functions of a Single Independent Variable  312
  Polynomials  313
  Logarithms  315
  Polynomial and Logarithmic Models of Test Scores and District Income  323
## CHAPTER 8

### 8.3 Interactions Between Independent Variables
- Interactions Between Two Binary Variables 324
- Interactions Between a Continuous and a Binary Variable 328
- Interactions Between Two Continuous Variables 332

### 8.4 Nonlinear Effects on Test Scores of the Student–Teacher Ratio
- Discussion of Regression Results 339
- Summary of Findings 343

### 8.5 Conclusion

APPENDIX 8.1 Regression Functions That Are Nonlinear in the Parameters 355

APPENDIX 8.2 Slopes and Elasticities for Nonlinear Regression Functions 359

---

## CHAPTER 9

### Assessing Studies Based on Multiple Regression

### 9.1 Internal and External Validity
- Threats to Internal Validity 362
- Threats to External Validity 363

### 9.2 Threats to Internal Validity of Multiple Regression Analysis
- Omitted Variable Bias 365
- Misspecification of the Functional Form of the Regression Function 367
- Measurement Error and Errors-in-Variables Bias 368
- Missing Data and Sample Selection 371
- Simultaneous Causality 372
- Sources of Inconsistency of OLS Standard Errors 375

### 9.3 Internal and External Validity When the Regression Is Used for Forecasting
- Using Regression Models for Forecasting 377
- Assessing the Validity of Regression Models for Forecasting 378

### 9.4 Example: Test Scores and Class Size
- External Validity 378
- Internal Validity 385
- Discussion and Implications 387

### 9.5 Conclusion

APPENDIX 9.1 The Massachusetts Elementary School Testing Data 395
## PART THREE  Further Topics in Regression Analysis

### CHAPTER 10  Regression with Panel Data  396

10.1 Panel Data  397
   Example: Traffic Deaths and Alcohol Taxes  398

10.2 Panel Data with Two Time Periods: “Before and After” Comparisons  400

10.3 Fixed Effects Regression  403
   The Fixed Effects Regression Model  403
   Estimation and Inference  405
   Application to Traffic Deaths  407

10.4 Regression with Time Fixed Effects  407
   Time Effects Only  408
   Both Entity and Time Fixed Effects  409

10.5 The Fixed Effects Regression Assumptions and Standard Errors for Fixed Effects Regression  411
   The Fixed Effects Regression Assumptions  411
   Standard Errors for Fixed Effects Regression  413

10.6 Drunk Driving Laws and Traffic Deaths  414

10.7 Conclusion  418
   APPENDIX 10.1 The State Traffic Fatality Data Set  426
   APPENDIX 10.2 Standard Errors for Fixed Effects Regression  426

### CHAPTER 11  Regression with a Binary Dependent Variable  431

11.1 Binary Dependent Variables and the Linear Probability Model  432
   Binary Dependent Variables  432
   The Linear Probability Model  434

11.2 Probit and Logit Regression  437
   Probit Regression  437
   Logit Regression  442
   Comparing the Linear Probability, Probit, and Logit Models  444

11.3 Estimation and Inference in the Logit and Probit Models  444
   Nonlinear Least Squares Estimation  445
11.4 Application to the Boston HMDA Data 448
11.5 Conclusion 455

APPENDIX 11.1 The Boston HMDA Data Set 464
APPENDIX 11.2 Maximum Likelihood Estimation 464
APPENDIX 11.3 Other Limited Dependent Variable Models 467

CHAPTER 12 Instrumental Variables Regression 470

12.1 The IV Estimator with a Single Regressor and a Single Instrument 471
The IV Model and Assumptions 471
The Two Stage Least Squares Estimator 472
Why Does IV Regression Work? 473
The Sampling Distribution of the TSLS Estimator 477
Application to the Demand for Cigarettes 479

12.2 The General IV Regression Model 481
TSLS in the General IV Model 483
Instrument Relevance and Exogeneity in the General IV Model 484
The IV Regression Assumptions and Sampling Distribution of the TSLS Estimator 485
Inference Using the TSLS Estimator 486
Application to the Demand for Cigarettes 487

12.3 Checking Instrument Validity 488
Assumption #1: Instrument Relevance 489
Assumption #2: Instrument Exogeneity 491

12.4 Application to the Demand for Cigarettes 494

12.5 Where Do Valid Instruments Come From? 499
Three Examples 500

12.6 Conclusion 504

APPENDIX 12.1 The Cigarette Consumption Panel Data Set 513
APPENDIX 12.2 Derivation of the Formula for the TSLS Estimator in Equation (12.4) 513
APPENDIX 12.3 Large-Sample Distribution of the TSLS Estimator  514
APPENDIX 12.4 Large-Sample Distribution of the TSLS Estimator When the Instrument Is Not Valid  515
APPENDIX 12.5 Instrumental Variables Analysis with Weak Instruments  517
APPENDIX 12.6 TSLS with Control Variables  519

CHAPTER 13 Experiments and Quasi-Experiments  521

13.1 Potential Outcomes, Causal Effects, and Idealized Experiments  522
  Potential Outcomes and the Average Causal Effect  522
  Econometric Methods for Analyzing Experimental Data  524

13.2 Threats to Validity of Experiments  525
  Threats to Internal Validity  525
  Threats to External Validity  529

13.3 Experimental Estimates of the Effect of Class Size Reductions  530
  Experimental Design  531
  Analysis of the STAR Data  532
  Comparison of the Observational and Experimental Estimates of Class Size Effects  537

13.4 Quasi-Experiments  539
  Examples  540
  The Differences-in-Differences Estimator  542
  Instrumental Variables Estimators  545
  Regression Discontinuity Estimators  546

13.5 Potential Problems with Quasi-Experiments  548
  Threats to Internal Validity  548
  Threats to External Validity  550

13.6 Experimental and Quasi-Experimental Estimates in Heterogeneous Populations  550
  OLS with Heterogeneous Causal Effects  551
  IV Regression with Heterogeneous Causal Effects  552
13.7 Conclusion 555

