EIGHTH EDITION

ECONOMETRIC ANALYSIS

William H. Greene

Percentiles of the Chi-Squared Distribution. Table Entry Is c Such That $\text{Prob}[\chi_n^2 \leq c] = P$

n	*.005*	*.010*	*.025*	*.050*	*.100*	*.250*	*.500*	*.750*	*.900*	*.950*	*.975*	*.990*	*.995*
1	.00004	.0002	.001	.004	.02	.10	.45	1.32	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.31	46.93
30	13.79	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
35	17.19	18.51	20.57	22.47	24.80	29.05	34.34	40.22	46.06	49.80	53.20	57.34	60.27
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.81	55.76	59.34	63.69	66.77
45	24.31	25.90	28.37	30.61	33.35	38.29	44.34	50.98	57.51	61.66	65.41	69.96	73.17
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33	56.33	63.17	67.50	71.42	76.15	79.49

EIGHTH EDITION

ECONOMETRIC ANALYSIS

William H. Greene

The Stern School of Business

New York University

Pearson

New York, NY

For Margaret and Richard Greene

Vice President, Business Publishing: Donna Battista
Director of Portfolio Management: Adrienne D' Ambrosio
Director, Courseware Portfolio Management: Ashley Dodge
Senior Sponsoring Editor: Neeraj Bhalla
Editorial Assistant: Courtney Paganelli
Vice President, Product Marketing: Roxanne McCarley
Director of Strategic Marketing: Brad Parkins
Strategic Marketing Manager: Deborah Strickland
Product Marketer: Tricia Murphy
Field Marketing Manager: Ramona Elmer
Product Marketing Assistant: Jessica Quazza
Vice President, Production and Digital Studio, Arts and Business: Etain O'Dea
Director of Production, Business: Jeff Holcomb
Managing Producer, Business: Alison Kalil
Content Producer: Sugandh Juneja
Operations Specialist: Carol Melville
Creative Director: Blair Brown
Manager, Learning Tools: Brian Surette
Content Developer, Learning Tools: Lindsey Sloan
Managing Producer, Digital Studio, Arts and Business: Diane Lombardo
Digital Studio Producer: Melissa Honig
Digital Studio Producer: Alana Coles
Digital Content Team Lead: Noel Lotz
Digital Content Project Lead: Courtney Kamauf
Full-Service Project Management and Composition: SPi Global
Interior Design: SPi Global
Cover Design: SPi Global
Cover Art: Jim Lozouski/Shutterstock
Printer/Binder: RRD Crawfordsville
Cover Printer: Phoenix/Hagerstown

Library of Congress Cataloging-in-Publication Data on File

1 17

Pearson

ISBN 10: 0-13-446136-3
ISBN 13: 978-0-13-446136-6

BRIEF CONTENTS

Examples and Applications
Preface

Part I The Linear Regression Model
Chapter 1 Econometrics 1
Chapter 2 The Linear Regression Model 12
Chapter 3 Least Squares Regression 28
Chapter 4 Estimating the Regression Model by Least Squares 54
Chapter 5 Hypothesis Tests and Model Selection 113
Chapter 6 Functional Form, Difference in Differences, and Structural Change 153
Chapter 7 Nonlinear, Semiparametric, and Nonparametric Regression Models 202
Chapter 8 Endogeneity and Instrumental Variable Estimation 242

Part II Generalized Regression Model and Equation Systems
Chapter 9 The Generalized Regression Model and Heteroscedasticity 297
Chapter 10 Systems of Regression Equations 326
Chapter 11 Models for Panel Data 373

Part III Estimation Methodology
Chapter 12 Estimation Frameworks in Econometrics 465
Chapter 13 Minimum Distance Estimation and the Generalized Method of Moments 488
Chapter 14 Maximum Likelihood Estimation 537
Chapter 15 Simulation-Based Estimation and Inference and Random Parameter Models 641
Chapter 16 Bayesian Estimation and Inference 694

Part IV Cross Sections, Panel Data, and Microeconometrics
Chapter 17 Binary Outcomes and Discrete Choices 725

Chapter 18 Multinomial Choices and Event Counts 826
Chapter 19 Limited Dependent Variables—Truncation, Censoring, and Sample Selection 918

Part V Time Series and Macroeconometrics

Chapter 20 Serial Correlation 981
Chapter 21 Nonstationary Data 1022

References 1054

Index 1098

Part VI Online Appendices

Appendix A Matrix Algebra A-1
Appendix B Probability and Distribution Theory B-1
Appendix C Estimation and Inference C-1
Appendix D Large-Sample Distribution Theory D-1
Appendix E Computation and Optimization E-1
Appendix F Data Sets Used in Applications F-1

Contents

Examples and Applications xxiv

Preface xxxv

Part I The Linear Regression Model

CHAPTER 1 Econometrics 1

1.1 Introduction 1
1.2 The Paradigm of Econometrics 1
1.3 The Practice of Econometrics 3
1.4 Microeconometrics and Macroeconometrics 4
1.5 Econometric Modeling 5
1.6 Plan of the Book 8
1.7 Preliminaries 9
1.7.1 Numerical Examples 9
1.7.2 Software and Replication 10
1.7.3 Notational Conventions 10

CHAPTER 2 The Linear Regression Model 12

2.1 Introduction 12
2.2 The Linear Regression Model 13
2.3 Assumptions of the Linear Regression Model 16
2.3.1 Linearity of the Regression Model 17
2.3.2 Full Rank 20
2.3.3 Regression 22
2.3.4 Homoscedastic and Nonautocorrelated Disturbances 23
2.3.5 Data Generating Process for the Regressors 25
2.3.6 Normality 25
2.3.7 Independence and Exogeneity 26
2.4 Summary and Conclusions 27

CHAPTER 3 Least Squares Regression 28

3.1 Introduction 28
3.2 Least Squares Regression 28

3.2.1 The Least Squares Coefficient Vector 29
3.2.2 Application: An Investment Equation 30
3.2.3 Algebraic Aspects of the Least Squares Solution 33
3.2.4 Projection 33
3.3 Partitioned Regression and Partial Regression 35
3.4 Partial Regression and Partial Correlation Coefficients 38
3.5 Goodness of Fit and the Analysis of Variance 41
3.5.1 The Adjusted R-Squared and a Measure of Fit 44
3.5.2 R-Squared and the Constant Term in the Model 47
3.5.3 Comparing Models 48
3.6 Linearly Transformed Regression 48
3.7 Summary and Conclusions 49

CHAPTER 4 Estimating the Regression Model by Least Squares 54
4.1 Introduction 54
4.2 Motivating Least Squares 55
4.2.1 Population Orthogonality Conditions 55
4.2.2 Minimum Mean Squared Error Predictor 56
4.2.3 Minimum Variance Linear Unbiased Estimation 57
4.3 Statistical Properties of the Least Squares Estimator 57
4.3.1 Unbiased Estimation 59
4.3.2 Omitted Variable Bias 59
4.3.3 Inclusion of Irrelevant Variables 61
4.3.4 Variance of the Least Squares Estimator 61
4.3.5 The Gauss–Markov Theorem 62
4.3.6 The Normality Assumption 63
4.4 Asymptotic Properties of the Least Squares Estimator 63
4.4.1 Consistency of the Least Squares Estimator of β 63
4.4.2 The Estimator of Asy. Var[b] 65
4.4.3 Asymptotic Normality of the Least Squares Estimator 66
4.4.4 Asymptotic Efficiency 67
4.4.5 Linear Projections 70
4.5 Robust Estimation and Inference 73
4.5.1 Consistency of the Least Squares Estimator 74
4.5.2 A Heteroscedasticity Robust Covariance Matrix for Least Squares 74
4.5.3 Robustness to Clustering 75
4.5.4 Bootstrapped Standard Errors with Clustered Data 77
4.6 Asymptotic Distribution of a Function of **b**: The Delta Method 78
4.7 Interval Estimation 81
4.7.1 Forming a Confidence Interval for a Coefficient 81
4.7.2 Confidence Interval for a Linear Combination of Coefficients: the Oaxaca Decomposition 83

4.8 Prediction and Forecasting 86
4.8.1 Prediction Intervals 86
4.8.2 Predicting y when the Regression Model Describes Log y 87
4.8.3 Prediction Interval for y when the Regression Model Describes Log y 88
4.8.4 Forecasting 92
4.9 Data Problems 93
4.9.1 Multicollinearity 94
4.9.2 Principal Components 97
4.9.3 Missing Values and Data Imputation 98
4.9.4 Measurement Error 102
4.9.5 Outliers and Influential Observations 104
4.10 Summary and Conclusions 107

CHAPTER 5 Hypothesis Tests and Model Selection 113
5.1 Introduction 113
5.2 Hypothesis Testing Methodology 113
5.2.1 Restrictions and Hypotheses 114
5.2.2 Nested Models 115
5.2.3 Testing Procedures 116
5.2.4 Size, Power, and Consistency of a Test 116
5.2.5 A Methodological Dilemma: Bayesian Versus Classical Testing 117
5.3 Three Approaches to Testing Hypotheses 117
5.3.1 Wald Tests Based on the Distance Measure 120
5.3.1.a Testing a Hypothesis About a Coefficient 120
5.3.1.b The F Statistic 123
5.3.2 Tests Based on the Fit of the Regression 126
5.3.2.a The Restricted Least Squares Estimator 126
5.3.2.b The Loss of Fit from Restricted Least Squares 127
5.3.2.c Testing the Significance of the Regression 129
5.3.2.d Solving Out the Restrictions and a Caution about R^2 *129*
5.3.3 Lagrange Multiplier Tests 130
5.4 Large-Sample Tests and Robust Inference 133
5.5 Testing Nonlinear Restrictions 136
5.6 Choosing Between Nonnested Models 138
5.6.1 Testing Nonnested Hypotheses 139
5.6.2 An Encompassing Model 140
5.6.3 Comprehensive Approach—The J Test 140
5.7 A Specification Test 141
5.8 Model Building—A General to Simple Strategy 143
5.8.1 Model Selection Criteria 143
5.8.2 Model Selection 144

5.8.3 Classical Model Selection 145
5.8.4 Bayesian Model Averaging 145
5.9 Summary and Conclusions 147

CHAPTER 6 Functional Form, Difference in Differences, and Structural Change 153

6.1 Introduction 153
6.2 Using Binary Variables 153
6.2.1 Binary Variables in Regression 153
6.2.2 Several Categories 157
6.2.3 Modeling Individual Heterogeneity 158
6.2.4 Sets of Categories 162
6.2.5 Threshold Effects and Categorical Variables 163
6.2.6 Transition Tables 164
6.3 Difference in Differences Regression 167
6.3.1 Treatment Effects 167
6.3.2 Examining the Effects of Discrete Policy Changes 172
6.4 Using Regression Kinks and Discontinuities to Analyze Social Policy 176
6.4.1 Regression Kinked Design 176
6.4.2 Regression Discontinuity Design 179
6.5 Nonlinearity in the Variables 183
6.5.1 Functional Forms 183
6.5.2 Interaction Effects 185
6.5.3 Identifying Nonlinearity 186
6.5.4 Intrinsically Linear Models 188
6.6 Structural Break and Parameter Variation 191
6.6.1 Different Parameter Vectors 191
6.6.2 Robust Tests of Structural Break with Unequal Variances 193
6.6.3 Pooling Regressions 195
6.7 Summary And Conclusions 197

CHAPTER 7 Nonlinear, Semiparametric, and Nonparametric Regression Models 202

7.1 Introduction 202
7.2 Nonlinear Regression Models 203
7.2.1 Assumptions of the Nonlinear Regression Model 203
7.2.2 The Nonlinear Least Squares Estimator 205
7.2.3 Large-Sample Properties of the Nonlinear Least Squares Estimator 207
7.2.4 Robust Covariance Matrix Estimation 210
7.2.5 Hypothesis Testing and Parametric Restrictions 211

7.2.6 Applications 212
7.2.7 Loglinear Models 215
7.2.8 Computing the Nonlinear Least Squares Estimator 222
7.3 Median and Quantile Regression 225
7.3.1 Least Absolute Deviations Estimation 226
7.3.2 Quantile Regression Models 228
7.4 Partially Linear Regression 234
7.5 Nonparametric Regression 235
7.6 Summary and Conclusions 238

