# 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>43</td>
</tr>
<tr>
<td>2</td>
<td>Review of Probability</td>
<td>56</td>
</tr>
<tr>
<td>3</td>
<td>Review of Statistics</td>
<td>106</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>149</td>
</tr>
<tr>
<td>5</td>
<td>Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals</td>
<td>186</td>
</tr>
<tr>
<td>6</td>
<td>Linear Regression with Multiple Regressors</td>
<td>221</td>
</tr>
<tr>
<td>7</td>
<td>Hypothesis Tests and Confidence Intervals in Multiple Regression</td>
<td>256</td>
</tr>
<tr>
<td>8</td>
<td>Nonlinear Regression Functions</td>
<td>294</td>
</tr>
<tr>
<td>9</td>
<td>Assessing Studies Based on Multiple Regression</td>
<td>354</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>389</td>
</tr>
<tr>
<td>11</td>
<td>Regression with a Binary Dependent Variable</td>
<td>423</td>
</tr>
<tr>
<td>12</td>
<td>Instrumental Variables Regression</td>
<td>461</td>
</tr>
<tr>
<td>13</td>
<td>Experiments and Quasi-Experiments</td>
<td>511</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>558</td>
</tr>
<tr>
<td>15</td>
<td>Estimation of Dynamic Causal Effects</td>
<td>625</td>
</tr>
<tr>
<td>16</td>
<td>Additional Topics in Time Series Regression</td>
<td>673</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>711</td>
</tr>
<tr>
<td>18</td>
<td>The Theory of Multiple Regression</td>
<td>739</td>
</tr>
</tbody>
</table>
## Contents

### Preface 29

### PART ONE  Introduction and Review

#### CHAPTER 1 Economic Questions and Data 43

1.1 Economic Questions We Examine 43

- Question #1: Does Reducing Class Size Improve Elementary School Education? 44
- Question #2: Is There Racial Discrimination in the Market for Home Loans? 45
- Question #3: How Much Do Cigarette Taxes Reduce Smoking? 45
- Question #4: What Will the Rate of Inflation Be Next Year? 46

1.2 Causal Effects and Idealized Experiments 47

- Estimation of Causal Effects 48
- Forecasting and Causality 49

1.3 Data: Sources and Types 49

- Experimental Versus Observational Data 49
- Cross-Sectional Data 50
- Time Series Data 51
- Panel Data 53

#### CHAPTER 2 Review of Probability 56

2.1 Random Variables and Probability Distributions 57

- Probabilities, the Sample Space, and Random Variables 57
- Probability Distribution of a Discrete Random Variable 58
- Probability Distribution of a Continuous Random Variable 60

2.2 Expected Values, Mean, and Variance 60

- The Expected Value of a Random Variable 60
- The Standard Deviation and Variance 63
- Mean and Variance of a Linear Function of a Random Variable 64
- Other Measures of the Shape of a Distribution 65
2.3 Two Random Variables 68
Joint and Marginal Distributions 68
Conditional Distributions 69
Independence 73
Covariance and Correlation 73
The Mean and Variance of Sums of Random Variables 74

2.4 The Normal, Chi-Squared, Student t, and F Distributions 78
The Normal Distribution 78
The Chi-Squared Distribution 83
The Student t Distribution 83
The F Distribution 84

2.5 Random Sampling and the Distribution of the Sample Average 85
Random Sampling 85
The Sampling Distribution of the Sample Average 86

2.6 Large-Sample Approximations to Sampling Distributions 89
The Law of Large Numbers and Consistency 90
The Central Limit Theorem 92
APPENDIX 2.1 Derivation of Results in Key Concept 2.3 104

CHAPTER 3 Review of Statistics 106

3.1 Estimation of the Population Mean 107
Estimators and Their Properties 107
Properties of $\bar{Y}$ 109
The Importance of Random Sampling 111

3.2 Hypothesis Tests Concerning the Population Mean 112
Null and Alternative Hypotheses 112
The $p$-Value 113
Calculating the $p$-Value When $\sigma_Y$ Is Known 114
The Sample Variance, Sample Standard Deviation, and Standard Error 115
Calculating the $p$-Value When $\sigma_Y$ Is Unknown 117
The $t$-Statistic 117
Hypothesis Testing with a Prespecified Significance Level 118
One-Sided Alternatives 120
CHAPTER 5 Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals 186

5.1 Testing Hypotheses About One of the Regression Coefficients 186
Two-Sided Hypotheses Concerning $\beta_1$ 187
One-Sided Hypotheses Concerning $\beta_1$ 190
Testing Hypotheses About the Intercept $\beta_0$ 192