APPENDIX 13.1 The Project STAR Data Set 564
APPENDIX 13.2 IV Estimation When the Causal Effect Varies Across Individuals 564
APPENDIX 13.3 The Potential Outcomes Framework for Analyzing Data from Experiments 566

PART FOUR Regression Analysis of Economic Time Series Data

CHAPTER 14 Introduction to Time Series Regression and Forecasting 568

14.1 Using Regression Models for Forecasting 569

14.2 Introduction to Time Series Data and Serial Correlation 570
   Real GDP in the United States 570
   Lags, First Differences, Logarithms, and Growth Rates 571
   Autocorrelation 574
   Other Examples of Economic Time Series 575

14.3 Autoregressions 577
   The First-Order Autoregressive Model 577
   The $p^{th}$-Order Autoregressive Model 580

14.4 Time Series Regression with Additional Predictors and the Autoregressive Distributed Lag Model 583
   Forecasting GDP Growth Using the Term Spread 583
   Stationarity 586
   Time Series Regression with Multiple Predictors 587
   Forecast Uncertainty and Forecast Intervals 590

14.5 Lag Length Selection Using Information Criteria 593
   Determining the Order of an Autoregression 593
   Lag Length Selection in Time Series Regression with Multiple Predictors 596

14.6 Nonstationarity I: Trends 597
   What Is a Trend? 597
   Problems Caused by Stochastic Trends 600
   Detecting Stochastic Trends: Testing for a Unit AR Root 602
   Avoiding the Problems Caused by Stochastic Trends 607
15.7 Is Exogeneity Plausible? Some Examples 670
   U.S. Income and Australian Exports 670
   Oil Prices and Inflation 671
   Monetary Policy and Inflation 672
   The Growth Rate of GDP and the Term Spread 672

15.8 Conclusion 673

**APPENDIX 15.1** The Orange Juice Data Set 680

**APPENDIX 15.2** The ADL Model and Generalized Least Squares in Lag Operator Notation 680

**CHAPTER 16 Additional Topics in Time Series Regression 684**

16.1 Vector Autoregressions 684
   The VAR Model 685
   A VAR Model of the Growth Rate of GDP and the Term Spread 688

16.2 Multiperiod Forecasts 689
   Iterated Multiperiod Forecasts 689
   Direct Multiperiod Forecasts 691
   Which Method Should You Use? 694

16.3 Orders of Integration and the DF-GLS Unit Root Test 695
   Other Models of Trends and Orders of Integration 695
   The DF-GLS Test for a Unit Root 697
   Why Do Unit Root Tests Have Nonnormal Distributions? 700

16.4 Cointegration 702
   Cointegration and Error Correction 702
   How Can You Tell Whether Two Variables Are Cointegrated? 704
   Estimation of Cointegrating Coefficients 705
   Extension to Multiple Cointegrated Variables 707
   Application to Interest Rates 708

16.5 Volatility Clustering and Autoregressive Conditional Heteroskedasticity 710
   Volatility Clustering 410
   Autoregressive Conditional Heteroskedasticity 712
   Application to Stock Price Volatility 713

16.6 Conclusion 716
PART FIVE  The Econometric Theory of Regression Analysis

CHAPTER 17  The Theory of Linear Regression with One Regressor  722

17.1  The Extended Least Squares Assumptions and the OLS Estimator  723
   The Extended Least Squares Assumptions  723
   The OLS Estimator  725

17.2  Fundamentals of Asymptotic Distribution Theory  725
   Convergence in Probability and the Law of Large Numbers  726
   The Central Limit Theorem and Convergence in Distribution  728
   Slutsky’s Theorem and the Continuous Mapping Theorem  729
   Application to the t-Statistic Based on the Sample Mean  730

17.3  Asymptotic Distribution of the OLS Estimator and t-Statistic  731
   Consistency and Asymptotic Normality of the OLS Estimators  731
   Consistency of Heteroskedasticity-Robust Standard Errors  731
   Asymptotic Normality of the Heteroskedasticity-Robust t-Statistic  733

17.4  Exact Sampling Distributions When the Errors Are Normally Distributed  733
   Distribution of  with Normal Errors  733
   Distribution of the Homoskedasticity-Only t-Statistic  735

17.5  Weighted Least Squares  736
   WLS with Known Heteroskedasticity  736
   WLS with Heteroskedasticity of Known Functional Form  737
   Heteroskedasticity-Robust Standard Errors or WLS?  740
   APPENDIX 17.1 The Normal and Related Distributions and Moments of Continuous Random Variables  746
   APPENDIX 17.2 Two Inequalities  749

CHAPTER 18  The Theory of Multiple Regression  751

18.1  The Linear Multiple Regression Model and OLS Estimator in Matrix Form  752
   The Multiple Regression Model in Matrix Notation  752
   The Extended Least Squares Assumptions  754
   The OLS Estimator  755
18.2 Asymptotic Distribution of the OLS Estimator and t-Statistic 756
   The Multivariate Central Limit Theorem 756
   Asymptotic Normality of \( \hat{\beta} \) 757
   Heteroskedasticity-Robust Standard Errors 758
   Confidence Intervals for Predicted Effects 759
   Asymptotic Distribution of the t-Statistic 759

18.3 Tests of Joint Hypotheses 759
   Joint Hypotheses in Matrix Notation 760
   Asymptotic Distribution of the F-Statistic 760
   Confidence Sets for Multiple Coefficients 761

18.4 Distribution of Regression Statistics with Normal Errors 762
   Matrix Representations of OLS Regression Statistics 762
   Distribution of \( \hat{\beta} \) with Normal Errors 763
   Distribution of \( s^2 \) 764
   Homoskedasticity-Only Standard Errors 764
   Distribution of the t-Statistic 765
   Distribution of the F-Statistic 765

18.5 Efficiency of the OLS Estimator with Homoskedastic Errors 766
   The Gauss–Markov Conditions for Multiple Regression 766
   Linear Conditionally Unbiased Estimators 766
   The Gauss–Markov Theorem for Multiple Regression 767

