Practical Econometrics
data collection, analysis, and application

Christiana E. Hilmer
San Diego State University

Michael J. Hilmer
San Diego State University
Table of Contents

PART ONE
THE BASICS  1

Chapter 1
An Introduction to Econometrics and Statistical Inference  1

Chapter Objectives  1
A Student’s Perspective  1
Big Picture Overview  1

1.1 Understand the Steps Involved in Conducting an Empirical Research Project  3

1.2 Understand the Meaning of the Term Econometrics  4

1.3 Understand the Relationship among Populations, Samples, and Statistical Inference  5
   Populations and Samples  5
   A Real-World Example of Statistical Inference: The Nielsen Ratings  8

1.4 Understand the Important Role that Sampling Distributions Play in Statistical Inference  9
   Additions to Our Empirical Research Toolkit  10
   Our New Empirical Tools in Practice: Using What We Have Learned in This Chapter  11
   Looking Ahead to Chapter 2  11
   Problems 11

Chapter 2
Collection and Management of Data  12

Chapter Objectives  12
A Student’s Perspective  12
Big Picture Overview  12

2.1 Consider Potential Sources of Data  13

2.2 Work Through an Example of the First Three Steps in Conducting an Empirical Research Project  16

2.3 Develop Data Management Skills  20

2.4 Understand Some Useful Excel Commands  22
   Installing the Data Analysis ToolPak  22
   Importing Data from the Web  24
   Creating New Worksheets  27
   Sorting Data from Lowest to Highest and Highest to Lowest  28
   Cut, Copy, and Paste Columns and Rows  29
   Use the Function Tool in Excel  29
   Copy Cell Entries Down a Column  30
   Use the Paste Special Command to Copy Values  31

Use the Paste Special Command to Transpose Columns  32
Additions to Our Empirical Research Toolkit  33
Our New Empirical Tools in Practice: Using What We Have Learned in This Chapter  34
Looking Ahead to Chapter 3  34
Problems 34
Exercises 35

Chapter 3
Summary Statistics  36

Chapter Objectives  36
A Student’s Perspective  36
Big Picture Overview  36

3.1 Construct Relative Frequency Histograms for a Given Variable  38
   Constructing a Relative Frequency Histogram  40

3.2 Calculate Measures of Central Tendency for a Given Variable  42
   The Sample Mean  43
   The Sample Median  44

3.3 Calculate Measures of Dispersion for a Given Variable  46
   Variance and Standard Deviation  46
   Percentiles  48
   The Five-Number Summary  48

3.4 Use Measures of Central Tendency and Dispersion for a Given Variable  51

3.5 Detect Whether Outliers for a Given Variable Are Present in Our Sample  55
   Detecting Outliers if the Data Set Is Symmetric  56
   Detecting Outliers if the Data Set Is Skewed  56

3.6 Construct Scatter Diagrams for the Relationship between Two Variables  58

3.7 Calculate the Covariance and the Correlation Coefficient for the Linear Relationship between y and x for Two Variables of Interest  59
   Additions to Our Empirical Research Toolkit  63
   Our New Empirical Tools in Practice: Using What We Have Learned in This Chapter  63
   Looking Ahead to Chapter 4  65
   Problems 65
   Exercises 68

XIII
PART TWO
LINEAR REGRESSION ANALYSIS 70

Chapter 4
Simple Linear Regression 70

Chapter Objectives 70
A Student’s Perspective 70
Big Picture Overview 71
Data to Be Analyzed: Our City Property Crime and CEO Compensation Samples 72
Data Analyzed in the Text 72
Data Analyzed in the Excel Boxes 72

4.1 Understand the Goals of Simple Linear Regression Analysis 73
4.2 Consider What the Random Error Component Contains 77
4.3 Define the Population Regression Model and the Sample Regression Function 78
4.4 Estimate the Sample Regression Function 80
4.5 Interpret the Estimated Sample Regression Function 83
4.6 Predict Outcomes Based on Our Estimated Sample Regression Function 84
4.7 Assess the Goodness-of-Fit of the Estimated Sample Regression Function 85
   Measure the Explained and Unexplained Variation in y 86
   Two Potential Measures of the Relative Goodness-of-Fit of Our Estimated Sample Regression Function 89