CHAPTER 8 Endogeneity and Instrumental Variable Estimation 242
8.1 Introduction 242
8.2 Assumptions of the Extended Model 246
8.3 Instrumental Variables Estimation 248
8.3.1 Least Squares 248
8.3.2 The Instrumental Variables Estimator 249
8.3.3 Estimating the Asymptotic Covariance Matrix 250
8.3.4 Motivating the Instrumental Variables Estimator 251
8.4 Two-Stage Least Squares, Control Functions, and Limited Information Maximum Likelihood 256
8.4.1 Two-Stage Least Squares 257
8.4.2 A Control Function Approach 259
8.4.3 Limited Information Maximum Likelihood 261
8.5 Endogenous Dummy Variables: Estimating Treatment Effects 262
8.5.1 Regression Analysis of Treatment Effects 266
8.5.2 Instrumental Variables 267
8.5.3 A Control Function Estimator 269
8.5.4 Propensity Score Matching 270
8.6 Hypothesis Tests 274
8.6.1 Testing Restrictions 274
8.6.2 Specification Tests 275
8.6.3 Testing for Endogeneity: The Hausman and Wu Specification Tests 276
8.6.4 A Test for Overidentification 277
8.7 Weak Instruments and LIML 279
8.8 Measurement Error 281
8.8.1 Least Squares Attenuation 282
8.8.2 Instrumental Variables Estimation 284
8.8.3 Proxy Variables 285
8.9 Nonlinear Instrumental Variables Estimation 288
8.10 Natural Experiments and the Search for Causal Effects 291
8.11 Summary and Conclusions 295

Part II Generalized Regression Model and Equation Systems

CHAPTER 9 The Generalized Regression Model and Heteroscedasticity 297

9.1 Introduction 297
9.2 Robust Least Squares Estimation and Inference 298
9.3 Properties of Least Squares and Instrumental Variables 301
9.3.1 Finite-Sample Properties of Least Squares 301
9.3.2 Asymptotic Properties of Least Squares 302
9.3.3 Heteroscedasticity and Var[**b**|**X**] *304*
9.3.4 Instrumental Variable Estimation 305
9.4 Efficient Estimation by Generalized Least Squares 306
9.4.1 Generalized Least Squares (GLS) 306
9.4.2 Feasible Generalized Least Squares (FGLS) 309
9.5 Heteroscedasticity and Weighted Least Squares 310
9.5.1 Weighted Least Squares 311
9.5.2 Weighted Least Squares with Known **Ω** *311*
9.5.3 Estimation When **Ω** *Contains Unknown Parameters 312*
9.6 Testing for Heteroscedasticity 313
9.6.1 White's General Test 314
9.6.2 The Lagrange Multiplier Test 314
9.7 Two Applications 315
9.7.1 Multiplicative Heteroscedasticity 315
9.7.2 Groupwise Heteroscedasticity 317
9.8 Summary and Conclusions 320

CHAPTER 10 Systems of Regression Equations 326

10.1 Introduction 326
10.2 The Seemingly Unrelated Regressions Model 328
10.2.1 Ordinary Least Squares And Robust Inference 330
10.2.2 Generalized Least Squares 332
10.2.3 Feasible Generalized Least Squares 333
10.2.4 Testing Hypotheses 334
10.2.5 The Pooled Model 336
10.3 Systems of Demand Equations: Singular Systems 339
10.3.1 Cobb–Douglas Cost Function 339
10.3.2 Flexible Functional Forms: The Translog Cost Function 342
10.4 Simultaneous Equations Models 346
10.4.1 Systems of Equations 347
10.4.2 A General Notation for Linear Simultaneous Equations Models 350
10.4.3 The Identification Problem 353
10.4.4 Single Equation Estimation and Inference 358
10.4.5 System Methods of Estimation 362
10.5 Summary and Conclusions 365

CHAPTER 11 Models for Panel Data 373

11.1 Introduction 373

11.2 Panel Data Modeling 374

11.2.1 General Modeling Framework for Analyzing Panel Data 375
11.2.2 Model Structures 376
11.2.3 Extensions 377
11.2.4 Balanced and Unbalanced Panels 377
11.2.5 Attrition and Unbalanced Panels 378
11.2.6 Well-Behaved Panel Data 382

11.3 The Pooled Regression Model 383

11.3.1 Least Squares Estimation of the Pooled Model 383
11.3.2 Robust Covariance Matrix Estimation and Bootstrapping 384
11.3.3 Clustering and Stratification 386
11.3.4 Robust Estimation Using Group Means 388
11.3.5 Estimation with First Differences 389
11.3.6 The Within- and Between-Groups Estimators 390

11.4 The Fixed Effects Model 393

11.4.1 Least Squares Estimation 393
11.4.2 A Robust Covariance Matrix for $\mathbf{b}_{LSDV}$ *396*
11.4.3 Testing the Significance of the Group Effects 397
11.4.4 Fixed Time and Group Effects 398
11.4.5 Reinterpreting the Within Estimator: Instrumental Variables and Control Functions 399
11.4.6 Parameter Heterogeneity 401

11.5 Random Effects 404

11.5.1 Least Squares Estimation 405
11.5.2 Generalized Least Squares 407
11.5.3 Feasible Generalized Least Squares Estimation of the Random Effects Model when $\mathbf{\Sigma}$ *is Unknown 408*
11.5.4 Robust Inference and Feasible Generalized Least Squares 409
11.5.5 Testing for Random Effects 410
11.5.6 Hausman's Specification Test for the Random Effects Model 414
11.5.7 Extending the Unobserved Effects Model: Mundlak's Approach 415
11.5.8 Extending the Random and Fixed Effects Models: Chamberlain's Approach 416

11.6 Nonspherical Disturbances and Robust Covariance Matrix Estimation 421

11.6.1 Heteroscedasticity in the Random Effects Model 421
11.6.2 Autocorrelation in Panel Data Models 422

11.7 Spatial Autocorrelation 422

11.8 Endogeneity 427

11.8.1 Instrumental Variable Estimation 427

11.8.2 Hausman and Taylor's Instrumental Variables Estimator 429

11.8.3 Consistent Estimation of Dynamic Panel Data Models: Anderson and Hsiao's Iv Estimator 433

11.8.4 Efficient Estimation of Dynamic Panel Data Models:The Arellano/Bond Estimators 436

11.8.5 Nonstationary Data and Panel Data Models 445

11.9 Nonlinear Regression with Panel Data 446

11.9.1 A Robust Covariance Matrix for Nonlinear Least Squares 446

11.9.2 Fixed Effects in Nonlinear Regression Models 447

11.9.3 Random Effects 449

11.10 Parameter Heterogeneity 450

11.10.1 A Random Coefficients Model 450

11.10.2 A Hierarchical Linear Model 453

11.10.3 Parameter Heterogeneity and Dynamic Panel Data Models 455

11.11 Summary and Conclusions 459

Part III Estimation Methodology

CHAPTER 12 Estimation Frameworks in Econometrics 465

12.1 Introduction 465

12.2 Parametric Estimation and Inference 467

12.2.1 Classical Likelihood-Based Estimation 467

12.2.2 Modeling Joint Distributions with Copula Functions 469

12.3 Semiparametric Estimation 472

12.3.1 Gmm Estimation in Econometrics 473

12.3.2 Maximum Empirical Likelihood Estimation 473

12.3.3 Least Absolute Deviations Estimation and Quantile Regression 475

12.3.4 Kernel Density Methods 475

12.3.5 Comparing Parametric and Semiparametric Analyses 476

12.4 Nonparametric Estimation 478

12.4.1 Kernel Density Estimation 478

12.5 Properties of Estimators 481

12.5.1 Statistical Properties of Estimators 481

12.5.2 Extremum Estimators 482

12.5.3 Assumptions for Asymptotic Properties of Extremum Estimators 483

12.5.4 Asymptotic Properties of Estimators 485

12.5.5 Testing Hypotheses 487

12.6 Summary and Conclusions 487

CHAPTER 13 Minimum Distance Estimation and the Generalized Method of Moments 488

13.1 Introduction 488

13.2 Consistent Estimation: The Method of Moments 489

13.2.1 Random Sampling and Estimating the Parameters of Distributions 490

13.2.2 Asymptotic Properties of the Method of Moments Estimator 493

13.2.3 Summary—The Method of Moments 496

13.3 Minimum Distance Estimation 496

13.4 The Generalized Method of Moments (GMM) Estimator 500

13.4.1 Estimation Based on Orthogonality Conditions 501

13.4.2 Generalizing the Method of Moments 502

13.4.3 Properties of the GMM Estimator 506

13.5 Testing Hypotheses in the GMM Framework 510

13.5.1 Testing the Validity of the Moment Restrictions 510

13.5.2 Gmm Wald Counterparts to the WALD, LM, and LR Tests 512

13.6 Gmm Estimation of Econometric Models 513

13.6.1 Single-Equation Linear Models 514

13.6.2 Single-Equation Nonlinear Models 519

13.6.3 Seemingly Unrelated Regression Equations 522

13.6.4 Gmm Estimation of Dynamic Panel Data Models 523

13.7 Summary and Conclusions 534

CHAPTER 14 Maximum Likelihood Estimation 537

14.1 Introduction 537

14.2 The Likelihood Function and Identification of the Parameters 537

14.3 Efficient Estimation: The Principle of Maximum Likelihood 539

14.4 Properties of Maximum Likelihood Estimators 541

14.4.1 Regularity Conditions 542

14.4.2 Properties of Regular Densities 543

14.4.3 The Likelihood Equation 544

14.4.4 The Information Matrix Equality 545

14.4.5 Asymptotic Properties of the Maximum Likelihood Estimator 545

14.4.5.a Consistency 545

14.4.5.b Asymptotic Normality 547

14.4.5.c Asymptotic Efficiency 548

14.4.5.d Invariance 548

14.4.5.e Conclusion 549

14.4.6 Estimating the Asymptotic Variance of the Maximum Likelihood Estimator 549

14.5 Conditional Likelihoods and Econometric Models 551

14.6 Hypothesis and Specification Tests and Fit Measures 552
14.6.1 The Likelihood Ratio Test 554
14.6.2 The Wald Test 555
14.6.3 The Lagrange Multiplier Test 557
14.6.4 An Application of the Likelihood-Based Test Procedures 558
14.6.5 Comparing Models and Computing Model Fit 560
14.6.6 Vuong's Test and the Kullback–Leibler Information Criterion 562
14.7 Two-Step Maximum Likelihood Estimation 564
14.8 Pseudo-Maximum Likelihood Estimation and Robust Asymptotic Covariance Matrices 570
14.8.1 A Robust Covariance Matrix Estimator for the MLE 570
14.8.2 Cluster Estimators 573
14.9 Maximum Likelihood Estimation of Linear Regression Models 576
14.9.1 Linear Regression Model with Normally Distributed Disturbances 576
14.9.2 Some Linear Models with Nonnormal Disturbances 578
14.9.3 Hypothesis Tests for Regression Models 580
14.10 The Generalized Regression Model 585
14.10.1 GLS With Known **Ω** *585*
14.10.2 Iterated Feasible GLS With Estimated Ω 586
14.10.3 Multiplicative Heteroscedasticity 586
14.10.4 The Method of Scoring 587
14.11 Nonlinear Regression Models and Quasi-Maximum Likelihood Estimation 591
14.11.1 Maximum Likelihood Estimation 592
14.11.2 Quasi-Maximum Likelihood Estimation 595
14.12 Systems of Regression Equations 600
14.12.1 The Pooled Model 600
14.12.2 The SUR Model 601
14.13 Simultaneous Equations Models 604
14.14 Panel Data Applications 605
14.14.1 ML Estimation of the Linear Random Effects Model 606
14.14.2 Nested Random Effects 609
14.14.3 Clustering Over More than One Level 612
14.14.4 Random Effects in Nonlinear Models: MLE Using Quadrature 613
14.14.5 Fixed Effects in Nonlinear Models: The Incidental Parameters Problem 617
14.15 Latent Class and Finite Mixture Models 622
14.15.1 A Finite Mixture Model 622
14.15.2 Modeling the Class Probabilities 624

14.15.3 Latent Class Regression Models 625
14.15.4 Predicting Class Membership and β_i *626*
14.15.5 Determining the Number of Classes 628
14.15.6 A Panel Data Application 628
14.15.7 A Semiparametric Random Effects Model 633
14.16 Summary and Conclusions 635

CHAPTER 15 Simulation-Based Estimation and Inference and Random Parameter Models 641

15.1 Introduction 641
15.2 Random Number Generation 643
15.2.1 Generating Pseudo-Random Numbers 643
15.2.2 Sampling from a Standard Uniform Population 644
15.2.3 Sampling from Continuous Distributions 645
15.2.4 Sampling from a Multivariate Normal Population 646
15.2.5 Sampling from Discrete Populations 646
15.3 Simulation-Based Statistical Inference: The Method of Krinsky and Robb 647
15.4 Bootstrapping Standard Errors and Confidence Intervals 650
15.4.1 Types of Bootstraps 651
15.4.2 Bias Reduction with Bootstrap Estimators 651
15.4.3 Bootstrapping Confidence Intervals 652
15.4.4 Bootstrapping with Panel Data: The Block Bootstrap 652
15.5 Monte Carlo Studies 653
15.5.1 A Monte Carlo Study: Behavior of a Test Statistic 655
15.5.2 A Monte Carlo Study: The Incidental Parameters Problem 656
15.6 Simulation-Based Estimation 660
15.6.1 Random Effects in a Nonlinear Model 661
15.6.2 Monte Carlo Integration 662
15.6.2a Halton Sequences and Random Draws for Simulation-Based Integration 664
15.6.2.b Computing Multivariate Normal Probabilities Using the GHK Simulator 666
15.6.3 Simulation-Based Estimation of Random Effects Models 668
15.7 A Random Parameters Linear Regression Model 673
15.8 Hierarchical Linear Models 678
15.9 Nonlinear Random Parameter Models 680
15.10 Individual Parameter Estimates 681
15.11 Mixed Models and Latent Class Models 689
15.12 Summary and Conclusions 692