5.2 Confidence Intervals for a Regression Coefficient 193

5.3 Regression When $X$ Is a Binary Variable 195
Interpretation of the Regression Coefficients 195

5.4 Heteroskedasticity and Homoskedasticity 197
What Are Heteroskedasticity and Homoskedasticity? 198
Mathematical Implications of Homoskedasticity 200
What Does This Mean in Practice? 201

5.5 The Theoretical Foundations of Ordinary Least Squares 203
Linear Conditionally Unbiased Estimators and the Gauss–Markov Theorem 204
Regression Estimators Other Than OLS 205
5.6 Using the $t$-Statistic in Regression When the Sample Size Is Small 206

The $t$-Statistic and the Student $t$ Distribution 206
Use of the Student $t$ Distribution in Practice 207

5.7 Conclusion 208

APPENDIX 5.1 Formulas for OLS Standard Errors 216
APPENDIX 5.2 The Gauss–Markov Conditions and a Proof of the Gauss–Markov Theorem 217

CHAPTER 6 Linear Regression with Multiple Regressors 221

6.1 Omitted Variable Bias 221
Definition of Omitted Variable Bias 222
A Formula for Omitted Variable Bias 224
Addressing Omitted Variable Bias by Dividing the Data into Groups 226

6.2 The Multiple Regression Model 228
The Population Regression Line 228
The Population Multiple Regression Model 229

6.3 The OLS Estimator in Multiple Regression 231
The OLS Estimator 232
Application to Test Scores and the Student–Teacher Ratio 233

6.4 Measures of Fit in Multiple Regression 235
The Standard Error of the Regression (SER) 235
The $R^2$ 235
The "Adjusted $R^2" 236
Application to Test Scores 237

6.5 The Least Squares Assumptions in Multiple Regression 238
Assumption #1: The Conditional Distribution of $u_i$ Given $X_{1i}, X_{2i}, \ldots, X_{ki}$ Has a Mean of Zero 238
Assumption #2: $(X_{1i}, X_{2i}, \ldots, X_{ki}, Y_i), i = 1, \ldots, n$, Are i.i.d. 238
Assumption #3: Large Outliers Are Unlikely 238
Assumption #4: No Perfect Multicollinearity 239

6.6 The Distribution of the OLS Estimators in Multiple Regression 240
6.7 Multicollinearity 241
    Examples of Perfect Multicollinearity 242
    Imperfect Multicollinearity 244

6.8 Conclusion 245

APPENDIX 6.1 Derivation of Equation (6.1) 253
APPENDIX 6.2 Distribution of the OLS Estimators When There Are Two Regressors and Homoskedastic Errors 254
APPENDIX 6.3 The Frisch–Waugh Theorem 254

CHAPTER 7 Hypothesis Tests and Confidence Intervals in Multiple Regression 256

7.1 Hypothesis Tests and Confidence Intervals for a Single Coefficient 256
    Standard Errors for the OLS Estimators 256
    Hypothesis Tests for a Single Coefficient 257
    Confidence Intervals for a Single Coefficient 258
    Application to Test Scores and the Student–Teacher Ratio 259

7.2 Tests of Joint Hypotheses 261
    Testing Hypotheses on Two or More Coefficients 261
    The F-Statistic 263
    Application to Test Scores and the Student–Teacher Ratio 265
    The Homoskedasticity-Only F-Statistic 266

7.3 Testing Single Restrictions Involving Multiple Coefficients 268

7.4 Confidence Sets for Multiple Coefficients 270

7.5 Model Specification for Multiple Regression 271
    Omitted Variable Bias in Multiple Regression 272
    The Role of Control Variables in Multiple Regression 272
    Model Specification in Theory and in Practice 275
    Interpreting the $R^2$ and the Adjusted $R^2$ in Practice 276

7.6 Analysis of the Test Score Data Set 277

7.7 Conclusion 282
APPENDIX 7.1 The Bonferroni Test of a Joint Hypothesis 290
APPENDIX 7.2 Conditional Mean Independence 292

CHAPTER 8 Nonlinear Regression Functions 294

8.1 A General Strategy for Modeling Nonlinear Regression Functions 296
Test Scores and District Income 296
The Effect on Y of a Change in X in Nonlinear Specifications 299
A General Approach to Modeling Nonlinearities Using Multiple Regression 304

8.2 Nonlinear Functions of a Single Independent Variable 304
Polynomials 305
Logarithms 307
Polynomial and Logarithmic Models of Test Scores and District Income 315

8.3 Interactions Between Independent Variables 316
Interactions Between Two Binary Variables 317
Interactions Between a Continuous and a Binary Variable 320
Interactions Between Two Continuous Variables 324

8.4 Nonlinear Effects on Test Scores of the Student–Teacher Ratio 328
Discussion of Regression Results 331
Summary of Findings 335