18.6 Generalized Least Squares 768
   The GLS Assumptions 769
   GLS When \( \Omega \) Is Known 771
   GLS When \( \Omega \) Contains Unknown Parameters 772
   The Zero Conditional Mean Assumption and GLS 772

18.7 Instrumental Variables and Generalized Method of Moments Estimation 774
   The IV Estimator in Matrix Form 775
   Asymptotic Distribution of the TSLS Estimator 776
   Properties of TSLS When the Errors Are Homoskedastic 777
   Generalized Method of Moments Estimation in Linear Models 780

APPENDIX 18.1 Summary of Matrix Algebra 792
APPENDIX 18.2 Multivariate Distributions 795
APPENDIX 18.3 Derivation of the Asymptotic Distribution of \( \hat{\beta} \) 797
APPENDIX 18.4 Derivations of Exact Distributions of OLS Test Statistics with Normal Errors  798
APPENDIX 18.5 Proof of the Gauss–Markov Theorem for Multiple Regression  799
APPENDIX 18.6 Proof of Selected Results for IV and GMM Estimation  800

Appendix  803
References  811
Glossary  817
Index  825
Key Concepts

PART ONE  Introduction and Review

1.1  Cross-Sectional, Time Series, and Panel Data  58
2.1  Expected Value and the Mean  66
2.2  Variance and Standard Deviation  67
2.3  Means, Variances, and Covariances of Sums of Random Variables  81
2.4  Computing Probabilities Involving Normal Random Variables  83
2.5  Simple Random Sampling and i.i.d. Random Variables  90
2.6  Convergence in Probability, Consistency, and the Law of Large Numbers  94
2.7  The Central Limit Theorem  98
3.1  Estimators and Estimates  113
3.2  Bias, Consistency, and Efficiency  114
3.3  Efficiency of $\hat{Y}$: $\hat{Y}$ Is BLUE  115
3.4  The Standard Error of $\hat{Y}$  121
3.5  The Terminology of Hypothesis Testing  124
3.6  Testing the Hypothesis $E(Y) = \mu_{Y,0}$ Against the Alternative $E(Y) \neq \mu_{Y,0}$  125
3.7  Confidence Intervals for the Population Mean  127

PART TWO  Fundamentals of Regression Analysis

4.1  Terminology for the Linear Regression Model with a Single Regressor  159
4.2  The OLS Estimator, Predicted Values, and Residuals  163
4.3  The Least Squares Assumptions  175
4.4  Large-Sample Distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$  177
5.1  General Form of the t-Statistic  193
5.2  Testing the Hypothesis $\beta_1 = \beta_{1,0}$ Against the Alternative $\beta_1 \neq \beta_{1,0}$  195
5.3  Confidence Interval for $\beta_1$  200
5.4  Heteroskedasticity and Homoskedasticity  205
5.5  The Gauss–Markov Theorem for $\hat{\beta}_1$  211
6.1  Omitted Variable Bias in Regression with a Single Regressor  231
6.2  The Multiple Regression Model  238
6.3  The OLS Estimators, Predicted Values, and Residuals in the Multiple Regression Model  240
6.4  The Least Squares Assumptions in the Multiple Regression Model  247
6.5  Large-Sample Distribution of $\hat{\beta}_0$, $\hat{\beta}_1$, \ldots, $\hat{\beta}_k$  248
7.1  Testing the Hypothesis $\beta_j = \beta_{j,0}$ Against the Alternative $\beta_j \neq \beta_{j,0}$  265
7.2  Confidence Intervals for a Single Coefficient in Multiple Regression  266

25
Key Concepts

7.3 Omitted Variable Bias in Multiple Regression 279
7.4 $R^2$ and $R^2$: What They Tell You—and What They Don’t 284
8.1 The Expected Change on $Y$ of a Change in $X_1$ in the Nonlinear Regression Model (8.3) 309
8.2 Logarithms in Regression: Three Cases 322
8.3 A Method for Interpreting Coefficients in Regressions with Binary Variables 327
8.4 Interactions Between Binary and Continuous Variables 330
8.5 Interactions in Multiple Regression 335
9.1 Internal and External Validity 362
9.2 Omitted Variable Bias: Should I Include More Variables in My Regression? 367
9.3 Functional Form Misspecification 368
9.4 Errors-in-Variables Bias 370
9.5 Sample Selection Bias 372
9.6 Simultaneous Causality Bias 375
9.7 Threats to the Internal Validity of a Multiple Regression Study 376

PART THREE Further Topics in Regression Analysis

10.1 Notation for Panel Data 397
10.2 The Fixed Effects Regression Model 405
10.3 The Fixed Effects Regression Assumptions 412
11.1 The Linear Probability Model 435
11.2 The Probit Model, Predicted Probabilities, and Estimated Effects 440
11.3 Logit Regression 442
12.1 The General Instrumental Variables Regression Model and Terminology 482
12.2 Two Stage Least Squares 484
12.3 The Two Conditions for Valid Instruments 485
12.4 The IV Regression Assumptions 486
12.5 A Rule of Thumb for Checking for Weak Instruments 490
12.6 The Overidentifying Restrictions Test (The $J$-Statistic) 494

PART FOUR Regression Analysis of Economic Time Series Data

14.1 Lags, First Differences, Logarithms, and Growth Rates 573
14.2 Autocorrelation (Serial Correlation) and Autocovariance 574
14.3 Autoregressions 581
14.4 The Autoregressive Distributed Lag Model 586
14.5  Stationarity 587
14.6  Time Series Regression with Multiple Predictors 588
14.7  Granger Causality Tests (Tests of Predictive Content) 589
14.8  The Augmented Dickey–Fuller Test for a Unit Autoregressive Root 605
14.9  The QLR Test for Coefficient Stability 612
14.10  Pseudo Out-of-Sample Forecasts 614
15.1  The Distributed Lag Model and Exogeneity 644
15.2  The Distributed Lag Model Assumptions 645
15.3  HAC Standard Errors 653
15.4  Estimation of Dynamic Multipliers Under Strict Exogeneity 662
16.1  Vector Autoregressions 685
16.2  Iterated Multiperiod Forecasts 692
16.3  Direct Multiperiod Forecasts 694
16.4  Orders of Integration, Differencing, and Stationarity 696
16.5  Cointegration 703