CHAPTER 16 Bayesian Estimation and Inference 694

16.1 Introduction 694
16.2 Bayes' Theorem and the Posterior Density 695
16.3 Bayesian Analysis of the Classical Regression Model 697
16.3.1 Analysis with a Noninformative Prior 698
16.3.2 Estimation with an Informative Prior Density 700
16.4 Bayesian Inference 703
16.4.1 Point Estimation 703
16.4.2 Interval Estimation 704
16.4.3 Hypothesis Testing 705
16.4.4 Large-Sample Results 707
16.5 Posterior Distributions and the Gibbs Sampler 707
16.6 Application: Binomial Probit Model 710
16.7 Panel Data Application: Individual Effects Models 713
16.8 Hierarchical Bayes Estimation of a Random Parameters Model 715
16.9 Summary and Conclusions 721

Part IV Cross Sections, Panel Data, and Microeconometrics

CHAPTER 17 Binary Outcomes and Discrete Choices 725

17.1 Introduction 725
17.2 Models for Binary Outcomes 728
17.2.1 Random Utility 729
17.2.2 The Latent Regression Model 730
17.2.3 Functional Form and Probability 731
17.2.4 Partial Effects in Binary Choice Models 734
17.2.5 Odds Ratios in Logit Models 736
17.2.6 The Linear Probability Model 740
17.3 Estimation and Inference for Binary Choice Models 742
17.3.1 Robust Covariance Matrix Estimation 744
17.3.2 Hypothesis Tests 746
17.3.3 Inference for Partial Effects 749
17.3.3.a The Delta Method 749
17.3.3.b An Adjustment to the Delta Method 751
17.3.3.c The Method of Krinsky and Robb 752
17.3.3.d Bootstrapping 752
17.3.4 Interaction Effects 755
17.4 Measuring Goodness of Fit for Binary Choice Models 757
17.4.1 Fit Measures Based on the Fitting Criterion 757
17.4.2 Fit Measures Based on Predicted Values 758
17.4.3 Summary of Fit Measures 760
17.5 Specification Analysis 762
17.5.1 Omitted Variables 763

17.5.2 Heteroscedasticity 764
17.5.3 Distributional Assumptions 766
17.5.4 Choice-Based Sampling 768
17.6 Treatment Effects and Endogenous Variables in Binary Choice Models 769
17.6.1 Endogenous Treatment Effect 770
17.6.2 Endogenous Continuous Variable 773
17.6.2.a IV and GMM Estimation 773
17.6.2.b Partial ML Estimation 774
17.6.2.c Full Information Maximum Likelihood Estimation 774
17.6.2.d Residual Inclusion and Control Functions 775
17.6.2.e A Control Function Estimator 775
17.6.3 Endogenous Sampling 777
17.7 Panel Data Models 780
17.7.1 The Pooled Estimator 781
17.7.2 Random Effects 782
17.7.3 Fixed Effects 785
17.7.3.a A Conditional Fixed Effects Estimator 787
17.7.3.b Mundlak's Approach, Variable Addition, and Bias Reduction 792
17.7.4 Dynamic Binary Choice Models 794
17.7.5 A Semiparametric Model for Individual Heterogeneity 797
17.7.6 Modeling Parameter Heterogeneity 798
17.7.7 Nonresponse, Attrition, and Inverse Probability Weighting 801
17.9 Spatial Binary Choice Models 804
17.9 The Bivariate Probit Model 807
17.9.1 Maximum Likelihood Estimation 808
17.9.2 Testing for Zero Correlation 811
17.9.3 Partial Effects 811
17.9.4 A Panel Data Model for Bivariate Binary Response 814
17.9.5 A Recursive Bivariate Probit Model 815
17.10 A Multivariate Probit Model 819
17.11 Summary and Conclusions 822

CHAPTER 18 Multinomial Choices and Event Counts 826
18.1 Introduction 826
18.2 Models for Unordered Multiple Choices 827
18.2.1 Random Utility Basis of the Multinomial Logit Model 827
18.2.2 The Multinomial Logit Model 829
18.2.3 The Conditional Logit Model 833
18.2.4 The Independence from Irrelevant Alternatives Assumption 834
18.2.5 Alternative Choice Models 835
18.2.5.a Heteroscedastic Extreme Value Model 836

18.2.5.b Multinomial Probit Model 836
18.2.5.c The Nested Logit Model 837
18.2.6 Modeling Heterogeneity 845
18.2.6.a The Mixed Logit Model 845
18.2.6.b A Generalized Mixed Logit Model 846
18.2.6.c Latent Classes 849
18.2.6.d Attribute Nonattendance 851
18.2.7 Estimating Willingness to Pay 853
18.2.8 Panel Data and Stated Choice Experiments 856
18.2.8.a The Mixed Logit Model 857
18.2.8.b Random Effects and the Nested Logit Model 858
18.2.8.c A Fixed Effects Multinomial Logit Model 859
18.2.9 Aggregate Market Share Data—The BLP Random Parameters Model 863
18.3 Random Utility Models for Ordered Choices 865
18.3.1 The Ordered Probit Model 869
18.3.2.A Specification Test for the Ordered Choice Model 872
18.3.3 Bivariate Ordered Probit Models 873
18.3.4 Panel Data Applications 875
18.3.4.a Ordered Probit Models with Fixed Effects 875
18.3.4.b Ordered Probit Models with Random Effects 877
18.3.5 Extensions of the Ordered Probit Model 881
18.3.5.a Threshold Models—Generalized Ordered Choice Models 881
18.3.5.b Thresholds and Heterogeneity—Anchoring Vignettes 883
18.4 Models for Counts of Events 884
18.4.1 The Poisson Regression Model 885
18.4.2 Measuring Goodness of Fit 887
18.4.3 Testing for Overdispersion 888
18.4.4 Heterogeneity and the Negative Binomial Regression Model 889
18.4.5 Functional Forms for Count Data Models 890
18.4.6 Truncation and Censoring in Models for Counts 894
18.4.7 Panel Data Models 898
18.4.7.a Robust Covariance Matrices for Pooled Estimators 898
18.4.7.b Fixed Effects 900
18.4.7.c Random Effects 902
18.4.8 Two-Part Models: Zero-Inflation and Hurdle Models 905
18.4.9 Endogenous Variables and Endogenous Participation 910
18.5 Summary and Conclusions 914

CHAPTER 19 Limited Dependent Variables–Truncation, Censoring, and Sample Selection 918

19.1 Introduction 918

19.2 Truncation 918

19.2.1 Truncated Distributions 919
19.2.2 Moments of Truncated Distributions 920
19.2.3 The Truncated Regression Model 922
19.2.4 The Stochastic Frontier Model 924

19.3 Censored Data 930

19.3.1 The Censored Normal Distribution 931
19.3.2 The Censored Regression (Tobit) Model 933
19.3.3 Estimation 936
19.3.4 Two-Part Models and Corner Solutions 938
19.3.5 Specification Issues 944
19.3.5.a Endogenous Right-Hand-Side Variables 944
19.3.5.b Heteroscedasticity 945
19.3.5.c Nonnormality 947
19.3.6 Panel Data Applications 948

19.4 Sample Selection and Incidental Truncation 949

19.4.1 Incidental Truncation in a Bivariate Distribution 949
19.4.2 Regression in a Model of Selection 950
19.4.3 Two-Step and Maximum Likelihood Estimation 953
19.4.4 Sample Selection in Nonlinear Models 957
19.4.5 Panel Data Applications of Sample Selection Models 961
19.4.5.a Common Effects in Sample Selection Models 961
19.4.5.b Attrition 964

19.5 Models for Duration 965

19.5.1 Models for Duration Data 966
19.5.2 Duration Data 966
19.5.3 A Regression-Like Approach: Parametric Models of Duration 967
19.5.3.a Theoretical Background 967
19.5.3.b Models of the Hazard Function 968
19.5.3.c Maximum Likelihood Estimation 970
19.5.3.d Exogenous Variables 971
19.5.3.e Heterogeneity 972
19.5.4 Nonparametric and Semiparametric Approaches 973

19.6 Summary and Conclusions 976

Part V Time Series and Macroeconometrics

CHAPTER 20 Serial Correlation 981

20.1 Introduction 981

20.2 The Analysis of Time-Series Data 984

20.3 Disturbance Processes 987

20.3.1 Characteristics of Disturbance Processes 987
20.3.2 Ar(1) Disturbances 989

20.4 Some Asymptotic Results for Analyzing Time-Series Data 990

20.4.1 *Convergence of Moments—The Ergodic Theorem* 991
20.4.2 *Convergence to Normality—A Central Limit Theorem* 994
20.5 Least Squares Estimation 996
20.5.1 *Asymptotic Properties of Least Squares* 996
20.5.2 *Estimating the Variance of the Least Squares Estimator* 998
20.6 Gmm Estimation 999
20.7 Testing for Autocorrelation 1000
20.7.1 *Lagrange Multiplier Test* 1000
20.7.2 *Box And Pierce's Test and Ljung's Refinement* 1001
20.7.3 *The Durbin–Watson Test* 1001
20.7.4 *Testing in the Presence of a Lagged Dependent Variable* 1002
20.7.5 *Summary of Testing Procedures* 1002
20.8 Efficient Estimation when is Known 1003
20.9 Estimation when is **Ω** Unknown 1004
20.9.1 *Ar(1) Disturbances* 1004
20.9.2 *Application: Estimation of a Model with Autocorrelation* 1005
20.9.3 *Estimation with a Lagged Dependent Variable* 1007
20.10 Autoregressive Conditional Heteroscedasticity 1010
20.10.1 *The ARCH(1) Model* 1011
20.10.2 *ARCH(q), ARCH-In-Mean, and Generalized ARCH Models* 1012
20.10.3 *Maximum Likelihood Estimation of the GARCH Model* 1014
20.10.4 *Testing for GARCH Effects* 1017
20.10.5 *Pseudo–Maximum Likelihood Estimation* 1018
20.11 Summary and Conclusions 1019

CHAPTER 21 Nonstationary Data 1022
21.1 Introduction 1022
21.2 Nonstationary Processes and Unit Roots 1022
21.2.1 *The Lag and Difference Operators* 1022
21.2.2 *Integrated Processes and Differencing* 1023
21.2.3 *Random Walks, Trends, and Spurious Regressions* 1026
21.2.4 *Tests for Unit Roots in Economic Data* 1028
21.2.5 *The Dickey–Fuller Tests* 1029
21.2.6 *The KPSS Test of Stationarity* 1038
21.3 Cointegration 1039
21.3.1 *Common Trends* 1043
21.3.2 *Error Correction and Var Representations* 1044
21.3.3 *Testing for Cointegration* 1045
21.3.4 *Estimating Cointegration Relationships* 1048
21.3.5 *Application: German Money Demand* 1048
21.3.5.a *Cointegration Analysis and a Long-Run Theoretical Model* 1049

21.3.5.b Testing for Model Instability 1050

21.4 Nonstationary Panel Data 1051

21.5 Summary and Conclusions 1052

References 1054

Index 1098

Part VI Online Appendices

Appendix A Matrix Algebra A-1

A.1 Terminology A-1

A.2 Algebraic Manipulation of Matrices A-2

A.2.1 Equality of Matrices A-2
A.2.2 Transposition A-2
A.2.3 Vectorization A-3
A.2.4 Matrix Addition A-3
A.2.5 Vector Multiplication A-3
A.2.6 A Notation for Rows and Columns of a Matrix A-3
A.2.7 Matrix Multiplication and Scalar Multiplication A-4
A.2.8 Sums of Values A-5
A.2.9 A Useful Idempotent Matrix A-6

A.3 Geometry of Matrices A-8

A.3.1 Vector Spaces A-8
A.3.2 Linear Combinations of Vectors and Basis Vectors A-9
A.3.3 Linear Dependence A-11
A.3.4 Subspaces A-12
A.3.5 Rank of a Matrix A-12
A.3.6 Determinant of a Matrix A-15
A.3.7 A Least Squares Problem A-16

A.4 Solution of a System of Linear Equations A-19

A.4.1 Systems of Linear Equations A-19
A.4.2 Inverse Matrices A-19
A.4.3 Nonhomogeneous Systems of Equations A-21
A.4.4 Solving the Least Squares Problem A-21

A.5 Partitioned Matrices A-22

A.5.1 Addition and Multiplication of Partitioned Matrices A-22
A.5.2 Determinants of Partitioned Matrices A-23
A.5.3 Inverses of Partitioned Matrices A-23
A.5.4 Deviations From Means A-23
A.5.5 Kronecker Products A-24