8.5 Conclusion 336

APPENDIX 8.1 Regression Functions That Are Nonlinear in the Parameters 348
APPENDIX 8.2 Slopes and Elasticities for Nonlinear Regression Functions 351

CHAPTER 9 Assessing Studies Based on Multiple Regression 354

9.1 Internal and External Validity 354
Threats to Internal Validity 355
Threats to External Validity 356
9.2 Threats to Internal Validity of Multiple Regression Analysis 358
Omitted Variable Bias 358
Misspecification of the Functional Form of the Regression Function 360
Measurement Error and Errors-in-Variables Bias 361
Missing Data and Sample Selection 364
Simultaneous Causality 366
Sources of Inconsistency of OLS Standard Errors 368

9.3 Internal and External Validity When the Regression Is Used for Forecasting 369
Using Regression Models for Forecasting 369
Assessing the Validity of Regression Models for Forecasting 371

9.4 Example: Test Scores and Class Size 371
External Validity 371
Internal Validity 378
Discussion and Implications 380

9.5 Conclusion 381
APPENDIX 9.1 The Massachusetts Elementary School Testing Data 387

PART THREE Further Topics in Regression Analysis
CHAPTER 10 Regression with Panel Data 389

10.1 Panel Data 390
Example: Traffic Deaths and Alcohol Taxes 391

10.2 Panel Data with Two Time Periods: "Before and After" Comparisons 393

10.3 Fixed Effects Regression 396
The Fixed Effects Regression Model 396
Estimation and Inference 398
Application to Traffic Deaths 400

10.4 Regression with Time Fixed Effects 400
Time Effects Only 401
Both Entity and Time Fixed Effects 402
10.5 The Fixed Effects Regression Assumptions and Standard Errors for Fixed Effects Regression  404
   The Fixed Effects Regression Assumptions  404
   Standard Errors for Fixed Effects Regression  406
10.6 Drunk Driving Laws and Traffic Deaths  407
10.7 Conclusion  411
APPENDIX 10.1 The State Traffic Fatality Data Set  418
APPENDIX 10.2 Standard Errors for Fixed Effects Regression  418

CHAPTER 11 Regression with a Binary Dependent Variable  423
11.1 Binary Dependent Variables and the Linear Probability Model  424
   Binary Dependent Variables  424
   The Linear Probability Model  426
11.2 Probit and Logit Regression  429
   Probit Regression  429
   Logit Regression  434
   Comparing the Linear Probability, Probit, and Logit Models  436
11.3 Estimation and Inference in the Logit and Probit Models  436
   Nonlinear Least Squares Estimation  437
   Maximum Likelihood Estimation  438
   Measures of Fit  439
11.4 Application to the Boston HMDA Data  440
11.5 Conclusion  447
APPENDIX 11.1 The Boston HMDA Data Set  455
APPENDIX 11.2 Maximum Likelihood Estimation  455
APPENDIX 11.3 Other Limited Dependent Variable Models  458

CHAPTER 12 Instrumental Variables Regression  461
12.1 The IV Estimator with a Single Regressor and a Single Instrument  462
   The IV Model and Assumptions  462
   The Two Stage Least Squares Estimator  463
13.2 Threats to Validity of Experiments 515
    Threats to Internal Validity 515
    Threats to External Validity 519

13.3 Experimental Estimates of the Effect of Class Size Reductions 520
    Experimental Design 520
    Analysis of the STAR Data 522
    Comparison of the Observational and Experimental Estimates of Class Size Effects 527

13.4 Quasi-Experiments 529
    Examples 530
    The Differences-in-Differences Estimator 532
    Instrumental Variables Estimators 536
    Regression Discontinuity Estimators 536

13.5 Potential Problems with Quasi-Experiments 538
    Threats to Internal Validity 538
    Threats to External Validity 540

13.6 Experimental and Quasi-Experimental Estimates in Heterogeneous Populations 540
    OLS with Heterogeneous Causal Effects 541
    IV Regression with Heterogeneous Causal Effects 542

13.7 Conclusion 545

APPENDIX 13.1 The Project STAR Data Set 554
APPENDIX 13.2 IV Estimation When the Causal Effect Varies Across Individuals 555
APPENDIX 13.3 The Potential Outcomes Framework for Analyzing Data from Experiments 556

PART FOUR Regression Analysis of Economic Time Series Data

CHAPTER 14 Introduction to Time Series Regression and Forecasting 558

14.1 Using Regression Models for Forecasting 559

14.2 Introduction to Time Series Data and Serial Correlation 560
    The Rates of Inflation and Unemployment in the United States 560
16.2 Multiperiod Forecasts 678
   Iterated Multiperiod Forecasts 678
   Direct Multiperiod Forecasts 680
   Which Method Should You Use? 683