PART FIVE  Regression Analysis of Economic Time Series Data

17.1  The Extended Least Squares Assumptions for Regression with a Single Regressor 724
18.1  The Extended Least Squares Assumptions in the Multiple Regression Model 753
18.2  The Multivariate Central Limit Theorem 757
18.3  Gauss–Markov Theorem for Multiple Regression 768
18.4  The GLS Assumptions 770
General Interest Boxes

The Distribution of Earnings in the United States in 2012  
A Bad Day on Wall Street  
Financial Diversification and Portfolios  
Off the Mark!  
The Gender Gap of Earnings of College Graduates in the United States  
A Way to Increase Voter Turnout  
The “Beta” of a Stock  
The Economic Value of a Year of Education: Homoskedasticity or Heteroskedasticity?  
The Mozart Effect: Omitted Variable Bias?  
The Return to Education and the Gender Gap  
The Demand for Economics Journals  
Do Stock Mutual Funds Outperform the Market?  
James Heckman and Daniel McFadden, Nobel Laureates  
Who Invented Instrumental Variables Regression?  
A Scary Regression  
The Externalities of Smoking  
The Hawthorne Effect  
What Is the Effect on Employment of the Minimum Wage?  
Can You Beat the Market? Part I  
The River of Blood  
Can You Beat the Market? Part II  
Orange Trees on the March  
NEWS FLASH: Commodity Traders Send Shivers through Disney World  
Nobel Laureates in Time Series Econometrics
Preface

Econometrics can be a fun course for both teacher and student. The real world of economics, business, and government is a complicated and messy place, full of competing ideas and questions that demand answers. Is it more effective to tackle drunk driving by passing tough laws or by increasing the tax on alcohol? Can you make money in the stock market by buying when prices are historically low, relative to earnings, or should you just sit tight, as the random walk theory of stock prices suggests? Can we improve elementary education by reducing class sizes, or should we simply have our children listen to Mozart for 10 minutes a day? Econometrics helps us sort out sound ideas from crazy ones and find quantitative answers to important quantitative questions. Econometrics opens a window on our complicated world that lets us see the relationships on which people, businesses, and governments base their decisions.

Introduction to Econometrics is designed for a first course in undergraduate econometrics. It is our experience that to make econometrics relevant in an introductory course, interesting applications must motivate the theory and the theory must match the applications. This simple principle represents a significant departure from the older generation of econometrics books, in which theoretical models and assumptions do not match the applications. It is no wonder that some students question the relevance of econometrics after they spend much of their time learning assumptions that they subsequently realize are unrealistic so that they must then learn “solutions” to “problems” that arise when the applications do not match the assumptions. We believe that it is far better to motivate the need for tools with a concrete application and then to provide a few simple assumptions that match the application. Because the theory is immediately relevant to the applications, this approach can make econometrics come alive.

New to the Third Edition

• Updated treatment of standard errors for panel data regression
• Discussion of when and why missing data can present a problem for regression analysis
• The use of regression discontinuity design as a method for analyzing quasi-experiments
• Updated discussion of weak instruments
• Discussion of the use and interpretation of control variables integrated into the core development of regression analysis
• Introduction of the “potential outcomes” framework for experimental data
• Additional general interest boxes
• Additional exercises, both pencil-and-paper and empirical

This third edition builds on the philosophy of the first and second editions that applications should drive the theory, not the other way around.

One substantial change in this edition concerns inference in regression with panel data (Chapter 10). In panel data, the data within an entity typically are correlated over time. For inference to be valid, standard errors must be computed using a method that is robust to this correlation. The chapter on panel data now uses one such method, clustered standard errors, from the outset. Clustered standard errors are the natural extension to panel data of the heteroskedasticity-robust standard errors introduced in the initial treatment of regression analysis in Part II. Recent research has shown that clustered standard errors have a number of desirable properties, which are now discussed in Chapter 10 and in a revised appendix to Chapter 10.

Another substantial set of changes concerns the treatment of experiments and quasi-experiments in Chapter 13. The discussion of differences-in-differences regression has been streamlined and draws directly on the multiple regression principles introduced in Part II. Chapter 13 now discusses regression discontinuity design, which is an intuitive and important framework for the analysis of quasi-experimental data. In addition, Chapter 13 now introduces the potential outcomes framework and relates this increasingly commonplace terminology to concepts that were introduced in Parts I and II.

This edition has a number of other significant changes. One is that it incorporates a precise but accessible treatment of control variables into the initial discussion of multiple regression. Chapter 7 now discusses conditions for control variables being successful in the sense that the coefficient on the variable of interest is unbiased even though the coefficients on the control variables generally are not. Other changes include a new discussion of missing data in Chapter 9, a new optional calculus-based appendix to Chapter 8 on slopes and elasticities of nonlinear regression functions, and an updated discussion in Chapter 12 of what to do if you have weak instruments. This edition also includes new general interest boxes, updated empirical examples, and additional exercises.
The Updated Third Edition

- The time series data used in Chapters 14–16 have been extended through the beginning of 2013 and now include the Great Recession.
- The empirical analysis in Chapter 14 now focuses on forecasting the growth rate of real GDP using the term spread, replacing the Phillips curve forecasts from earlier editions.
- Several new empirical exercises have been added to each chapter. Rather than include all of the empirical exercises in the text, we have moved many of them to the Companion Website, www.pearsonglobaleditions.com/Stock_Watson. This has two main advantages: first, we can offer more and more in-depth exercises, and second, we can add and update exercises between editions. We encourage you to browse the empirical exercises available on the Companion Website.

Features of This Book

*Introduction to Econometrics* differs from other textbooks in three main ways. First, we integrate real-world questions and data into the development of the theory, and we take seriously the substantive findings of the resulting empirical analysis. Second, our choice of topics reflects modern theory and practice. Third, we provide theory and assumptions that match the applications. Our aim is to teach students to become sophisticated consumers of econometrics and to do so at a level of mathematics appropriate for an introductory course.