A.6 Characteristic Roots And Vectors *A-24*

A.6.1 The Characteristic Equation A-25
A.6.2 Characteristic Vectors A-25
A.6.3 General Results for Characteristic Roots And Vectors A-26

A.6.4 Diagonalization and Spectral Decomposition of a Matrix A-26
A.6.5 Rank of a Matrix A-27
A.6.6 Condition Number of a Matrix A-28
A.6.7 Trace of a Matrix A-29
A.6.8 Determinant of a Matrix A-30
A.6.9 Powers of a Matrix A-30
A.6.10 Idempotent Matrices A-32
A.6.11 Factoring a Matrix: The Cholesky Decomposition A-32
A.6.12 Singular Value Decomposition A-33
A.6.13 Qr Decomposition A-33
A.6.14 The Generalized Inverse of a Matrix A-33
A.7 Quadratic Forms And Definite Matrices *A-34*
A.7.1 Nonnegative Definite Matrices A-35
A.7.2 Idempotent Quadratic Forms A-36
A.7.3 Comparing Matrices A-37
A.8 Calculus And Matrix Algebra 15 *A-37*
A.8.1 Differentiation and the Taylor Series A-37
A.8.2 Optimization A-41
A.8.3 Constrained Optimization A-43
A.8.4 Transformations A-45

Appendix B Probability and Distribution Theory B-1
B.1 Introduction B-1
B.2 Random Variables B-1
B.2.1 Probability Distributions B-2
B.2.2 Cumulative Distribution Function B-2
B.3 Expectations of a Random Variable B-3
B.4 Some Specific Probability Distributions B-6
B.4.1 The Normal and Skew Normal Distributions B-6
B.4.2 The Chi-Squared, t, and F Distributions B-8
B.4.3 Distributions with Large Degrees of Freedom B-11
B.4.4 Size Distributions: The Lognormal Distribution B-12
B.4.5 The Gamma and Exponential Distributions B-13
B.4.6 The Beta Distribution B-13
B.4.7 The Logistic Distribution B-14
B.4.8 The Wishart Distribution B-14
B.4.9 Discrete Random Variables B-15
B.5 The Distribution of a Function of a Random Variable B-15
B.6 Representations of a Probability Distribution B-18
B.7 Joint Distributions B-19
B.7.1 Marginal Distributions B-20
B.7.2 Expectations in a Joint Distribution B-20
B.7.3 Covariance and Correlation B-21

B.7.4 Distribution of a Function of Bivariate Random Variables B-22
B.8 Conditioning in a Bivariate Distribution B-23
B.8.1 Regression: The Conditional Mean B-24
B.8.2 Conditional Variance B-24
B.8.3 Relationships among Marginal and Conditional Moments B-24
B.8.4 The Analysis of Variance B-26
B.8.5 Linear Projection B-27
B.9 The Bivariate Normal Distribution B-28
B.10 Multivariate Distributions B-29
B.10.1 Moments B-29
B.10.2 Sets of Linear Functions B-30
B.10.3 Nonlinear Functions: The Delta Method B-31
B.11 The Multivariate Normal Distribution B-31
B.11.1 Marginal and Conditional Normal Distributions B-32
B.11.2 The Classical Normal Linear Regression Model B-33
B.11.3 Linear Functions of a Normal Vector B-33
B.11.4 Quadratic Forms in a Standard Normal Vector B-34
B.11.5 The F Distribution B-36
B.11.6 A Full Rank Quadratic Form B-36
B.11.7 Independence of a Linear and a Quadratic Form B-38

Appendix C Estimation and Inference C-1
C.1 Introduction C-1
C.2 Samples and Random Sampling C-1
C.3 Descriptive Statistics C-2
C.4 Statistics as Estimators—Sampling Distributions C-6
C.5 Point Estimation of Parameters C-9
C.5.1 Estimation in a Finite Sample C-9
C.5.2 Efficient Unbiased Estimation C-12
C.6 Interval Estimation C-14
C.7 Hypothesis Testing C-16
C.7.1 Classical Testing Procedures C-16
C.7.2 Tests Based on Confidence Intervals C-19
C.7.3 Specification Tests D-1

Appendix D Large-Sample Distribution Theory D-1
D.1 Introduction D-1
D.2 Large-Sample Distribution Theory 1 D-2
D.2.1 Convergence in Probability D-2
D.2.2 Other forms of Convergence and Laws of Large Numbers D-5
D.2.3 Convergence of Functions D-9
D.2.4 Convergence to a Random Variable D-10

D.2.5 Convergence in Distribution: Limiting Distributions D-11
D.2.6 Central Limit Theorems D-14
D.2.7 The Delta Method D-19
D.3 Asymptotic Distributions D-19
D.3.1 Asymptotic Distribution of a Nonlinear Function D-21
D.3.2 Asymptotic Expectations D-22
D.4 Sequences and the Order of a Sequence D-24

Appendix E Computation and Optimization E-1
E.1 Introduction E-1
E.2 Computation in Econometrics E-1
E.2.1 Computing Integrals E-2
E.2.2 The Standard Normal Cumulative Distribution Function E-2
E.2.3 The Gamma and Related Functions E-3
E.2.4 Approximating Integrals by Quadrature E-4
E.3 Optimization E-5
E.3.1 Algorithms E-7
E.3.2 Computing Derivatives E-7
E.3.3 Gradient Methods E-9
E.3.4 Aspects of Maximum Likelihood Estimation E-12
E.3.5 Optimization with Constraints E-14
E.3.6 Some Practical Considerations E-15
E.3.7 The EM Algorithm E-17
E.4 Examples E-19
E.4.1 Function of one Parameter E-19
E.4.2 Function of two Parameters: The Gamma Distribution E-20
E.4.3 A Concentrated Log-Likelihood Function E-21

Appendix F Data Sets Used in Applications F-1

EXAMPLES AND APPLICATIONS

CHAPTER 1 Econometrics 1

Example 1.1 Behavioral Models and the Nobel Laureates 2
Example 1.2 Keynes's Consumption Function 5

CHAPTER 2 The Linear Regression Model 12

Example 2.1 Keynes's Consumption Function 14
Example 2.2 Earnings and Education 15
Example 2.3 The U.S. Gasoline Market 19
Example 2.4 The Translog Model 19
Example 2.5 Short Rank 20
Example 2.6 An Inestimable Model 21
Example 2.7 Nonzero Conditional Mean of the Disturbances 22

CHAPTER 3 Least Squares Regression 28

Example 3.1 Partial Correlations 41
Example 3.2 Fit of a Consumption Function 44
Example 3.3 Analysis of Variance for the Investment Equation 44
Example 3.4 Art Appreciation 48

CHAPTER 4 Estimating the Regression Model by Least Squares 54

Example 4.1 The Sampling Distribution of a Least Squares Estimator 58
Example 4.2 Omitted Variable in a Demand Equation 59
Example 4.3 Least Squares Vs. Least Absolute Deviations—A Monte Carlo Study 68
Example 4.4 Linear Projection: A Sampling Experiment 72
Example 4.5 Robust Inference about the Art Market 76
Example 4.6 Clustering and Block Bootstrapping 78
Example 4.7 Nonlinear Functions of Parameters: The Delta Method 80
Example 4.8 Confidence Interval for the Income Elasticity of Demand for Gasoline 83
Example 4.9 Oaxaca Decomposition of Home Sale Prices 85
Example 4.10 Pricing Art 90
Example 4.11 Multicollinearity in the Longley Data 95
Example 4.12 Predicting Movie Success 97
Example 4.13 Imputation in the Survey of Consumer Finances 16 101

CHAPTER 5 Hypothesis Tests and Model Selection 113

Example 5.1 Art Appreciation 121
Example 5.2 Earnings Equation 122
Example 5.3 Restricted Investment Equation 124
Example 5.4 F Test for the Earnings Equation 129
Example 5.5 Production Functions 130
Example 5.6 A Long-Run Marginal Propensity to Consume 137
Example 5.7 J Test for a Consumption Function 141
Example 5.8 Size of a RESET Test 142
Example 5.9 Bayesian Averaging of Classical Estimates 147

CHAPTER 6 Functional Form, Difference in Differences, and Structural Change 153

Example 6.1 Dummy Variable in an Earnings Equation 154
Example 6.2 Value of a Signature 155
Example 6.3 Gender and Time Effects in a Log Wage Equation 156
Example 6.4 Genre Effects on Movie Box Office Receipts 158
Example 6.5 Sports Economics: Using Dummy Variables for Unobserved Heterogeneity 5 160
Example 6.6 Analysis of Covariance 162
Example 6.7 Education Thresholds in a Log Wage Equation 165
Example 6.8 SAT Scores 169
Example 6.9 A Natural Experiment: The Mariel Boatlift 169
Example 6.10 Effect of the Minimum Wage 170
Example 6.11 Difference in Differences Analysis of a Price Fixing Conspiracy 13 172
Example 6.12 Policy Analysis Using Kinked Regressions 178
Example 6.13 The Treatment Effect of Compulsory Schooling 180
Example 6.14 Interest Elasticity of Mortgage Demand 180
Example 6.15 Quadratic Regression 184
Example 6.16 Partial Effects in a Model with Interactions 186
Example 6.17 Functional Form for a Nonlinear Cost Function 187
Example 6.18 Intrinsically Linear Regression 189
Example 6.19 CES Production Function 190
Example 6.20 Structural Break in the Gasoline Market 192
Example 6.21 Sample Partitioning by Gender 194
Example 6.22 The World Health Report 194
Example 6.23 Pooling in a Log Wage Model 196

CHAPTER 7 Nonlinear, Semiparametric, and Nonparametric Regression Models 202

Example 7.1 CES Production Function 203
Example 7.2 Identification in a Translog Demand System 204
Example 7.3 First-Order Conditions for a Nonlinear Model 206
Example 7.4 Analysis of a Nonlinear Consumption Function 213

Example 7.5 The Box–Cox Transformation 214
Example 7.6 Interaction Effects in a Loglinear Model for Income 216
Example 7.7 Generalized Linear Models for the Distribution of Healthcare Costs 221
Example 7.8 Linearized Regression 223
Example 7.9 Nonlinear Least Squares 224
Example 7.10 LAD Estimation of a Cobb–Douglas Production Function 228
Example 7.11 Quantile Regression for Smoking Behavior 230
Example 7.12 Income Elasticity of Credit Card Expenditures 231
Example 7.13 Partially Linear Translog Cost Function 235
Example 7.14 A Nonparametric Average Cost Function 237

CHAPTER 8 Endogeneity and Instrumental Variable Estimation 242

Example 8.1 Models with Endogenous Right-Hand-Side Variables 242
Example 8.2 Instrumental Variable Analysis 252
Example 8.3 Streams as Instruments 254
Example 8.4 Instrumental Variable in Regression 255
Example 8.5 Instrumental Variable Estimation of a Labor Supply Equation 258
Example 8.6 German Labor Market Interventions 265
Example 8.7 Treatment Effects on Earnings 266
Example 8.8 The Oregon Health Insurance Experiment 266
Example 8.9 The Effect of Counseling on Financial Management 266
Example 8.10 Treatment Effects on Earnings 271
Example 8.5 Labor Supply Model (Continued) 277
Example 8.11 Overidentification of the Labor Supply Equation 279
Example 8.12 Income and Education in a Study of Twins 286
Example 8.13 Instrumental Variables Estimates of the Consumption Function 291
Example 8.14 Does Television Watching Cause Autism? 292
Example 8.15 Is Season of Birth a Valid Instrument? 294

CHAPTER 9 The Generalized Regression Model and Heteroscedasticity 297

Example 9.1 Heteroscedastic Regression and the White Estimator 300
Example 9.2 Testing for Heteroscedasticity 315
Example 9.3 Multiplicative Heteroscedasticity 315
Example 9.4 Groupwise Heteroscedasticity 318

CHAPTER 10 Systems of Regression Equations 326

Example 10.1 A Regional Production Model for Public Capital 336
Example 10.2 Cobb–Douglas Cost Function 340
Example 10.3 A Cost Function for U.S. Manufacturing 344
Example 10.4. Reverse Causality and Endogeneity in Health 347

Example 10.5 Structure and Reduced Form in a Small Macroeconomic Model 351
Example 10.6 Identification of a Supply and Demand Model 355
Example 10.7 The Rank Condition and a Two-Equation Model 357
Example 10.8 Simultaneity in Health Production 360
Example 10.9 Klein's Model I 364