4.8 Understand How to Read Regression Output in Excel 94
4.9 Understand the Difference between Correlation and Causation 96
Additions to Our Empirical Research Toolkit 98
Our New Empirical Tools in Practice: Using What We Have Learned in This Chapter 98
Looking Ahead to Chapter 5 99
Problems 101
Exercises 102
References 102

Chapter 5
Hypothesis Testing for Linear Regression Analysis 103

Chapter Objectives 103
A Student’s Perspective 103
Big Picture Overview 103

5.1 Construct Sampling Distributions 105
5.2 Understand Desirable Properties of Simple Linear Regression Estimators 108
5.3 Understand the Simple Linear Regression Assumptions Required for OLS to be the Best Linear Unbiased Estimator 111
5.4 Understand How to Conduct Hypothesis Tests in Linear Regression Analysis 115
   Method 1: Construct Confidence Intervals around the Population Parameter 115
   Method 2: Compare Calculated Test Statistics with Predetermined Critical Values 119
   Method 3: Calculate and Compare p-values with Predetermined Levels of Significance 121
5.5 Conduct Hypothesis Tests for the Overall Statistical Significance of the Sample Regression Function 123
5.6 Conduct Hypothesis Tests for the Statistical Significance of the Slope Coefficient 125
   Calculate the Standard Error of the Estimated Slope Coefficient 125
   Test for the Individual Significance of the Slope Coefficient 126
5.7 Understand How to Read Regression Output in Excel for the Purpose of Hypothesis Testing 128
5.8 Construct Confidence Intervals around the Predicted Value of y 131
Additions to Our Empirical Research Toolkit 135
Our New Empirical Tools in Practice: Using What We Have Learned in This Chapter 135
Looking Ahead to Chapter 6 136
Problems 136
Exercises 139

Appendix 5A
Common Theoretical Probability Distributions 141

Chapter 6
Multiple Linear Regression Analysis 147

Chapter Objectives 147
A Student’s Perspective 147
Big Picture Overview 148
Data to Be Analyzed: Our MLB Position Player and International GDP Samples 149
Data Analyzed in the Text 149
Data Analyzed in the Excel Boxes 150

6.1 Understand the Goals of Multiple Linear Regression Analysis 151
6.2 Understand the “Holding All Other Independent Variables Constant” Condition in Multiple Linear Regression Analysis 153
6.3 Understand the Multiple Linear Regression Assumptions Required for OLS to Be Blue 155
6.4 Interpret Multiple Linear Regression Output in Excel 157
6.5 Assess the Goodness-of-Fit of the Sample Multiple Linear Regression Function 160
   The Coefficient of Determination (R^2) 160
   The Adjusted R^2 (R^2) 161
   Standard Error of the Sample Regression Function 162
6.6 Perform Hypothesis Tests for the Overall Significance of the Sample Regression Function 164
6.7 Perform Hypothesis Tests for the Individual Significance of a Slope Coefficient 166
6.8 Perform Hypothesis Tests for the Joint Significance of a Subset of Slope Coefficients 170
6.9 Perform the Chow Test for Structural Differences Between Two Subsets of Data 173
Additions to Our Empirical Research Toolkit 175
Our New Empirical Tools in Practice: Using What We Have Learned in This Chapter 176
Looking Ahead to Chapter 7 177
Problems 177
Exercises 181

Chapter 7
Qualitative Variables and Nonlinearities in Multiple Linear Regression Analysis 183
Chapter Objectives 183
A Student's Perspective 183
Big Picture Overview 184
7.1 Construct and Use Qualitative Independent Variables 184
  Binary Dummy Variables 185
  Categorical Variables 191
  Categorical Variables as a Series of Dummy Variables 195
7.2 Construct and Use Interaction Effects 199
7.3 Control for Nonlinear Relationships 204
  Quadratic Effects 204
  Interaction Effects between Two Quantitative Variables 210
7.4 Estimate Marginal Effects as Percent Changes and Elasticities 215
  The Log-Linear Model 215
  The Log-Log Model 217
7.5 Estimate a More Fully Specified Model 219
Additions to Our Empirical Research Toolkit 222
Our New Empirical Tools in Practice: Using What We Have Learned in This Chapter 222
Looking Ahead to Chapter 8 222
Problems 224
Exercises 228

