

Contents

Introduction	xvii
I Experiments	1
1 Ordinary Least Squares	3
1.1 Introduction	3
1.2 Estimating the Causal Effect	3
1.2.1 Graphing the Causal Effect	3
1.2.2 A Linear Causal Model	4
1.2.3 Simulation of the Causal Effect	4
1.2.4 Averaging to Estimate the Causal Effect	5
1.2.5 Assumptions of the OLS Model	7
1.3 Matrix Algebra of the OLS Model	7
1.3.1 Standard Algebra of the OLS Model	8
1.3.2 Algebraic OLS Estimator in R	9
1.3.3 Using Matrices	9
1.3.4 Multiplying Matrices in R	10
1.3.5 Matrix Estimator of OLS	11
1.3.6 Matrix Estimator of OLS in R	12
1.4 Least Squares Method for OLS	13
1.4.1 Moment Estimation	13
1.4.2 Algebra of Least Squares	14
1.4.3 Estimating Least Squares in R	14
1.4.4 The <code>lm()</code> Function	15
1.5 Measuring Uncertainty	16
1.5.1 Data Simulations	16
1.5.2 Introduction to the Bootstrap	18
1.5.3 Bootstrap in R	18
1.5.4 Standard Errors	19
1.6 Returns to Schooling	20
1.6.1 A Linear Model of Returns to Schooling	20
1.6.2 NLSM Data	20
1.6.3 Plotting Returns to Schooling	21
1.6.4 Estimating Returns to Schooling	22
1.7 Discussion and Further Reading	23

2	Multiple Regression	25
2.1	Introduction	25
2.2	Long and Short Regression	25
2.2.1	Using Short Regression	25
2.2.2	Independent Explanatory Variables	26
2.2.3	Dependent Explanatory Variables	27
2.2.4	Simulation with Multiple Explanatory Variables	27
2.2.5	Matrix Algebra of Short Regression	30
2.3	Collinearity and Multicollinearity	31
2.3.1	Matrix Algebra of Multicollinearity	32
2.3.2	Understanding Multicollinearity with R	32
2.4	Returns to Schooling	33
2.4.1	Multiple Regression of Returns to Schooling	33
2.4.2	NLSM Data	34
2.4.3	OLS Estimates of Returns to Schooling	34
2.5	Causal Pathways	36
2.5.1	Dual Path Model	36
2.5.2	Simulation of Dual Path Model	38
2.5.3	Dual Path Estimator Versus Long Regression	39
2.5.4	Matrix Algebra of the Dual Path Estimator	42
2.5.5	Dual Path Estimator in R	43
2.6	Are Bankers Racist or Greedy?	43
2.6.1	Boston HMDA Data	44
2.6.2	Causal Pathways of Discrimination	44
2.6.3	Estimating the Direct Effect	45
2.6.4	Adding in More Variables	46
2.6.5	Bootstrap Dual Path Estimator in R	47
2.6.6	Policy Implications of Dual Path Estimates	48
2.7	Discussion and Further Reading	49
3	Instrumental Variables	51
3.1	Introduction	51
3.2	A Confounded Model	52
3.2.1	Confounded Model DAG	52
3.2.2	Confounded Linear Model	53
3.2.3	Simulation of Confounded Data	54
3.3	IV Estimator	54
3.3.1	Graph Algebra of IV Estimator	55
3.3.2	Properties of IV Estimator	56
3.3.3	IV Estimator with Standard Algebra	56
3.3.4	Simulation of an IV Estimator	56
3.3.5	IV Estimator with Matrix Algebra	57
3.3.6	Two-Stage Least Squares	58
3.3.7	IV Estimator in R	59
3.3.8	Bootstrap IV Estimator for R	59

3.4	Returns to Schooling	61
3.4.1	Distance to College as an Instrument	62
3.4.2	NLSM Data	62
3.4.3	Simple IV Estimates of Returns to Schooling	63
3.4.4	Matrix Algebra IV Estimates of Returns to Schooling	64
3.4.5	Concerns with Distance to College	65
3.5	Instrument Validity	66
3.5.1	Test of Instrument Validity	67
3.5.2	Test of Instrument Validity in R	68
3.6	Better LATE than Nothing	69
3.6.1	Heterogeneous Effects	69
3.6.2	Local Average Treatment Effect	70
3.6.3	LATE Estimator	72
3.6.4	LATE Estimates of Returns to Schooling	72
3.7	Discussion and Further Reading	73
4	Bounds Estimation	75
4.1	Introduction	75
4.2	Potential Outcomes	76
4.2.1	Model of Potential Outcomes	76
4.2.2	Simulation of Impossible Data	76
4.2.3	Distribution of the Treatment Effect	77
4.3	Average Treatment Effect	79
4.3.1	ATE and Its Derivation	79
4.3.2	ATE and Do Operators	80
4.3.3	ATE and Unconfoundedness	81
4.3.4	ATE and Simulated Data	82
4.4	Kolmogorov Bounds	82
4.4.1	Kolmogorov's Conjecture	82
4.4.2	Kolmogorov Bounds in R	83
4.5	Do "Nudges" Increase Savings?	84
4.5.1	Field Experiment Data	85
4.5.2	Bounds on the Distribution of Balance Changes	86
4.5.3	Intent To Treat Discussion	87
4.6	Manski Bounds	87
4.6.1	Confounded Model	87
4.6.2	Simulation of Manski Bounds	88
4.6.3	Bounding the Average Treatment Effect	88
4.6.4	Natural Bounds of the Average Treatment Effect	89
4.6.5	Natural Bounds with Simulated Data	90
4.6.6	Are Natural Bounds Useless?	90
4.6.7	Bounds with Exogenous Variation	91
4.6.8	Exogenous Variation in Simulated Data	92
4.6.9	Bounds with Monotonicity	92
4.6.10	Bounds with Monotonicity in Simulated Data	93