CHAPTER 11 Models for Panel Data 373

Example 11.1 A Rotating Panel: The Survey of Income and Program Participation (SIPP) Data 378
Example 11.2 Attrition and Inverse Probability Weighting in a Model for Health 378
Example 11.3 Attrition and Sample Selection in an Earnings Model for Physicians 380
Example 11.4 Wage Equation 385
Example 11.5 Robust Estimators of the Wage Equation 389
Example 11.6 Analysis of Covariance and the World Health Organization (WHO) Data 392
Example 11.7 Fixed Effects Estimates of a Wage Equation 397
Example 11.8 Two-Way Fixed Effects with Unbalanced Panel Data 399
Example 11.9 Heterogeneity in Time Trends in an Aggregate Production Function 402
Example 11.10 Test for Random Effects 411
Example 11.11 Estimates of the Random Effects Model 412
Example 11.12 Hausman and Variable Addition Tests for Fixed versus Random Effects 416
Example 11.13 Hospital Costs 419
Example 11.14 Spatial Autocorrelation in Real Estate Sales 424
Example 11.15 Spatial Lags in Health Expenditures 426
Example 11.16 Endogenous Income in a Health Production Model 429
Example 11.17 The Returns to Schooling 432
Example 11.18 The Returns to Schooling 433
Example 11.19 Dynamic Labor Supply Equation 443
Example 11.20 Health Care Utilization 446
Example 11.21 Exponential Model with Fixed Effects 448
Example 11.22 Random Coefficients Model 452
Example 11.23 Fannie Mae's Pass Through 453
Example 11.24 Dynamic Panel Data Models 455
Example 11.25 A Mixed Fixed Growth Model for Developing Countries 459

CHAPTER 12 Estimation Frameworks in Econometrics 465

Example 12.3 Joint Modeling of a Pair of Event Counts 472
Example 12.4 The Formula That Killed Wall Street 6 472
Example 12.5 Semiparametric Estimator for Binary Choice Models 475

Example 12.6 A Model of Vacation Expenditures 476
Example 12.1 The Linear Regression Model 468
Example 12.2 The Stochastic Frontier Model 468

CHAPTER 13 Minimum Distance Estimation and the Generalized Method of Moments 488

Example 13.1 Euler Equations and Life Cycle Consumption 488
Example 13.2 Method of Moments Estimator for N[μ, σ^2] 490
Example 13.3 Inverse Gaussian (Wald) Distribution 491
Example 13.4 Mixture of Normal Distributions 491
Example 13.5 Gamma Distribution 493
Example 13.5 (Continued) 495
Example 13.6 Minimum Distance Estimation of a Hospital Cost Function 498
Example 13.7 GMM Estimation of a Nonlinear Regression Model 504
Example 13.8 Empirical Moment Equation for Instrumental Variables 507
Example 13.9 Overidentifying Restrictions 511
Example 13.10 GMM Estimation of a Dynamic Panel Data Model of Local Government Expenditures 530

CHAPTER 14 Maximum Likelihood Estimation 537

Example 14.1 Identification of Parameters 538
Example 14.2 Log-Likelihood Function and Likelihood Equations for the Normal Distribution 541
Example 14.3 Information Matrix for the Normal Distribution 548
Example 14.4 Variance Estimators for an MLE 550
Example 14.5 Two-Step ML Estimation 567
Example 14.6 A Regression with Nonnormal Disturbances 572
Example 14.7 Cluster Robust Standard Errors 574
Example 14.8 Logistic, t, and Skew Normal Disturbances 579
Example 14.9 Testing for Constant Returns to Scale 584
Example 14.10 Multiplicative Heteroscedasticity 589
Example 14.11 Maximum Likelihood Estimation of Gasoline Demand 590
Example 14.12 Identification in a Loglinear Regression Model 591
Example 14.13 Geometric Regression Model for Doctor Visits 597
Example 14.14 Ml Estimates of a Seemingly Unrelated Regressions Model 602
Example 14.15 Maximum Likelihood and FGLS Estimates of a Wage Equation 608
Example 14.16 Statewide Productivity 610
Example 14.17 Random Effects Geometric Regression Model 617
Example 14.18 Fixed and Random Effects Geometric Regression 621
Example 14.19 A Normal Mixture Model for Grade Point Averages 623

Example 14.20 Latent Class Regression Model for Grade Point Averages 625
Example 14.21 Predicting Class Probabilities 627
Example 14.22 A Latent Class Two-Part Model for Health Care Utilization 630
Example 14.23 Latent Class Models for Health Care Utilization 631
Example 14.24 Semiparametric Random Effects Model 634

CHAPTER 15 Simulation-Based Estimation and Inference and Random Parameter Models 641

Example 15.1 Inferring the Sampling Distribution of the Least Squares Estimator 641
Example 15.2 Bootstrapping the Variance of the LAD Estimator 641
Example 15.3 Least Simulated Sum of Squares 642
Example 15.4 Long-Run Elasticities 648
Example 15.5 Bootstrapping the Variance of the Median 651
Example 15.6 Block Bootstrapping Standard Errors and Confidence Intervals in a Panel 653
Example 15.7 Monte Carlo Study of the Mean Versus the Median 654
Example 15.8 Fractional Moments of the Truncated Normal Distribution 663
Example 15.9 Estimating the Lognormal Mean 666
Example 15.10 Poisson Regression Model with Random Effects 672
Example 15.11 Maximum Simulated Likelihood Estimation of the Random Effects Linear Regression Model 672
Example 15.12 Random Parameters Wage Equation 675
Example 15.13 Least Simulated Sum of Squares Estimates of a Production Function Model 677
Example 15.14 Hierarchical Linear Model of Home Prices 679
Example 15.15 Individual State Estimates of a Private Capital Coefficient 684
Example 15.16 Mixed Linear Model for Wages 685
Example 15.17 Maximum Simulated Likelihood Estimation of a Binary Choice Model 689

CHAPTER 16 Bayesian Estimation and Inference 694

Example 16.1 Bayesian Estimation of a Probability 696
Example 16.2 Estimation with a Conjugate Prior 701
Example 16.3 Bayesian Estimate of the Marginal Propensity to Consume 703
Example 16.4 Posterior Odds for the Classical Regression Model 706
Example 16.5 Gibbs Sampling from the Normal Distribution 708
Example 16.6 Gibbs Sampler for a Probit Model 712
Example 16.7 Bayesian and Classical Estimation of Heterogeneity in the Returns to Education 717

CHAPTER 17 Binary Outcomes and Discrete Choices 725

Example 17.1 Labor Force Participation Model 728
Example 17.2 Structural Equations for a Binary Choice Model 730
Example 17.3 Probability Models 737
Example 17.4 The Light Bulb Puzzle: Examining Partial Effects 739
Example 17.5 Cheating in the Chicago School System—An LPM 741
Example 17.6 Robust Covariance Matrices for Probit and LPM Estimators 745
Example 17.7 Testing for Structural Break in a Logit Model 748
Example 17.8 Standard Errors for Partial Effects 752
Example 17.9 Hypothesis Tests About Partial Effects 753
Example 17.10 Confidence Intervals for Partial Effects 754
Example 17.11 Inference About Odds Ratios 754
Example 17.12 Interaction Effect 757
Example 17.13 Prediction with a Probit Model 760
Example 17.14 Fit Measures for a Logit Model 761
Example 17.15 Specification Test in a Labor Force Participation Model 765
Example 17.16 Distributional Assumptions 767
Example 17.17 Credit Scoring 768
Example 17.18 An Incentive Program for Quality Medical Care 771
Example 17.19 Moral Hazard in German Health Care 772
Example 17.20 Labor Supply Model 776
Example 17.21 Cardholder Status and Default Behavior 779
Example 17.22 Binary Choice Models for Panel Data 789
Example 17.23 Fixed Effects Logit Model: Magazine Prices Revisited 789
Example 17.24 Panel Data Random Effects Estimators 793
Example 17.25 A Dynamic Model for Labor Force Participation and Disability 796
Example 17.26 An Intertemporal Labor Force Participation Equation 796
Example 17.27 Semiparametric Models of Heterogeneity 797
Example 17.28 Parameter Heterogeneity in a Binary Choice Model 799
Example 17.29 Nonresponse in the GSOEP Sample 802
Example 17.30 A Spatial Logit Model for Auto Supplier Locations 806
Example 17.31 Tetrachoric Correlation 810
Example 17.32 Bivariate Probit Model for Health Care Utilization 813
Example 17.33 Bivariate Random Effects Model for Doctor and Hospital Visits 814
Example 17.34 The Impact of Catholic School Attendance on High School Performance 817
Example 17.35 Gender Economics Courses at Liberal Arts Colleges 817
Example 17.36 A Multivariate Probit Model for Product Innovations 820

CHAPTER 18 Multinomial Choices and Event Counts 826

Example 18.1 Hollingshead Scale of Occupations 831
Example 18.2 Home Heating Systems 832

Example 18.3 Multinomial Choice Model for Travel Mode 839
Example 18.4 Using Mixed Logit to Evaluate a Rebate Program 847
Example 18.5 Latent Class Analysis of the Demand for Green Energy 849
Example 18.6 Malaria Control During Pregnancy 852
Example 18.7 Willingness to Pay for Renewable Energy 855
Example 18.8 Stated Choice Experiment: Preference for Electricity Supplier 860
Example 18.9 Health Insurance Market 865
Example 18.10 Movie Ratings 867
Example 18.11 Rating Assignments 870
Example 18.12 Brant Test for an Ordered Probit Model of Health Satisfaction 873
Example 18.13 Calculus and Intermediate Economics Courses 873
Example 18.14 Health Satisfaction 877
Example 18.15 A Dynamic Ordered Choice Model: 878
Example 18.16 Count Data Models for Doctor Visits 892
Example 18.17 Major Derogatory Reports 896
Example 18.18 Extramarital Affairs 897
Example 18.19 Panel Data Models for Doctor Visits 904
Example 18.20 Zero-Inflation Models for Major Derogatory Reports 906
Example 18.21 Hurdle Models for Doctor Visits 909
Example 18.22 Endogenous Treatment in Health Care Utilization 913

CHAPTER 19 Limited Dependent VariablesÑTruncation, Censoring, and Sample Selection 918

Example 19.1 Truncated Uniform Distribution 920
Example 19.2 A Truncated Lognormal Income Distribution 921
Example 19.3 Stochastic Cost Frontier for Swiss Railroads 928
Example 19.4 Censored Random Variable 933
Example 19.5 Estimated Tobit Equations for Hours Worked 937
Example 19.6 Two-Part Model For Extramarital Affairs 942
Example 19.7 Multiplicative Heteroscedasticity in the Tobit Model 946
Example 19.8 Incidental Truncation 949
Example 19.9 A Model of Labor Supply 950
Example 19.10 Female Labor Supply 956
Example 19.11 A Mover-Stayer Model for Migration 957
Example 19.12 Doctor Visits and Insurance 958
Example 19.13 Survival Models for Strike Duration 975
Example 19.14 Time Until Retirement 976

CHAPTER 20 Serial Correlation 981

Example 20.1 Money Demand Equation 981
Example 20.2 Autocorrelation Induced by Misspecification of the Model 982

Example 20.3 Negative Autocorrelation in the Phillips Curve 983
Example 20.4 Autocorrelation Function for the Rate of Inflation 988
Example 20.5 Autocorrelation Consistent Covariance Estimation 999
Example 20.6 Test for Autocorrelation 1001
Example 20.7 Dynamically Complete Regression 1009
Example 20.8 Stochastic Volatility 1011
Example 20.9 GARCH Model for Exchange Rate Volatility 1017

CHAPTER 21 Nonstationary Data 1022
Example 21.1 A Nonstationary Series 1024
Example 21.2 Tests for Unit Roots 1030
Example 21.3 Augmented Dickey–Fuller Test for a Unit Root in GDP 1037
Example 21.4 Is there a Unit Root in GDP? 1039
Example 21.5 Cointegration in Consumption and Output 1040
Example 21.6 Several Cointegrated Series 1041
Example 21.7 Multiple Cointegrating Vectors 1043
Example 21.8 Cointegration in Consumption and Output 1046

Online Appendix C Estimation and Inference C-1
Example C.1 Descriptive Statistics for a Random Sample C-4
Example C.2 Kernel Density Estimator for the Income Data C-5
Example C.3 Sampling Distribution of A Sample Mean C-7
Example C.4 Sampling Distribution of the Sample Minimum C-7
Example C.5 Mean Squared Error of The Sample Variance C-11
Example C.6 Likelihood Functions for Exponential and Normal Distributions C-12
Example C.7 Variance Bound for the Poisson Distribution C-13
Example C.8 Confidence Intervals for the Normal Mean C-14
Example C.9 Estimated Confidence Intervals for a Normal Mean and Variance C-15
Example C.10 Testing a Hypothesis About a Mean C-17
Example C.11 Consistent Test About a Mean C-19
Example C.12 Testing A Hypothesis About a Mean with a Confidence Interval C-19
Example C.13 One-Sided Test About a Mean D-1

Online Appendix D Large-Sample Distribution Theory D-1
Example D.1 Mean Square Convergence of the Sample Minimum in Exponential Sampling D-4
Example D.2 Estimating a Function of the Mean D-5
Example D.3 Probability Limit of a Function of $\bar{x}$ and s^2 D-9
Example D.4 Limiting Distribution of t_{n-2} D-12
Example D.5 The F Distribution D-14
Example D.6 The Lindeberg–Levy Central Limit Theorem D-16

Example D.7 Asymptotic Distribution of the Mean of an Exponential Sample D-20
Example D.8 Asymptotic Inefficiency of the Median In Normal Sampling D-21
Example D.9 Asymptotic Distribution of a Function of Two Estimators D-22
Example D.10 Asymptotic Moments of the Normal Sample Variance D-23

PREFACE

ECONOMETRIC ANALYSIS

Econometric Analysis is a broad introduction to the field of econometrics. This field grows continually. A (not complete) list of journals devoted at least in part to econometrics now includes: *Econometric Reviews; Econometric Theory; Econometrica; Econometrics; Econometrics and Statistics; The Econometrics Journal; Empirical Economics; Foundations and Trends in Econometrics; The Journal of Applied Econometrics; The Journal of Business and Economic Statistics; The Journal of Choice Modelling; The Journal of Econometric Methods; The Journal of Econometrics; The Journal of Time Series Analysis; The Review of Economics and Statistics*. Constructing a textbook-style survey to introduce the topic at a graduate level has become increasingly ambitious. Nonetheless, that is what I seek to do here. This text attempts to present, at an entry graduate level, enough of the topics in econometrics that a student can comfortably move on from here to practice or to more advanced study. For example, the literature on "Treatment Effects" is already vast, rapidly growing, complex in the extreme, and occasionally even contradictory. But, there are a few bedrock principles presented in Chapter 8 that (I hope) can help the interested practitioner or student get started as they wade into this segment of the literature. The book is intended as a bridge between an introduction to econometrics and the professional literature.