Real-World Questions and Data

We organize each methodological topic around an important real-world question that demands a specific numerical answer. For example, we teach single-variable regression, multiple regression, and functional form analysis in the context of estimating the effect of school inputs on school outputs. (Do smaller elementary school class sizes produce higher test scores?) We teach panel data methods in the context of analyzing the effect of drunk driving laws on traffic fatalities. We use possible racial discrimination in the market for home loans as the empirical application for teaching regression with a binary dependent variable (logit and probit). We teach instrumental variable estimation in the context of estimating the demand elasticity for cigarettes. Although these examples involve economic reasoning, all
can be understood with only a single introductory course in economics, and many can be understood without any previous economics coursework. Thus the instructor can focus on teaching econometrics, not microeconomics or macroeconomics.

We treat all our empirical applications seriously and in a way that shows students how they can learn from data but at the same time be self-critical and aware of the limitations of empirical analyses. Through each application, we teach students to explore alternative specifications and thereby to assess whether their substantive findings are robust. The questions asked in the empirical applications are important, and we provide serious and, we think, credible answers. We encourage students and instructors to disagree, however, and invite them to reanalyze the data, which are provided on the textbook’s Companion Website (www.pearsonglobaleditions.com/Stock_Watson).

Contemporary Choice of Topics

Econometrics has come a long way since the 1980s. The topics we cover reflect the best of contemporary applied econometrics. One can only do so much in an introductory course, so we focus on procedures and tests that are commonly used in practice. For example:

- **Instrumental variables regression.** We present instrumental variables regression as a general method for handling correlation between the error term and a regressor, which can arise for many reasons, including omitted variables and simultaneous causality. The two assumptions for a valid instrument—exogeneity and relevance—are given equal billing. We follow that presentation with an extended discussion of where instruments come from and with tests of overidentifying restrictions and diagnostics for weak instruments, and we explain what to do if these diagnostics suggest problems.

- **Program evaluation.** An increasing number of econometric studies analyze either randomized controlled experiments or quasi-experiments, also known as natural experiments. We address these topics, often collectively referred to as program evaluation, in Chapter 13. We present this research strategy as an alternative approach to the problems of omitted variables, simultaneous causality, and selection, and we assess both the strengths and the weaknesses of studies using experimental or quasi-experimental data.

- **Forecasting.** The chapter on forecasting (Chapter 14) considers univariate (autoregressive) and multivariate forecasts using time series regression, not large simultaneous equation structural models. We focus on simple and reliable tools, such as autoregressions and model selection via an information
criterion, that work well in practice. This chapter also features a practically oriented treatment of stochastic trends (unit roots), unit root tests, tests for structural breaks (at known and unknown dates), and pseudo out-of-sample forecasting, all in the context of developing stable and reliable time series forecasting models.

- **Time series regression.** We make a clear distinction between two very different applications of time series regression: forecasting and estimation of dynamic causal effects. The chapter on causal inference using time series data (Chapter 15) pays careful attention to when different estimation methods, including generalized least squares, will or will not lead to valid causal inferences and when it is advisable to estimate dynamic regressions using OLS with heteroskedasticity- and autocorrelation-consistent standard errors.

### Theory That Matches Applications

Although econometric tools are best motivated by empirical applications, students need to learn enough econometric theory to understand the strengths and limitations of those tools. We provide a modern treatment in which the fit between theory and applications is as tight as possible, while keeping the mathematics at a level that requires only algebra.

Modern empirical applications share some common characteristics: The data sets typically are large (hundreds of observations, often more); regressors are not fixed over repeated samples but rather are collected by random sampling (or some other mechanism that makes them random); the data are not normally distributed; and there is no a priori reason to think that the errors are homoskedastic (although often there are reasons to think that they are heteroskedastic).

These observations lead to important differences between the theoretical development in this textbook and other textbooks:

- **Large-sample approach.** Because data sets are large, from the outset we use large-sample normal approximations to sampling distributions for hypothesis testing and confidence intervals. In our experience, it takes less time to teach the rudiments of large-sample approximations than to teach the Student \( t \) and exact \( F \) distributions, degrees-of-freedom corrections, and so forth. This large-sample approach also saves students the frustration of discovering that, because of nonnormal errors, the exact distribution theory they just mastered is irrelevant. Once taught in the context of the sample mean, the large-sample approach to hypothesis testing and confidence intervals carries directly through multiple regression analysis, logit and probit, instrumental variables estimation, and time series methods.
• **Random sampling.** Because regressors are rarely fixed in econometric applications, from the outset we treat data on all variables (dependent and independent) as the result of random sampling. This assumption matches our initial applications to cross-sectional data, it extends readily to panel and time series data, and because of our large-sample approach, it poses no additional conceptual or mathematical difficulties.

• **Heteroskedasticity.** Applied econometricians routinely use heteroskedasticity-robust standard errors to eliminate worries about whether heteroskedasticity is present or not. In this book, we move beyond treating heteroskedasticity as an exception or a “problem” to be “solved”; instead, we allow for heteroskedasticity from the outset and simply use heteroskedasticity-robust standard errors. We present homoskedasticity as a special case that provides a theoretical motivation for OLS.

### Skilled Producers, Sophisticated Consumers

We hope that students using this book will become sophisticated consumers of empirical analysis. To do so, they must learn not only how to use the tools of regression analysis but also how to assess the validity of empirical analyses presented to them.

Our approach to teaching how to assess an empirical study is threefold. First, immediately after introducing the main tools of regression analysis, we devote Chapter 9 to the threats to internal and external validity of an empirical study. This chapter discusses data problems and issues of generalizing findings to other settings. It also examines the main threats to regression analysis, including omitted variables, functional form misspecification, errors-in-variables, selection, and simultaneity—ways to recognize these threats in practice.

Second, we apply these methods for assessing empirical studies to the empirical analysis of the ongoing examples in the book. We do so by considering alternative specifications and by systematically addressing the various threats to validity of the analyses presented in the book.

Third, to become sophisticated consumers, students need firsthand experience as producers. Active learning beats passive learning, and econometrics is an ideal course for active learning. For this reason, the textbook website features data sets, software, and suggestions for empirical exercises of different scopes.