Chapter 8
Model Selection in Multiple Linear Regression Analysis 230
Chapter Objectives 230
A Student's Perspective 230
Big Picture Overview 231
8.1 Understand the Problem Presented by Omitted Variable Bias 232
8.2 Understand the Problem Presented by Including an Irrelevant Variable 233
8.3 Understand the Problem Presented by Missing Data 235
8.4 Understand the Problem Presented by Outliers 238
8.5 Perform the Reset Test for the Inclusion of Higher-Order Polynomials 240
8.6 Perform the Davidson-MacKinnon Test for Choosing among Non-Nested Alternatives 243
8.7 Consider How to Implement the “Eye Test” to Judge the Sample Regression Function 246
8.8 Consider What It Means for a p-value to be Just Above a Given Significance Level 248
Additions to Our Empirical Research Toolkit 249
Our New Empirical Tools in Practice: Using What We Have Learned in This Chapter 249
Looking Ahead to Chapter 9 250
Problems 251
Exercises 253

PART THREE
VIOLATIONS OF ASSUMPTIONS 255

Chapter 9
Heteroskedasticity 255
Chapter Objectives 255
A Student's Perspective 255
Big Picture Overview 257
  Our Empirical Example: The Relationship between Income and Expenditures 258
Data to Be Analyzed: Our California Home Mortgage Application Sample 260
9.1 Understand Methods for Detecting Heteroskedasticity 262
  The Informal Method for Detecting Heteroskedasticity 262
  Formal Methods for Detecting Heteroskedasticity 264
9.2 Correct for Heteroskedasticity 272
  Weighted Least Squares 272
  A Different Assumed Form of Heteroskedasticity 275
  White's Heteroskedastic Consistent Standard Errors 275
Additions to Our Empirical Research Toolkit 278
Our New Empirical Tools in Practice: Using What We Have Learned in This Chapter 278
Looking Ahead to Chapter 10 280
Problems 280
Exercises 282
Chapter 10
Time-Series Analysis 284
Chapter Objectives 284
A Student’s Perspective 284
Big Picture Overview 284
Data to Be Analyzed: Our U.S. Houses Sold Data, 1986Q2–2005Q4 287
10.1 Understand the Assumptions Required for OLS to Be the Best Linear Unbiased Estimator for Time-Series Data 289
10.2 Understand Stationarity and Weak Dependence 290
Stationarity in Time Series 290
Weakly Dependent Time Series 290
10.3 Estimate Static Time-Series Models 291
10.4 Estimate Distributed Lag Models 292
10.5 Understand and Account for Time Trends and Seasonality 294
Time Trends 295
Seasonality 300
10.6 Test for Structural Breaks in the Data 301
10.7 Understand the Problem Presented by Spurious Regression 304
10.8 Learn to Perform Forecasting 306
Additions to Our Empirical Research Toolkit 308
Our New Empirical Tools in Practice: Using What We Have Learned in This Chapter 309
Looking Ahead to Chapter 11 310
Problems 311
Exercises 311
Reference 312

Chapter 11
Autocorrelation 313
Chapter Objectives 313
A Student’s Perspective 313
Big Picture Overview 313
11.1 Understand the Autoregressive Structure of the Error Term 316
The AR(1) Process 316
The AR(2) Process 316
The AR(1,4) Process 316
11.2 Understand Methods for Detecting Autocorrelation 316
Informal Methods for Detecting Autocorrelation 317
Formal Methods for Detecting Autocorrelation 318
11.3 Understand How to Correct for Autocorrelation 325
The Cochrane-Orcutt Method for AR(1) Processes 325
The Prais-Winsten Method for AR(1) Processes 332
Newey-West Robust Standard Errors 335
11.4 Understand Unit Roots and Cointegration 336
Unit Roots 336
Cointegration 338
Additions to Our Empirical Research Toolkit 340
Our New Empirical Tools in Practice: Using What We Have Learned in This Chapter 340
Looking Ahead to Chapter 12 341
Problems 342
Exercises 343