The book has two objectives. The first is to introduce students to *applied econometrics*, including basic techniques in linear regression analysis and some of the rich variety of models that are used when the linear model proves inadequate or inappropriate. Modern software has made complicated modeling very easy to put into practice. The second objective is to present sufficient *theoretical background* so that the reader will (1) understand the advanced techniques that are made so simple in modern software and (2) recognize new variants of the models learned about here as merely natural extensions that fit within a common body of principles. This book contains a substantial amount of theoretical material, such as that on the GMM, maximum likelihood estimation, and asymptotic results for regression models.

One overriding purpose has motivated all eight editions of *Econometric Analysis*. The vast majority of readers of this book will be users, not developers, of econometrics. I believe that it is not sufficient to teach econometrics by reciting (and proving) the theories of estimation and inference. Although the often-subtle theory is extremely important, the application is equally crucial. To that end, I have provided hundreds of worked numerical examples and extracts from applications in the received empirical literature in many fields. My purpose in writing this work, and in my continuing efforts to update it, is to show readers how to *do* econometric analysis. But, I also believe that readers want (and need) to know what is going on behind the curtain when they use ever more sophisticated modern software for ever more complex econometric analyses.

I have taught econometrics at the level of *Econometric Analysis* at NYU for many years. I ask my students to learn how to use a (any) modern econometrics program as part of their study. I've lost track of the number of my students who recount to me their disappointment in a previous course in which they were taught how to use software, but not the theory and motivation of the techniques. In October, 2014, Google Scholar published its list of the 100 most cited works over all fields and all time. (www.nature.com/polopoly_fs/7.21245!/file/GoogleScholartop100.xlsx). *Econometric Analysis*, the only work in econometrics on the list, ranked number 34 with 48,100 citations. (As of this writing, November 2016, the number of citations to the first 7 editions in all languages approaches 60,000.) I take this extremely gratifying result as evidence that there are readers in many fields who agree that the practice of econometrics calls for an understanding of *why*, as well as *how* to use the tools in modern software. This book is for them.

THE EIGHTH EDITION OF *ECONOMETRIC ANALYSIS*

This text is intended for a one-year graduate course for social scientists. Prerequisites should include calculus, mathematical statistics, and an introduction to econometrics at the level of, say, Gujarati and Porter's (2011) *Basic Econometrics*, Stock and Watson's (2014) *Introduction to Econometrics*, Kennedy's (2008) *Guide to Econometrics*, or Wooldridge's (2015) *Introductory Econometrics: A Modern Approach*. I assume, for example, that the reader has already learned about the basics of econometric methodology including the fundamental role of economic and statistical assumptions; the distinctions between cross-section, time-series, and panel data sets; and the essential ingredients of estimation, inference, and prediction with the multiple linear regression model. Self-contained (for our purposes) summaries of the matrix algebra, mathematical statistics, and statistical theory used throughout the book are given in Appendices A through D. I rely heavily on matrix algebra throughout. This may be a bit daunting to some early on but matrix algebra is an indispensable tool and I hope the reader will come to agree that it is a means to an end, not an end in itself. With matrices, the unity of a variety of results will emerge without being obscured by a curtain of summation signs. Appendix E and Chapter 15 contain a description of numerical methods that will be useful to practicing econometricians (and to us in the later chapters of the book).

Estimation of advanced nonlinear models is now as routine as least squares. I have included five chapters on estimation methods used in current research and five chapters on applications in micro- and macroeconometrics. The nonlinear models used in these fields are now the staples of the applied econometrics literature. As a consequence, this book also contains a fair amount of material that will extend beyond many first courses in econometrics. Once again, I have included this in the hope of laying a foundation for study of the professional literature in these areas.

PLAN OF THE BOOK

The arrangement of the book is as follows:

Part I begins the formal development of econometrics with its fundamental pillar, the *linear multiple regression model*. Estimation and inference with the linear least squares estimator are analyzed in Chapters 2 through 6. The *nonlinear regression model* is introduced in Chapter 7 along with quantile, semi- and nonparametric regression, all as extensions of the familiar linear model. *Instrumental variables estimation* is developed in Chapter 8.

Part II presents three major extensions of the regression model. Chapter 9 presents the consequences of relaxing one of the main assumptions of the linear model, homoscedastic nonautocorrelated disturbances, to introduce the *generalized regression model*. The focus here is on heteroscedasticity; autocorrelation is mentioned, but a detailed treatment is deferred to Chapter 20 in the context of time-series data. Chapter 10 introduces systems of regression equations, in principle, as the approach to modeling simultaneously a set of random variables and, in practical terms, as an extension of the generalized linear regression model. Finally, *panel data methods*, primarily fixed and random effects models of heterogeneity, are presented in Chapter 11.

The second half of the book is devoted to topics that extend the linear regression model in many directions. Beginning with Chapter 12, we proceed to the more involved methods of analysis that contemporary researchers use in analysis of "real-world" data. Chapters 12 to 16 in Part III present different estimation methodologies. Chapter 12 presents an overview by making the distinctions between *parametric*, *semiparametric* and *nonparametric methods*. The leading application of semiparametric estimation in the current literature is the *generalized method of moments (GMM) estimator* presented in Chapter 13. This technique provides the platform for much of modern econometrics. *Maximum likelihood estimation* is developed in Chapter 14. *Monte Carlo* and *simulation-based methods* such as *bootstrapping* that have become a major component of current research are developed in Chapter 15. Finally, *Bayesian methods* are introduced in Chapter 16.

Parts IV and V develop two major subfields of econometric methods, *microeconometrics*, which is typically based on cross-section and panel data, and *macroeconometrics*, which is usually associated with analysis of time-series data. In Part IV, Chapters 17 to 19 are concerned with models of *discrete choice, censoring, truncation, sample selection, duration* and the analysis of *counts of events*. In Part V, Chapters 20 and 21, we consider two topics in time-series analysis, models of *serial correlation* and regression models for *nonstationary data*—the usual substance of macroeconomic analysis.

REVISIONS

With only a couple exceptions noted below, I have retained the broad outline of the text. I have revised the presentation throughout the book (including this preface) to streamline the development of topics, in some cases (I hope), to improve the clarity of the derivations. Major revisions include:

- I have moved the material related to "causal inference" forward to the early chapters of the book – these topics are now taught earlier in the graduate sequence than heretofore and I've placed them in the context of the models and methods where they appear rather than as separate topics in the more advanced sections of the seventh edition. Difference in difference regression as a method, and regression discontinuity designs now appear in Chapter 6 with the discussion of functional forms and in the context of extensive applications extracted from the literature. The analysis of treatment effects has all been moved from Chapter 19 (on censoring and truncation) to Chapter 8 on endogeneity under the heading of "Endogenous

Dummy Variables." Chapter 8, as a whole, now includes a much more detailed treatment of instrumental variable methods.

- I have added many new examples, some as extracts from applications in the received literature, and others as worked numerical examples. I have drawn applications from many different fields including industrial organization, transportation, health economics, popular culture and sports, urban development and labor economics.
- Chapter 10 on systems of equations has been shifted (yet further) from its early emphasis on formal simultaneous linear equations models to systems of regression equations and the leading application, the single endogenous variable in a two equation recursive model – this is the implicit form of the regression model that contains one "endogenous" variable.
- The use of robust estimation and inference methods has been woven more extensively into the general methodology, in practice and throughout this text. The ideas of robust estimation and inference are introduced immediately with the linear regression model in Chapters 4 and 5, rather than as accommodations to nonspherical disturbances in Chapter 9. The role that a robust variance estimator will play in the Wald statistic is developed immediately when the result is first presented in Chapter 5.
- Chapters 4 (Least Squares), 6 (Functional Forms), 8 (Endogeneity), 10 (Equation Systems) and 11 (Panel Data) have been heavily revised to emphasize both contemporary econometric methods and the applications.
- I have moved Appendices A-F to the Companion Web site, at www.pearsonhighered.com/greene, that accompanies this text. Students can access them at no cost.

The first semester of study in a course based on Econometric Analysis would focus on Chapters 1-6 (the linear regression model), 8 (endogeneity and causal modeling), and possibly some of 11 (panel data). Most of the revisions in the eighth edition appear in these chapters.

SOFTWARE AND DATA

There are many computer programs that are widely used for the computations described in this book. All were written by econometricians or statisticians, and in general, all are regularly updated to incorporate new developments in applied econometrics. A sampling of the most widely used packages and Web sites where you can find information about them are

EViews	www.eviews.com	(QMS, Irvine, CA)
Gauss	www.aptech.com	(Aptech Systems, Kent, WA)
LIMDEP	www.limdep.com	(Econometric Software, Plainview, NY)
MATLAB	www.mathworks.com	(Mathworks, Natick, MA)
NLOGIT	www.nlogit.com	(Econometric Software, Plainview, NY)
R	www.r-project.org/	(The R Project for Statistical Computing)
RATS	www.estima.com	(Estima, Evanston, IL)
SAS	www.sas.com	(SAS, Cary, NC)
Shazam	econometrics.com	(Northwest Econometrics Ltd., Gibsons, Canada)
Stata	www.stata.com	(Stata, College Station, TX)

A more extensive list of computer software used for econometric analysis can be found at the resource Web site, http://www.oswego.edu/~economic/econsoftware.htm.

With only a few exceptions, the computations described in this book can be carried out with any of the packages listed. *NLOGIT* was used for the computations in most of the applications. This text contains no instruction on using any particular program or language. Many authors have produced *RATS*, *LIMDEP/NLOGIT*, *EViews*, *SAS*, or *Stata* code for some of the applications, including, in a few cases, in the documentation for their computer programs. There are also quite a few volumes now specifically devoted to econometrics associated with particular packages, such as Cameron and Trivedi's (2009) companion to their treatise on microeconometrics.

The data sets used in the examples are also available on the Web site for the text, http://people.stern.nyu.edu/wgreene/Text/econometricanalysis.htm. Throughout the text, these data sets are referred to "Table Fn.m," for example Table F4.1. The "F" refers to Appendix F available on the Companion web site which contains descriptions of the data sets. The actual data are posted in generic ASCII and portable formats on the Web site with the other supplementary materials for the text. There are now thousands of interesting Web sites containing software, data sets, papers, and commentary on econometrics. It would be hopeless to attempt any kind of a survey. One code/data site that is particularly agreeably structured and well targeted for readers of this book is the data archive for the *Journal of Applied Econometrics* (*JAE*). They have archived all the nonconfidential data sets used in their publications since 1988 (with some gaps before 1995). This useful site can be found at http://qed.econ.queensu.ca/jae/. Several of the examples in the text use the *JAE* data sets. Where we have done so, we direct the reader to the *JAE*'s Web site, rather than our own, for replication. Other journals have begun to ask their authors to provide code and data to encourage replication. Another easy-to-navigate site for aggregate data on the U.S. economy is https://datahub.io/dataset/economagic.

ACKNOWLEDGMENTS

It is a pleasure to express my appreciation to those who have influenced this work. I remain grateful to Arthur Goldberger (dec.), Arnold Zellner (dec.), Dennis Aigner, Bill Becker, and Laurits Christensen for their encouragement and guidance. After eight editions of this book, the number of individuals who have significantly improved it through their comments, criticisms, and encouragement has become far too large for me to thank each of them individually. I am grateful for their help and I hope that all of them see their contribution to this edition. Any number of people have submitted tips about the text. You can find many of them listed in the errata pages on the text Web site, http://people.stern.nyu.edu/wgreene/Text/econometricanalysis.htm, in particular: David Hoaglin, University of Massachusetts; Randall Campbell, Mississippi State University; Carter Hill, Louisiana State University; and Tom Doan, Estima Corp. I would also like to thank two colleagues who have worked on translations of *Econometric Analysis*, Marina Turuntseva (the Russian edition) and Umit Senesen (the Turkish translation). I must also acknowledge the mail I've received from hundreds of readers and practitioners from the world over who have given me a view into topics and questions that practitioners are interested in, and have provided a vast trove of helpful material for my econometrics courses.