### Approach to Mathematics and Level of Rigor

Our aim is for students to develop a sophisticated understanding of the tools of modern regression analysis, whether the course is taught at a “high” or a “low” level of mathematics. Parts I through IV of the text (which cover the substantive
material) are accessible to students with only precalculus mathematics. Parts I through IV have fewer equations and more applications than many introductory econometrics books and far fewer equations than books aimed at mathematical sections of undergraduate courses. But more equations do not imply a more sophisticated treatment. In our experience, a more mathematical treatment does not lead to a deeper understanding for most students.

That said, different students learn differently, and for mathematically well-prepared students, learning can be enhanced by a more explicitly mathematical treatment. Part V therefore contains an introduction to econometric theory that is appropriate for students with a stronger mathematical background. When the mathematical chapters in Part V are used in conjunction with the material in Parts I through IV, this book is suitable for advanced undergraduate or master’s level econometrics courses.

Contents and Organization

There are five parts to Introduction to Econometrics. This textbook assumes that the student has had a course in probability and statistics, although we review that material in Part I. We cover the core material of regression analysis in Part II. Parts III, IV, and V present additional topics that build on the core treatment in Part II.

Part I

Chapter 1 introduces econometrics and stresses the importance of providing quantitative answers to quantitative questions. It discusses the concept of causality in statistical studies and surveys the different types of data encountered in econometrics. Material from probability and statistics is reviewed in Chapters 2 and 3, respectively; whether these chapters are taught in a given course or are simply provided as a reference depends on the background of the students.

Part II

Chapter 4 introduces regression with a single regressor and ordinary least squares (OLS) estimation, and Chapter 5 discusses hypothesis tests and confidence intervals in the regression model with a single regressor. In Chapter 6, students learn how they can address omitted variable bias using multiple regression, thereby estimating the effect of one independent variable while holding other independent variables constant. Chapter 7 covers hypothesis tests, including $F$-tests, and confidence intervals in multiple regression. In Chapter 8, the linear regression model is
extended to models with nonlinear population regression functions, with a focus on regression functions that are linear in the parameters (so that the parameters can be estimated by OLS). In Chapter 9, students step back and learn how to identify the strengths and limitations of regression studies, seeing in the process how to apply the concepts of internal and external validity.

### Part III
Part III presents extensions of regression methods. In Chapter 10, students learn how to use panel data to control for unobserved variables that are constant over time. Chapter 11 covers regression with a binary dependent variable. Chapter 12 shows how instrumental variables regression can be used to address a variety of problems that produce correlation between the error term and the regressor, and examines how one might find and evaluate valid instruments. Chapter 13 introduces students to the analysis of data from experiments and quasi-, or natural, experiments, topics often referred to as “program evaluation.”

### Part IV
Part IV takes up regression with time series data. Chapter 14 focuses on forecasting and introduces various modern tools for analyzing time series regressions, such as unit root tests and tests for stability. Chapter 15 discusses the use of time series data to estimate causal relations. Chapter 16 presents some more advanced tools for time series analysis, including models of conditional heteroskedasticity.

### Part V
Part V is an introduction to econometric theory. This part is more than an appendix that fills in mathematical details omitted from the text. Rather, it is a self-contained treatment of the econometric theory of estimation and inference in the linear regression model. Chapter 17 develops the theory of regression analysis for a single regressor; the exposition does not use matrix algebra, although it does demand a higher level of mathematical sophistication than the rest of the text. Chapter 18 presents and studies the multiple regression model, instrumental variables regression, and generalized method of moments estimation of the linear model, all in matrix form.

### Prerequisites Within the Book
Because different instructors like to emphasize different material, we wrote this book with diverse teaching preferences in mind. To the maximum extent possible,
the chapters in Parts III, IV, and V are “stand-alone” in the sense that they do not require first teaching all the preceding chapters. The specific prerequisites for each chapter are described in Table I. Although we have found that the sequence of topics adopted in the textbook works well in our own courses, the chapters are written in a way that allows instructors to present topics in a different order if they so desire.

Sample Courses

This book accommodates several different course structures.

| TABLE I Guide to Prerequisites for Special-Topic Chapters in Parts III, IV, and V |
|-------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Chapter           | Part I 1–3      | Part II 4–7, 9  | Part III 8     | Part IV 10.1, 10.2 | Part IV 12.1, 12.2 | Part V 14.1–14.4 | Part V 14.5–14.8 |
| 10                | X^a             | X^a             | X               |                  |                  |                  |                  |
| 11                | X^a             | X^a             | X               |                  |                  |                  |                  |
| 12.1, 12.2        | X^a             | X^a             | X               |                  |                  |                  |                  |
| 12.3–12.6         | X^a             | X^a             | X               | X                | X                |                  |                  |
| 13                | X^a             | X^a             | X               | X                | X                |                  |                  |
| 14                | X^a             | X^a             | b               |                  |                  |                  |                  |
| 15                | X^a             | X^a             | b               |                  | X                |                  |                  |
| 16                | X^a             | X^a             | b               |                  | X                | X                |                  |
| 17                | X               | X               | X               |                  |                  |                  |                  |
| 18                | X               | X               | X               |                  |                  |                  | X                |

This table shows the minimum prerequisites needed to cover the material in a given chapter. For example, estimation of dynamic causal effects with time series data (Chapter 15) first requires Part I (as needed, depending on student preparation, and except as noted in footnote a), Part II (except for Chapter 8; see footnote b), and Sections 14.1 through 14.4.

^aChapters 10 through 16 use exclusively large-sample approximations to sampling distributions, so the optional Sections 3.6 (the Student t distribution for testing means) and 5.6 (the Student t distribution for testing regression coefficients) can be skipped.