PART 4
ADVANCED TOPICS IN ECONOMETRICS 345

Chapter 12
Limited Dependent Variable Analysis 345
Chapter Objectives 345
A Student’s Perspective 345
Big Picture Overview 346
Data to Be Analyzed: Our 2010 House Election Data 347
12.1 Estimate Models with Binary Dependent Variables 349
The Linear Probability Model 349
The Logit Model 351
The Probit Model 354
Comparing the Three Estimators 356
12.2 Estimate Models with Categorical Dependent Variables 358
A New Data Set: Analyzing Educational Attainment Using Our SIPP Education Data 359
The Multinomial Logit 361
The Multinomial Probit 364
The Ordered Probit 365
Additions to Our Empirical Research Toolkit 367
Our New Empirical Tools in Practice: Using What We Have Learned in This Chapter 368
Looking Ahead to Chapter 13 368
Problems 369
Exercises 370

Chapter 13
Panel Data 371
Chapter Objectives 371
A Student’s Perspective 371
Big Picture Overview 372
13.1 Understand the Nature of Panel Data 373
Data to Be Analyzed: Our NFL Team Value Panel 374
13.2 Employ Pooled Cross-Section Analysis 376
Pooled Cross-Section Analysis with Year Dummies 377
13.3 Estimate Panel Data Models 380
First-Differenced Data in a Two-Period Model 380
Fixed-Effects Panel Data Models 382
Random-Effects Panel Data Models 385
Additions to Our Empirical Research Toolkit 387
Chapter 14
Instrumental Variables for Simultaneous Equations, Endogenous Independent Variables, and Measurement Error 390

Chapter Objectives 390
A Student’s Perspective 390
Big Picture Overview 391
14.1 Use Two-Stage Least Squares to Identify Simultaneous Demand and Supply Equations 392
Data to Be Analyzed: Our U.S. Gasoline Sales Data 393
14.2 Use Two-Stage Least Squares to Correct for Endogeneity of an Independent Variable 399
Our Empirical Example: The Effect of a Doctor’s Advice to Reduce Drinking 400
Data to Be Analyzed: Our Doctor Advice Data 401
14.3 Use Two-Stage Least Squares to Correct for Measurement Error 405
Measurement Error in the Dependent Variable 406
Measurement Error in an Independent Variable 406
Our Empirical Example: Using a Spouse’s Responses to Control for Measurement Error in an Individual’s Self-Reported Drinking 407
Additions to Our Empirical Research Toolkit 410
Our New Empirical Tools in Practice: Using What We Have Learned in This Chapter 410
Looking Ahead to Chapter 15 411
Problems 411
Exercises 413

Chapter 15
Quantile Regression, Count Data, Sample Selection Bias, and Quasi-Experimental Methods 415

Chapter Objectives 415
A Student’s Perspective 415
Big Picture Overview 417
15.1 Estimate Quantile Regression 418
15.2 Estimate Models with Non-Negative Count Data 420
Our Empirical Example: Early-Career Publications by Economics PhDs 420
Data to Be Analyzed: Our Newly Minted Economics PhD Publication Data 421
The Poisson Model 423
The Negative Binomial Model 425
Choosing between the Poisson and the Negative Binomial Models 427
15.3 Control for Sample-Selection Bias 428
Data to Be Analyzed: Our CPS Salary Data 430
15.4 Use Quasi-Experimental Methods 433
Our Empirical Example: Changes in State Speed Limits 434
Data to Be Analyzed: Our State Traffic Fatality Data 434
Additions to Our Empirical Research Toolkit 438
Our New Empirical Tools in Practice: Using What We Have Learned in This Chapter 438
Looking Ahead to Chapter 16 439
Problems 439
Exercises 439

Chapter 16
How to Conduct and Write Up an Empirical Research Project 441

Chapter Objectives 441
A Student’s Perspective 441
Big Picture Overview 441
16.1 General Approach to Conducting an Empirical Research Project 442
Collecting Data for the Dependent Variables 445
Collecting Data for the Independent Variables 448
16.2 General Approach to Writing Up an Empirical Research Project 457
16.3 An Example Write-Up of Our Movie Box-Office Project 461
Lights, Camera, Ticket Sales: An Analysis of the Determinants of Domestic Box-Office Gross 461
1. Introduction 461
2. Data Description 462
3. Empirical Results 463
4. Conclusion 465
References 466

Appendix A
Data Collection 469

Appendix B
Stata Commands 493

Appendix C
Statistical Tables 515

Index 519