16.3 Orders of Integration and the DF-GLS Unit Root Test 684
   Other Models of Trends and Orders of Integration 684
   The DF-GLS Test for a Unit Root 686
   Why Do Unit Root Tests Have Nonnormal Distributions? 689

16.4 Cointegration 691
   Cointegration and Error Correction 691
   How Can You Tell Whether Two Variables Are Cointegrated? 693
   Estimation of Cointegrating Coefficients 696
   Extension to Multiple Cointegrated Variables 697
   Application to Interest Rates 698

16.5 Volatility Clustering and Autoregressive Conditional Heteroskedasticity 701
   Volatility Clustering 701
   Autoregressive Conditional Heteroskedasticity 702
   Application to Stock Price Volatility 704

16.6 Conclusion 704

APPENDIX 16.1 U.S. Financial Data Used in Chapter 16 710

PART FIVE The Econometric Theory of Regression Analysis

CHAPTER 17 The Theory of Linear Regression with One Regressor 711

17.1 The Extended Least Squares Assumptions and the OLS Estimator 712
   The Extended Least Squares Assumptions 712
   The OLS Estimator 714

17.2 Fundamentals of Asymptotic Distribution Theory 714
   Convergence in Probability and the Law of Large Numbers 715
   The Central Limit Theorem and Convergence in Distribution 717
Slutsky's Theorem and the Continuous Mapping Theorem 718
Application to the t-Statistic Based on the Sample Mean 719

17.3 Asymptotic Distribution of the OLS Estimator and t-Statistic 720
Consistency and Asymptotic Normality of the OLS Estimators 720
Consistency of Heteroskedasticity-Robust Standard Errors 720
Asymptotic Normality of the Heteroskedasticity-Robust t-Statistic 722

17.4 Exact Sampling Distributions When the Errors Are Normally Distributed 722
Distribution of \( \hat{\beta}_1 \) with Normal Errors 722
Distribution of the Homoskedasticity-Only t-Statistic 724

17.5 Weighted Least Squares 725
WLS with Known Heteroskedasticity 725
WLS with Heteroskedasticity of Known Functional Form 726
Heteroskedasticity-Robust Standard Errors or WLS? 729
APPENDIX 17.1 The Normal and Related Distributions and Moments of Continuous Random Variables 734
APPENDIX 17.2 Two Inequalities 737

CHAPTER 18 The Theory of Multiple Regression 739

18.1 The Linear Multiple Regression Model and OLS Estimator in Matrix Form 740
The Multiple Regression Model in Matrix Notation 740
The Extended Least Squares Assumptions 742
The OLS Estimator 743

18.2 Asymptotic Distribution of the OLS Estimator and t-Statistic 744
The Multivariate Central Limit Theorem 744
Asymptotic Normality of \( \hat{\beta} \) 745
Heteroskedasticity-Robust Standard Errors 746
Confidence Intervals for Predicted Effects 747
Asymptotic Distribution of the t-Statistic 747

18.3 Tests of Joint Hypotheses 748
Joint Hypotheses in Matrix Notation 748
Asymptotic Distribution of the F-Statistic 748
Confidence Sets for Multiple Coefficients 749
18.4 Distribution of Regression Statistics with Normal Errors 750
   Matrix Representations of OLS Regression Statistics 750
   Distribution of $\hat{\beta}$ for Normal Errors 751
   Distribution of $s^2$ 752
   Homoskedasticity-Only Standard Errors 752
   Distribution of the $t$-Statistic 753
   Distribution of the $F$-Statistic 753

18.5 Efficiency of the OLS Estimator with Homoskedastic Errors 754
   The Gauss–Markov Conditions for Multiple Regression 754
   Linear Conditionally Unbiased Estimators 754
   The Gauss–Markov Theorem for Multiple Regression 755

18.6 Generalized Least Squares 756
   The GLS Assumptions 757
   GLS When $\Omega$ Is Known 759
   GLS When $\Omega$ Contains Unknown Parameters 760
   The Zero Conditional Mean Assumption and GLS 760

18.7 Instrumental Variables and Generalized Method of Moments Estimation 762
   The IV Estimator in Matrix Form 763
   Asymptotic Distribution of the TSLS Estimator 764
   Properties of TSLS When the Errors Are Homoskedastic 765
   Generalized Method of Moments Estimation in Linear Models 768

APPENDIX 18.1 Summary of Matrix Algebra 779
APPENDIX 18.2 Multivariate Distributions 783
APPENDIX 18.3 Derivation of the Asymptotic Distribution of $\hat{\beta}$ 784
APPENDIX 18.4 Derivations of Exact Distributions of OLS Test Statistics with Normal Errors 785
APPENDIX 18.5 Proof of the Gauss–Markov Theorem for Multiple Regression 786
APPENDIX 18.6 Proof of Selected Results for IV and GMM Estimation 788