^bChapters 14 through 16 (the time series chapters) can be taught without first teaching Chapter 8 (nonlinear regression functions) if the instructor pauses to explain the use of logarithmic transformations to approximate percentage changes.
Standard Introductory Econometrics
This course introduces econometrics (Chapter 1) and reviews probability and statistics as needed (Chapters 2 and 3). It then moves on to regression with a single regressor, multiple regression, the basics of functional form analysis, and the evaluation of regression studies (all Part II). The course proceeds to cover regression with panel data (Chapter 10), regression with a limited dependent variable (Chapter 11), and instrumental variables regression (Chapter 12), as time permits. The course concludes with experiments and quasi-experiments in Chapter 13, topics that provide an opportunity to return to the questions of estimating causal effects raised at the beginning of the semester and to recapitulate core regression methods. Prerequisites: Algebra II and introductory statistics.

Introductory Econometrics with Time Series and Forecasting Applications
Like a standard introductory course, this course covers all of Part I (as needed) and Part II. Optionally, the course next provides a brief introduction to panel data (Sections 10.1 and 10.2) and takes up instrumental variables regression (Chapter 12, or just Sections 12.1 and 12.2). The course then proceeds to Part IV, covering forecasting (Chapter 14) and estimation of dynamic causal effects (Chapter 15). If time permits, the course can include some advanced topics in time series analysis such as volatility clustering and conditional heteroskedasticity (Section 16.5). Prerequisites: Algebra II and introductory statistics.

Applied Time Series Analysis and Forecasting
This book also can be used for a short course on applied time series and forecasting, for which a course on regression analysis is a prerequisite. Some time is spent reviewing the tools of basic regression analysis in Part II, depending on student preparation. The course then moves directly to Part IV and works through forecasting (Chapter 14), estimation of dynamic causal effects (Chapter 15), and advanced topics in time series analysis (Chapter 16), including vector autoregressions and conditional heteroskedasticity. An important component of this course is hands-on forecasting exercises, available to instructors on the book’s accompanying website. Prerequisites: Algebra II and basic introductory econometrics or the equivalent.

Introduction to Econometric Theory
This book is also suitable for an advanced undergraduate course in which the students have a strong mathematical preparation or for a master’s level course in
econometrics. The course briefly reviews the theory of statistics and probability as necessary (Part I). The course introduces regression analysis using the nonmathematical, applications-based treatment of Part II. This introduction is followed by the theoretical development in Chapters 17 and 18 (through Section 18.5). The course then takes up regression with a limited dependent variable (Chapter 11) and maximum likelihood estimation (Appendix 11.2). Next, the course optionally turns to instrumental variables regression and generalized method of moments (Chapter 12 and Section 18.7), time series methods (Chapter 14), and the estimation of causal effects using time series data and generalized least squares (Chapter 15 and Section 18.6). Prerequisites: Calculus and introductory statistics. Chapter 18 assumes previous exposure to matrix algebra.

Pedagogical Features

This textbook has a variety of pedagogical features aimed at helping students understand, retain, and apply the essential ideas. Chapter introductions provide real-world grounding and motivation, as well as brief road maps highlighting the sequence of the discussion. Key terms are boldfaced and defined in context throughout each chapter, and Key Concept boxes at regular intervals recap the central ideas. General interest boxes provide interesting excursions into related topics and highlight real-world studies that use the methods or concepts being discussed in the text. A Summary concluding each chapter serves as a helpful framework for reviewing the main points of coverage. The questions in the Review the Concepts section check students’ understanding of the core content, Exercises give more intensive practice working with the concepts and techniques introduced in the chapter, and Empirical Exercises allow students to apply what they have learned to answer real-world empirical questions. At the end of the textbook, the Appendix provides statistical tables, the References section lists sources for further reading, and a Glossary conveniently defines many key terms in the book.

Supplements to Accompany the Textbook

The online supplements accompanying the third edition update of Introduction to Econometrics include the Instructor’s Resource Manual, Test Bank, and PowerPoint® slides with text figures, tables, and Key Concepts. The Instructor’s Resource Manual includes solutions to all the end-of-chapter exercises, while the Test Bank, offered in Testgen, provides a rich supply of easily edited test problems and
questions of various types to meet specific course needs. These resources are available for download from the Instructor’s Resource Center at www.pearsonglobaleditions.com/Stock_Watson.

Companion Website

The Companion Website, found at www.pearsonglobaleditions.com/Stock_Watson, provides a wide range of additional resources for students and faculty. These resources include more and more in depth empirical exercises, data sets for the empirical exercises, replication files for empirical results reported in the text, practice quizzes, answers to end-of-chapter Review the Concepts questions and Exercises, and EViews tutorials.

MyEconLab

The third edition update is accompanied by a robust MyEconLab course. The MyEconLab course includes all the Review the Concepts questions as well as some Exercises and Empirical Exercises. In addition, the enhanced eText available in MyEconLab for the third edition update includes URL links from the Exercises and Empirical Exercises to questions in the MyEconLab course and to the data that accompanies them. To register for MyEconLab and to learn more, log on to www.myeconlab.com.

Acknowledgments

A great many people contributed to the first edition of this book. Our biggest debts of gratitude are to our colleagues at Harvard and Princeton who used early drafts of this book in their classrooms. At Harvard’s Kennedy School of Government, Suzanne Cooper provided invaluable suggestions and detailed comments on multiple drafts. As a coteacher with one of the authors (Stock), she also helped vet much of the material in this book while it was being developed for a required course for master’s students at the Kennedy School. We are also indebted to two other Kennedy School colleagues, Alberto Abadie and Sue Dynarski, for their patient explanations of quasi-experiments and the field of program evaluation and for their detailed comments on early drafts of the text. At Princeton, Eli Tamer taught from an early draft and also provided helpful comments on the penultimate draft of the book.
We also owe much to many of our friends and colleagues in econometrics who spent time talking with us about the substance of this book and who collectively made so many helpful suggestions. Bruce Hansen (University of Wisconsin–Madison) and Bo Honore (Princeton) provided helpful feedback on very early outlines and preliminary versions of the core material in Part II. Joshua Angrist (MIT) and Guido Imbens (University of California, Berkeley) provided thoughtful suggestions about our treatment of materials on program evaluation. Our presentation of the material on time series has benefited from discussions with Yacine Ait-Sahalia (Princeton), Graham Elliott (University of California, San Diego), Andrew Harvey (Cambridge University), and Christopher Sims (Princeton). Finally, many people made helpful suggestions on parts of the manuscript close to their area of expertise: Don Andrews (Yale), John Bound (University of Michigan), Gregory Chow (Princeton), Thomas Downes (Tufts), David Drukker (StataCorp.), Jean Baldwin Grossman (Princeton), Eric Hanushek (Hoover Institution), James Heckman (University of Chicago), Han Hong (Princeton), Caroline Hoxby (Harvard), Alan Krueger (Princeton), Steven Levitt (University of Chicago), Richard Light (Harvard), David Neumark (Michigan State University), Joseph Newhouse (Harvard), Pierre Perron (Boston University), Kenneth Warner (University of Michigan), and Richard Zeckhauser (Harvard).

