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
Quantitative Questions, Quantitative Answers 51

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

Stock, James H.
Introduction to econometrics, update 2015

digitalisiert durch:
IDS Basel Bern

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 $\overline{Y}$ 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 σ_Y Is Known 119
The Sample Variance, Sample Standard Deviation, and Standard Error 120
Calculating the p-Value When σ_Y 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 $\overline{Y}$ Is the Least Squares Estimator of μ_Y 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 R^2 167
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 u_i Given X_i Has a Mean of Zero 170
Assumption #2: (X_i, Y_i), $i = 1,\ldots, n$, Are Independently and Identically Distributed 172
Assumption #3: Large Outliers Are Unlikely 173
Use of the Least Squares Assumptions 174

4.5 Sampling Distribution of the OLS Estimators 175
The Sampling Distribution of the OLS Estimators 176

4.6 Conclusion 179

APPENDIX 4.1 The California Test Score Data Set 187

APPENDIX 4.2 Derivation of the OLS Estimators 187

APPENDIX 4.3 Sampling Distribution of the OLS Estimator 188

CHAPTER 5 Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals 192

5.1 Testing Hypotheses About One of the Regression Coefficients 192
Two-Sided Hypotheses Concerning β_1 193
One-Sided Hypotheses Concerning β_1 196
Testing Hypotheses About the Intercept β_0 198

5.2 Confidence Intervals for a Regression Coefficient 199

5.3 Regression When X Is a Binary Variable 201
Interpretation of the Regression Coefficients 201

5.4 Heteroskedasticity and Homoskedasticity 203
What Are Heteroskedasticity and Homoskedasticity? 204
Mathematical Implications of Homoskedasticity 206
What Does This Mean in Practice? 207

5.5 The Theoretical Foundations of Ordinary Least Squares 209
Linear Conditionally Unbiased Estimators and the Gauss–Markov Theorem 210
Regression Estimators Other Than OLS 211

5.6 Using the t-Statistic in Regression When the Sample Size Is Small 212
The t-Statistic and the Student t Distribution 212
Use of the Student t Distribution in Practice 213

5.7 Conclusion 214

APPENDIX 5.1 Formulas for OLS Standard Errors 223

APPENDIX 5.2 The Gauss–Markov Conditions and a Proof of the Gauss–Markov Theorem 224

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

8.3 Interactions Between Independent Variables 324
Interactions Between Two Binary Variables 325
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 339
Discussion of Regression Results 339
Summary of Findings 343
8.5 Conclusion 344
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 361
9.1 Internal and External Validity 361
Threats to Internal Validity 362
Threats to External Validity 363
9.2 Threats to Internal Validity of Multiple Regression Analysis 365
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 377
Using Regression Models for Forecasting 377
Assessing the Validity of Regression Models for Forecasting 378
9.4 Example: Test Scores and Class Size 378
External Validity 378
Internal Validity 385
Discussion and Implications 387
9.5 Conclusion 388
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

Maximum Likelihood Estimation 446
Measures of Fit 447

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

14.7 Nonstationarity II: Breaks 607
What Is a Break? 608
Testing for Breaks 608
Pseudo Out-of-Sample Forecasting 613
Avoiding the Problems Caused by Breaks 619
14.8 Conclusion 619
APPENDIX 14.1 Time Series Data Used in Chapter 14 629
APPENDIX 14.2 Stationarity in the AR(1) Model 630
APPENDIX 14.3 Lag Operator Notation 631
APPENDIX 14.4 ARMA Models 632
APPENDIX 14.5 Consistency of the BIC Lag Length Estimator 633

CHAPTER 15 Estimation of Dynamic Causal Effects 635
15.1 An Initial Taste of the Orange Juice Data 636
15.2 Dynamic Causal Effects 639
Causal Effects and Time Series Data 639
Two Types of Exogeneity 642
15.3 Estimation of Dynamic Causal Effects with Exogenous Regressors 643
The Distributed Lag Model Assumptions 644
Autocorrelated u_t, Standard Errors, and Inference 645
Dynamic Multipliers and Cumulative Dynamic Multipliers 646
15.4 Heteroskedasticity- and Autocorrelation-Consistent Standard Errors 647
Distribution of the OLS Estimator with Autocorrelated Errors 602
HAC Standard Errors 650
15.5 Estimation of Dynamic Causal Effects with Strictly Exogenous Regressors 652
The Distributed Lag Model with AR(1) Errors 653
OLS Estimation of the ADL Model 656
GLS Estimation 657
The Distributed Lag Model with Additional Lags and AR(p) Errors 659
15.6 Orange Juice Prices and Cold Weather 662

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 $\hat{\beta}_1$ 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 $s_{\hat{u}}^2$ 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 Ω Is Known 771
GLS When Ω 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