I also acknowledge the many reviewers of my work whose careful reading has vastly improved the book through this edition: Scott Atkinson, University of Georgia; Badi Baltagi, Syracuse University; Neal Beck, New York University; William E. Becker (Ret.), Indiana University; Eric J. Belasko, Texas Tech University; Anil Bera, University of Illinois; John Burkett, University of Rhode Island; Leonard Carlson, Emory University; Frank Chaloupka, University of Illinois at Chicago; Chris Cornwell, University of Georgia; Craig Depken II, University of Texas at Arlington; Frank Diebold, University of Pennsylvania; Edward Dwyer, Clemson University; Michael Ellis, Wesleyan University; Martin Evans, Georgetown University; Vahagn Galstyan, Trinity College Dublin; Paul Glewwe, University of Minnesota; Ed Greenberg, Washington University at St. Louis; Miguel Herce, University of North Carolina; Joseph Hilbe, Arizona State University; Dr. Uwe Jensen, Christian-Albrecht University; K. Rao Kadiyala, Purdue University; William Lott, University of Connecticut; Thomas L. Marsh, Washington State University; Edward Mathis, Villanova University; Mary McGarvey, University of Nebraska–Lincoln; Ed Melnick, New York University; Thad Mirer, State University of New York at Albany; Cyril Pasche, University of Geneva; Paul Ruud, University of California at Berkeley; Sherrie Rhine, Federal Deposit Insurance Corp.; Terry G. Seaks (Ret.), University of North Carolina at Greensboro; Donald Snyder, California State University at Los Angeles; Steven Stern, University of Virginia; Houston Stokes, University of Illinois at Chicago; Dmitrios Thomakos, Columbia University; Paul Wachtel, New York University; Mary Beth Walker, Georgia State University; Mark Watson, Harvard University; and Kenneth West, University of Wisconsin. My numerous discussions with Bruce McCullough of Drexel University have improved Appendix E and at the same time increased my appreciation for numerical analysis. I am especially grateful to Jan Kiviet of the University of Amsterdam, who subjected my third edition to a microscopic examination and provided literally scores of suggestions, virtually all of which appear herein. Professor Pedro Bacao, University of Coimbra, Portugal, and Mark Strahan of Sand Hill Econometrics and Umit Senesen of Istanbul Technical University did likewise with the sixth and seventh editions.

I would also like to thank the many people at Pearson Education who have put this book together with me: Adrienne D' Ambrosio, Neeraj Bhalla, Sugandh Juneja, and Nicole Suddeth and the composition team at SPi Global.

For over 25 years since the first edition, I've enjoyed the generous support and encouragement of many people, some close to me, especially my family, and many not so close. I'm especially grateful for the help, support and priceless encouragement of my wife, Sherrie Rhine, whose unending enthusiasm for this project has made it much less daunting, and much more fun.

William H. Greene
February 2017

1
ECONOMETRICS

1.1 INTRODUCTION

This book will present an introductory survey of econometrics. We will discuss the fundamental ideas that define the methodology and examine a large number of specific models, tools, and methods that econometricians use in analyzing data. This chapter will introduce the central ideas that are the paradigm of econometrics. Section 1.2 defines the field and notes the role that theory plays in motivating econometric practice. Sections 1.3 and 1.4 discuss the types of applications that are the focus of econometric analyses. The process of econometric modeling is presented in Section 1.5 with a classic application, Keynes's consumption function. A broad outline of the text is presented in Section 1.6. Section 1.7 notes some specific aspects of the presentation, including the use of numerical examples and the mathematical notation that will be used throughout the text.

1.2 THE PARADIGM OF ECONOMETRICS

> In the first issue of *Econometrica*, the Econometric Society stated that its main object shall be to promote studies that aim at a unification of the theoretical-quantitative and the empirical-quantitative approach to economic problems and that are penetrated by constructive and rigorous thinking similar to that which has come to dominate the natural sciences. . . . But there are several aspects of the quantitative approach to economics, and no single one of these aspects taken by itself, should be confounded with econometrics. Thus, econometrics is by no means the same as economic statistics. Nor is it identical with what we call general economic theory, although a considerable portion of this theory has a definitely quantitative character. Nor should econometrics be taken as synonomous [sic] with the application of mathematics to economics. Experience has shown that each of these three viewpoints, that of statistics, economic theory, and mathematics, is a necessary, but not by itself a sufficient, condition for a real understanding of the quantitative relations in modern economic life. It is the *unification* of all three that is powerful. And it is this unification that constitutes econometrics.

The Society responded to an unprecedented accumulation of statistical information. It saw a need to establish a body of principles that could organize what would otherwise become a bewildering mass of data. Neither the pillars nor the objectives of econometrics have changed in the years since this editorial appeared. Econometrics concerns itself with the application of mathematical statistics and the tools of statistical inference to the empirical measurement of relationships postulated by an underlying theory.

It is interesting to observe the response to a contemporary, likewise unprecedented accumulation of massive amounts of quantitative information in the form of "Big Data." Consider the following assessment of what Kitchin (2014) sees as a paradigm shift in the analysis of data.

> This article examines how the availability of Big Data, coupled with new data analytics, challenges established epistemologies across the sciences, social sciences and humanities, and assesses the extent to which they are engendering paradigm shifts across multiple disciplines. In particular, it critically explores new forms of empiricism that declare 'the end of theory,' the creation of data-driven rather than knowledge-driven science, and the development of digital humanities and computational social sciences that propose radically different ways to make sense of culture, history, economy and society. It is argued that: (1) Big Data and new data analytics are disruptive innovations which are reconfiguring in many instances how research is conducted; and (2) there is an urgent need for wider critical reflection within the academy on the epistemological implications of the unfolding data revolution, a task that has barely begun to be tackled despite the rapid changes in research practices presently taking place.

We note the suggestion that data-driven analytics are proposed to replace theory (and econometrics as envisioned by Frisch) for providing the organizing principles to guide empirical research. (We will examine an example in Chapter 18 where we consider analyzing survey data with ordered choice models. Also, see Varian (2014) for a more balanced view.) The focus is driven partly by the startling computational power that would have been unavailable to Frisch. It seems likely that the success of this new paradigm will turn at least partly on the questions pursued. Whether the interesting features of an underlying data-generating process can be revealed by appealing to the data themselves without a theoretical platform seems to be a prospect raised by the author. The article does focus on the role of an underlying theory in empirical research—this is a central pillar of econometric methodology. As of this writing, the success story of Big Data analysis is still being written.

The crucial role that econometrics plays in economics has grown over time. The Nobel Prize in Economics has recognized this contribution with numerous awards to econometricians, including the first which was given to (the same) Ragnar Frisch in 1969. Lawrence Klein in 1980, Trygve Haavelmo in 1989, James Heckman and Daniel McFadden in 2000, Robert Engle and Clive Granger in 2003. Christopher Sims in 2011 and Lars Hansen in 2013 were recognized for their empirical research. The 2000 prize was noteworthy in that it celebrated the work of two scientists whose research was devoted to the marriage of behavioral theory and econometric modeling.

Example 1.1 Behavioral Models and the Nobel Laureates

The pioneering work by both James Heckman and Dan McFadden rests firmly on a theoretical foundation of utility maximization.

For Heckman's, we begin with the standard theory of household utility maximization over consumption and leisure. The textbook model of utility maximization produces a demand for leisure time that translates into a supply function of labor. When home production (i.e., work in the home as opposed to the outside, formal labor market) is considered in the calculus, then desired *hours* of (formal) labor can be negative. An important conditioning variable is the *reservation wage*—the wage rate that will induce formal labor market participation. On the demand side of the labor market, we have firms that offer market wages that respond to such attributes as age, education, and experience. What can we learn about labor supply behavior based on observed market wages, these attributes, and observed hours in the formal market? Less than it might seem, intuitively because our observed data omit half the market—the data on formal labor market activity are not randomly drawn from the whole population.

Heckman's observations about this implicit truncation of the distribution of hours or wages revolutionized the analysis of labor markets. Parallel interpretations have since guided analyses in every area of the social sciences. The analysis of policy interventions such as education initiatives, job training and employment policies, health insurance programs, market creation, financial regulation, and a host of others is heavily influenced by Heckman's pioneering idea that when participation is part of the behavior being studied, the analyst must be cognizant of the impact of common influences in both the presence of the intervention and the outcome. We will visit the literature on sample selection and treatment/program evaluation in Chapters 5, 6, 8 and 19.

Textbook presentations of the theories of demand for goods that produce utility, because they deal in continuous variables, are conspicuously silent on the kinds of discrete choices that consumers make every day—what brand of product to choose, whether to buy a large commodity such as a car or a refrigerator, how to travel to work, whether to rent or buy a home, where to live, what candidate to vote for, and so on. Nonetheless, a model of *random utility* defined over the alternatives available to the consumer provides a theoretically sound platform for studying such choices. Important variables include, as always, income and relative prices. What can we learn about underlying preference structures from the discrete choices that consumers make? What must be assumed about these preferences to allow this kind of inference? What kinds of statistical models will allow us to draw inferences about preferences? McFadden's work on how commuters choose to travel to work, and on the underlying theory appropriate to this kind of modeling, has guided empirical research in discrete consumer choices for several decades. We will examine McFadden's models of discrete choice in Chapter 18.

1.3 THE PRACTICE OF ECONOMETRICS

We can make a useful distinction between *theoretical econometrics* and *applied econometrics*. Theorists develop new techniques for estimation and hypothesis testing and analyze the consequences of applying particular methods when the assumptions that justify those methods are not met. Applied econometricians are the users of these techniques and the analysts of data (real world and simulated). The distinction is far from sharp; practitioners routinely develop new analytical tools for the purposes of the study that they are involved in. This text contains a large amount of econometric theory, but it is directed toward applied econometrics. We have attempted to survey techniques, admittedly some quite elaborate and intricate, that have seen wide use in the field.

Applied econometric methods will be used for estimation of important quantities, analysis of economic outcomes such as policy changes, markets or individual behavior, testing theories, and for forecasting. The last of these is an art and science in itself that is the subject of a vast library of sources. Although we will briefly discuss some aspects of forecasting, our interest in this text will be on estimation and analysis of models. The presentation, where there is a distinction to be made, will contain a blend of microeconometric and macroeconometric techniques and applications. It is also necessary to distinguish between *time-series analysis* (which is not our focus) and methods that primarily use *time-series data*. The former is, like forecasting, a growth industry served by its own literature in many fields. While we will employ some of the techniques of time-series analysis, we will spend relatively little time developing first principles.

1.4 MICROECONOMETRICS AND MACROECONOMETRICS

The connection between underlying behavioral models and the modern practice of econometrics is increasingly strong. Another distinction is made between *microeconometrics* and *macroeconometrics*. The former is characterized by its analysis of cross section and panel data and by its focus on individual consumers, firms, and micro-level decision makers. Practitioners rely heavily on the theoretical tools of microeconomics including utility maximization, profit maximization, and market equilibrium. The analyses are directed at subtle, difficult questions that often require intricate formulations. A few applications are as follows:

- What are the likely effects on labor supply behavior of proposed negative income taxes? [Ashenfelter and Heckman (1974)]
- Does attending an elite college bring an expected payoff in expected lifetime income sufficient to justify the higher tuition? [Kreuger and Dale (1999) and Kreuger (2000)]
- Does a voluntary training program produce tangible benefits? Can these benefits be accurately measured? [Angrist (2001)]
- Does an increase in the minimum wage lead to reduced employment? [Card and Krueger (1994)]
- Do smaller class sizes bring real benefits in student performance? [Hanuschek (1999), Hoxby (2000), and Angrist and Lavy (1999)]
- Does the presence of health insurance induce individuals to make heavier use of the health care system—is moral hazard a measurable problem? [Riphahn et al. (2003)]
- Did the intervention addressing anticompetitive behavior of a group of 50 boarding schools by the UK Office of Fair Trading produce a measurable impact on fees charged? [Pesaresi, Flanagan, Scott, and Tragear (2015)]

Macroeconometrics is involved in the analysis of time-series data, usually of broad aggregates such as price levels, the money supply, exchange rates, output, investment, economic growth, and so on. The boundaries are not sharp. For example, an application that we will examine in this text concerns spending patterns of municipalities, which rests somewhere between the two fields. The very large field of financial econometrics is concerned with long time-series data and occasionally vast panel data sets, but with a sharply focused orientation toward models of individual behavior. The analysis of market returns and exchange rate behavior is neither exclusively macro- nor microeconometric. [We will not be spending any time in this text on financial econometrics. For those with an interest in this field, We would recommend the celebrated work by Campbell, Lo, and Mackinlay (1997), or for a more time-series–oriented approach, Tsay (2005).]