Many people were very generous in providing us with data. The California test score data were constructed with the assistance of Les Axelrod of the Standards and Assessments Division, California Department of Education. We are grateful to Charlie DePascale, Student Assessment Services, Massachusetts Department of Education, for his help with aspects of the Massachusetts test score data set. Christopher Ruhm (University of North Carolina, Greensboro) graciously provided us with his data set on drunk driving laws and traffic fatalities. The research department at the Federal Reserve Bank of Boston deserves thanks for putting together its data on racial discrimination in mortgage lending; we particularly thank Geoffrey Tootell for providing us with the updated version of the data set we use in Chapter 9 and Lynn Browne for explaining its policy context. We thank Jonathan Gruber (MIT) for sharing his data on cigarette sales, which we analyze in Chapter 12, and Alan Krueger (Princeton) for his help with the Tennessee STAR data that we analyze in Chapter 13.

We thank several people for carefully checking the page proof for errors. Kerry Griffin and Yair Listokin read the entire manuscript, and Andrew Fraker, Ori Heffetz, Amber Henry, Hong Li, Alessandro Tarozzi, and Matt Watson worked through several chapters.

In the first edition, we benefited from the help of an exceptional development editor, Jane Tufts, whose creativity, hard work, and attention to detail improved
the book in many ways, large and small. Pearson provided us with first-rate sup-
port, starting with our excellent editor, Sylvia Mallory, and extending through the
entire publishing team. Jane and Sylvia patiently taught us a lot about writing,
organization, and presentation, and their efforts are evident on every page of this
book. We extend our thanks to the superb Pearson team, who worked with us on
the second edition: Adrienne D’Ambrosio (senior acquisitions editor), Bridget
Page (associate media producer), Charles Spaulding (senior designer), Nancy
Fenton (managing editor) and her selection of Nancy Freihofer and Thompson
Steele Inc. who handled the entire production process, Heather McNally (sup-
plements coordinator), and Denise Clinton (editor-in-chief). Finally, we had the
benefit of Kay Ueno’s skilled editing in the second edition. We are also grate-
ful to the excellent third edition Pearson team of Adrienne D’Ambrosio, Nancy
Fenton, and Jill Kolongowski, as well as Mary Sanger, the project manager with
Nesbitt Graphics. We also wish to thank the Pearson team who worked on the
third edition update: Christina Masturzo, Carolyn Philips, Liz Napolitano, and
Heidi Allgair, project manager with Cenveo® Publisher Services.

We also received a great deal of help and suggestions from faculty, students,
and researchers as we prepared the third edition and its update. The changes
made in the third edition incorporate or reflect suggestions, corrections, com-
ments, data, and help provided by a number of researchers and instructors: Don-
ald Andrews (Yale University), Jushan Bai (Columbia), James Cobbe (Florida
State University), Susan Dynarski (University of Michigan), Nicole Eichelberger
(Texas Tech University), Boyd Fjeldsted (University of Utah), Martina Grunow,
Daniel Hamermesh (University of Texas–Austin), Keisuke Hirano (University
of Arizona), Bo Honore (Princeton University), Guido Imbens (Harvard Uni-
versity), Manfred Keil (Claremont McKenna College), David Laibson (Harvard
University), David Lee (Princeton University), Brigitte Madrian (Harvard Uni-
versity), Jorge Marquez (University of Maryland), Karen Bennett Mathis (Flor-
da Department of Citrus), Alan Mehlenbacher (University of Victoria), Ulrich
Müller (Princeton University), Serena Ng (Columbia University), Harry Patrinos
(World Bank), Zhuan Pei (Brandeis University), Peter Summers (Texas Tech
University), Andrey Vasnov (University of Sydney), and Douglas Young (Mont-
tana State University). We also benefited from student input from F. Hoces dela
Guardia and Carrie Wilson.

Thoughtful reviews for the third edition were prepared for Addison-Wesley
by Steve DeLoach (Elon University), Jeffrey DeSimone (University of Texas at
Arlington), Gary V. Engelhardt (Syracuse University), Luca Flabbi (Georgetown
University), Steffen Habermalz (Northwestern University), Carolyn J. Heinrich
(University of Wisconsin–Madison), Emma M. Iglesias-Vazquez (Michigan State
University), Carlos Lamarche (University of Oklahoma), Vicki A. McCracken (Washington State University), Claudiney M. Pereira (Tulane University), and John T. Warner (Clemson University). We also received very helpful input on draft revisions of Chapters 7 and 10 from John Berdell (DePaul University), Janet Kohlhase (University of Houston), Aprajit Mahajan (Stanford University), Xia Meng (Brandeis University), and Chan Shen (Georgetown University).

Above all, we are indebted to our families for their endurance throughout this project. Writing this book took a long time, and for them, the project must have seemed endless. They, more than anyone else, bore the burden of this commitment, and for their help and support we are deeply grateful.

Pearson would also like to thank and acknowledge Anisha Sharma (University of Oxford) and Samprit Chakrabarti (International School of Business and Media) for their contributions and reviews for the Global Edition, and Indrani Chkarborty (Institute of Development Studies), Gagari Chakrabarti (Presidency University), Sukanya Das (Madras School of Economics), and Abhik Kumar Mukherjee (St. Xavier’s College) for their guidance and recommendations to improve the global content.