Macroeconomic model builders rely on the interactions between economic agents and policy makers. For example:

- Does a monetary policy regime that is strongly oriented toward controlling inflation impose a real cost in terms of lost output on the U.S. economy? [Cecchetti and Rich (2001)]
- Did 2001's largest federal tax cut in U.S. history contribute to or dampen the concurrent recession? Or was it irrelevant?

Each of these analyses would depart from a formal model of the process underlying the observed data.

The techniques used in econometrics have been employed in a widening variety of fields, including political methodology, sociology,[1] health economics, medical research (e.g., how do we handle attrition from medical treatment studies?) environmental economics, economic geography, transportation engineering, and numerous others. Practitioners in these fields and many more are all heavy users of the techniques described in this text.

1.5 ECONOMETRIC MODELING

Econometric analysis usually begins with a statement of a theoretical proposition. Consider, for example, a classic application by one of Frisch's contemporaries:

Example 1.2 Keynes's Consumption Function

From Keynes's (1936) *General Theory of Employment, Interest and Money*:

We shall therefore define what we shall call the propensity to consume as the functional relationship f between X, a given level of income, and C, the expenditure on consumption out of the level of income, so that $C = f(X)$.

The amount that the community spends on consumption depends (i) partly on the amount of its income, (ii) partly on other objective attendant circumstances, and (iii) partly on the subjective needs and the psychological propensities and habits of the individuals composing it. The fundamental psychological law upon which we are entitled to depend with great confidence, both a priori from our knowledge of human nature and from the detailed facts of experience, is that men are disposed, as a rule and on the average, to increase their consumption as their income increases, but not by as much as the increase in their income. That is, . . . dC/dX is positive and less than unity.

But, apart from short period changes in the level of income, it is also obvious that a higher absolute level of income will tend as a rule to widen the gap between income and consumption. . . . These reasons will lead, as a rule, to a greater proportion of income being saved as real income increases.

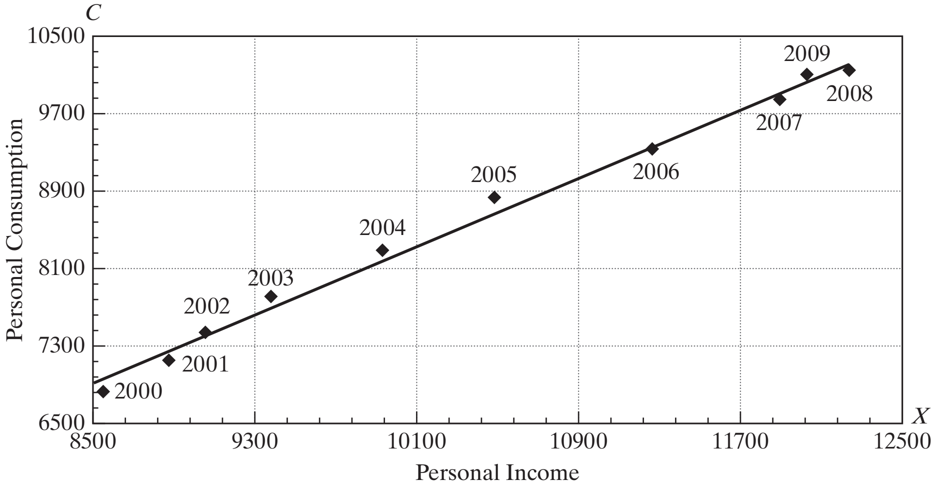
The theory asserts a relationship between consumption and income, $C = f(X)$, and claims in the second paragraph that the marginal propensity to consume (MPC), dC/dX, is between zero and one.[2] The final paragraph asserts that the average propensity to consume (APC), C/X, falls as income rises, or $d(C/X)/dX = (\text{MPC} - \text{APC})/X < 0$. It follows that MPC < APC. The

---

[1] See, for example, Long (1997) and DeMaris (2004).

[2] Modern economists are rarely this confident about their theories. More contemporary applications generally begin from first principles and behavioral axioms, rather than simple observation.

FIGURE 1.1 Aggregate U.S. Consumption and Income Data, 2000–2009.



most common formulation of the consumption function is a linear relationship, $C = \alpha + X\beta$, that satisfies Keynes's "laws" if β lies between zero and one and if α is greater than zero.

These theoretical propositions provide the basis for an econometric study. Given an appropriate data set, we could investigate whether the theory appears to be consistent with the observed "facts." For example, we could see whether the linear specification appears to be a satisfactory description of the relationship between consumption and income, and, if so, whether α is positive and β is between zero and one. Some issues that might be studied are (1) whether this relationship is stable through time or whether the parameters of the relationship change from one generation to the next (a change in the average propensity to save, $1 - \text{APC}$, might represent a fundamental change in the behavior of consumers in the economy); (2) whether there are systematic differences in the relationship across different countries, and, if so, what explains these differences; and (3) whether there are other factors that would improve the ability of the model to explain the relationship between consumption and income. For example, Figure 1.1 presents aggregate consumption and personal income in constant dollars for the United States for the 10 years of 2000–2009. (See Appendix Table F1.1.) Apparently, at least superficially, the data (the facts) are consistent with the theory. The relationship appears to be linear, albeit only approximately, the intercept of a line that lies close to most of the points is positive and the slope is less than one, although not by much. (However, if the line is fit by linear least squares regression, the intercept is negative, not positive.) Moreover, observers might disagree on what is meant by relationship in this description.

Economic theories such as Keynes's are typically sharp and unambiguous. Models of demand, production, labor supply, individual choice, educational attainment, income and wages, investment, market equilibrium, and aggregate consumption all specify precise, *deterministic relationships*. Dependent and independent variables are identified, a functional form is specified, and in most cases, at least a qualitative statement is made about the directions of effects that occur when independent variables in the model change. The model is only a simplification of reality. It will include the salient features of the relationship of interest but will leave unaccounted for influences that might well be present but are regarded as unimportant.

Correlations among economic variables are easily observable through descriptive statistics and techniques such as linear regression methods. The ultimate goal of the econometric model builder is often to uncover the deeper causal connections through elaborate structural, behavioral models. Note, for example, Keynes's use of the behavior of a *representative consumer* to motivate the behavior of macroeconomic variables, such as income and consumption. Heckman's model of labor supply noted in Example 1.1 is framed in a model of individual behavior. Berry, Levinsohn, and Pakes's (1995) detailed model of equilibrium pricing in the automobile market is another.

No model could hope to encompass the myriad essentially random aspects of economic life. It is thus also necessary to incorporate stochastic elements. As a consequence, observations on a variable will display variation attributable not only to differences in variables that are explicitly accounted for in the model, but also to the randomness of human behavior and the interaction of countless minor influences that are not. It is understood that the introduction of a random disturbance into a deterministic model is not intended merely to paper over its inadequacies. It is essential to examine the results of the study, in an ex post analysis, to ensure that the allegedly random, unexplained factor is truly unexplainable. If it is not, the model is, in fact, inadequate.[3] The stochastic element endows the model with its statistical properties. Observations on the variable(s) under study are thus taken to be the outcomes of a random process. With a sufficiently detailed stochastic structure and adequate data, the analysis will become a matter of deducing the properties of a probability distribution. The tools and methods of mathematical statistics will provide the operating principles.

A model (or theory) can never truly be confirmed unless it is made so broad as to include every possibility. But it may be subjected to ever more rigorous scrutiny and, in the face of contradictory evidence, refuted. A deterministic theory will be invalidated by a single contradictory observation. The introduction of stochastic elements into the model changes it from an exact statement to a probabilistic description about expected outcomes and carries with it an important implication. Only a preponderance of contradictory evidence can convincingly invalidate the probabilistic model, and what constitutes a preponderance of evidence is a matter of interpretation. Thus, the probabilistic model is less precise but at the same time, more robust.[4]

The process of econometric analysis departs from the specification of a theoretical relationship. We initially proceed on the optimistic assumption that we can obtain precise measurements on all the variables in a correctly specified model. If the ideal conditions are met at every step, the subsequent analysis will be routine. Unfortunately, they rarely are. Some of the difficulties one can expect to encounter are the following:

- The data may be badly measured or may correspond only vaguely to the variables in the model. "The interest rate" is one example.

[3] In the example given earlier, the estimated constant term in the linear least squares regression is negative. Is the theory wrong, or is the finding due to random fluctuation in the data? Another possibility is that the theory is broadly correct, but the world changed between 1936 when Keynes devised his theory and 2000–2009 when the data (outcomes) were generated. Or, perhaps linear least squares is not the appropriate technique to use for this model, and that is responsible for the inconvenient result (the negative intercept).

[4] See Keuzenkamp and Magnus (1995) for a lengthy symposium on testing in econometrics.

- Some of the variables may be inherently unmeasurable. "Expectations" is a case in point.
- The theory may make only a rough guess as to the correct form of the model, if it makes any at all, and we may be forced to choose from an embarrassingly long menu of possibilities.
- The assumed stochastic properties of the random terms in the model may be demonstrably violated, which may call into question the methods of estimation and inference procedures we have used.
- Some relevant variables may be missing from the model.
- The conditions under which data are collected lead to a sample of observations that is systematically unrepresentative of the population we wish to study.

The ensuing steps of the analysis consist of coping with these problems and attempting to extract whatever information is likely to be present in such obviously imperfect data. The methodology is that of mathematical statistics and economic theory. The product is an econometric model.

1.6 PLAN OF THE BOOK

Our objective in this survey is to develop in detail a set of tools, then use those tools in applications. The following set of applications will include many that readers will use in practice. But it is not exhaustive. We will attempt to present our results in sufficient generality that the tools we develop here can be extended to other kinds of situations and applications not described here.

One possible approach is to organize (and orient) the areas of study by the type of data being analyzed—cross section, panel, discrete data, then time series being the obvious organization.

Alternatively, we could distinguish at the outset between micro- and macroeconometrics.[5] Ultimately, all of these will require a common set of tools, including, for example, the multiple regression model, the use of moment conditions for estimation, instrumental variables (IV), and maximum likelihood estimation. With that in mind, the organization of this book is as follows: The first half of the text develops fundamental results that are common to all the applications. The concept of multiple regression and the linear regression model in particular constitutes the underlying platform of most modeling, even if the linear model itself is not ultimately used as the empirical specification. This part of the text concludes with developments of IV estimation and the general topic of panel data modeling. The latter pulls together many features of modern econometrics, such as, again, IV estimation, modeling heterogeneity, and a rich variety of extensions of the linear model. The second half of the text presents a variety

[5] An excellent reference on the former that is at a more advanced level than this text is Cameron and Trivedi (2005). There does not appear to be available a counterpart, large-scale pedagogical survey of macroeconometrics that includes both econometric theory and applications. The numerous more focused studies include books such as Bardsen et al. (2005).

of topics. Part III is an overview of estimation methods. Finally, Parts IV and V present results from microeconometrics and macroeconometrics, respectively. The broad outline is as follows:

I. **Regression Modeling**
Chapters 2 through 6 present the multiple linear regression model. We will discuss specification, estimation, and statistical inference. This part develops the ideas of estimation, robust analysis, functional form, and principles of model specification.

II. **Generalized Regression, Instrumental Variables, and Panel Data**
Chapter 7 extends the regression model to nonlinear functional forms. The method of instrumental variables is presented in Chapter 8. Chapters 9 and 10 introduce the generalized regression model and systems of regression models. This section ends with Chapter 11 on panel data methods.

III. **Estimation Methods**
Chapters 12 through 16 present general results on different methods of estimation including GMM, maximum likelihood, and simulation-based methods. Various estimation frameworks, including non- and semiparametric and Bayesian estimation, are presented in Chapters 12 and 16.

IV. **Microeconometric Methods**
Chapters 17 through 19 are about microeconometrics, discrete choice modeling, limited dependent variables, and the analysis of data on events—how many occur in a given setting and when they occur. Chapters 17 through 19 are devoted to methods more suited to cross sections and panel data sets.

V. **Macroeconometric Methods**
Chapters 20 and 21 focus on time-series modeling and macroeconometrics.

VI. **Background Materials**
Appendices A through E present background material on tools used in econometrics including matrix algebra, probability and distribution theory, estimation, and asymptotic distribution theory. Appendix E presents results on computation. The data sets used in the numerical examples are described in Appendix F. The actual data sets and other supplementary materials can be downloaded from the author's Web site for the text: http://people.stern.nyu.edu/wgreene/Text/.

1.7 PRELIMINARIES

Before beginning, we note some specific aspects of the presentation in the text.

1.7.1 NUMERICAL EXAMPLES

There are many numerical examples given throughout the discussion. Most of these are either self-contained exercises or extracts from published studies. In general, their purpose is to provide a limited application to illustrate a method or model. The reader can replicate them with the data sets provided. This will generally not entail attempting to replicate the full published study. Rather, we use the data sets to provide applications that relate to the published study in a limited fashion that also focuses on a particular