4.7	More Guns, Less Crime?	94
4.7.1	Crime Data	94
4.7.2	ATE of RTC Laws under Unconfoundedness	95
4.7.3	Natural Bounds on ATE of RTC Laws	95
4.7.4	Bounds on ATE of RTC Laws with Exogenous Variation	97
4.7.5	Bounds on ATE of RTC Laws with Monotonicity	98
4.8	Discussion and Further Reading	98
II	Structural Estimation	101
5	Estimating Demand	103
5.1	Introduction	103
5.2	Revealed Preference	104
5.2.1	Modeling Demand	104
5.2.2	Simulating Demand	105
5.2.3	Revealing Demand	106
5.3	Discrete Choice	106
5.3.1	Simple Discrete Choice Model	107
5.3.2	Simulating Discrete Choice	108
5.3.3	Modeling Discrete Choice	108
5.4	Maximum Likelihood	110
5.4.1	Binomial Likelihood	110
5.4.2	Binomial Likelihood in $\mathbf{R}$	112
5.4.3	OLS with Maximum Likelihood	112
5.4.4	Maximum Likelihood OLS in $\mathbf{R}$	114
5.4.5	Probit	115
5.4.6	Probit in $\mathbf{R}$	116
5.4.7	Generalized Linear Model	116
5.5	McFadden's Random Utility Model	117
5.5.1	Model of Demand	117
5.5.2	Probit and Logit Estimators	118
5.5.3	Simulation with Probit and Logit Estimators	119
5.6	Multinomial Choice	121
5.6.1	Multinomial Choice Model	121
5.6.2	Multinomial Probit	122
5.6.3	Multinomial Probit in $\mathbf{R}$	124
5.6.4	Multinomial Logit	125
5.6.5	Multinomial Logit in $\mathbf{R}$	125
5.6.6	Simulating Multinomial Choice	126
5.7	Demand for Rail	127
5.7.1	National Household Travel Survey	128
5.7.2	Demand for Cars	129
5.7.3	Estimating Demand for Rail	131
5.7.4	Predicting Demand for Rail	131
5.8	Discussion and Further Reading	132

6	Estimating Selection Models	135
6.1	Introduction	135
6.2	Modeling Censored Data	136
6.2.1	A Model of Censored Data	136
6.2.2	Simulation of Censored Data	136
6.2.3	Latent Value Model	138
6.2.4	Tobit Estimator	140
6.2.5	Tobit Estimator in R	140
6.3	Censoring Due to Minimum Wages	141
6.3.1	National Longitudinal Survey of Youth 1997	142
6.3.2	Tobit Estimates	143
6.4	Modeling Selected Data	144
6.4.1	A Selection Model	145
6.4.2	Simulation of a Selection Model	145
6.4.3	Heckman Model	146
6.4.4	Heckman Estimator	148
6.4.5	Heckman Estimator in R	149
6.5	Analyzing the Gender Wage Gap	150
6.5.1	NLSY97 Data	150
6.5.2	Choosing to Work	152
6.5.3	Heckman Estimates of Gender Gap	152
6.6	Back to School Returns	155
6.6.1	NLSM Data	155
6.6.2	College vs. No College	157
6.6.3	Choosing College	158
6.6.4	Heckman Estimates of Returns to Schooling	158
6.6.5	Effect of College	161
6.7	Discussion and Further Reading	162
7	Demand Estimation with IV	163
7.1	Introduction	163
7.2	Modeling Competition	164
7.2.1	Competition is a Game	164
7.2.2	Hotelling's Line	165
7.2.3	Nash Equilibrium	167
7.3	Estimating Demand in Hotelling's Model	168
7.3.1	Simulation of Hotelling Model	168
7.3.2	Prices are Endogenous	168
7.3.3	Cost Shifters	169
7.3.4	Demand Shifters	171
7.4	Berry Model of Demand	171
7.4.1	Choosing Prices	172
7.4.2	A Problem with Cost Shifters	172
7.4.3	Empirical Model of Demand	173
7.4.4	Inverting Demand	173

7.4.5	Demand Shifters to Estimate Supply	174
7.4.6	Demand Estimation from Supply Estimates	175
7.5	Introduction of Apple Cinnamon Cheerios	175
7.5.1	Dominick's Data for Cereal	176
7.5.2	Instrument for Price of Cereal	177
7.5.3	Demand for Apple Cinnamon Cheerios	178
7.5.4	Value of Apple Cinnamon Cheerios	181
7.6	Discussion and Further Reading	181
8	Estimating Games	183
8.1	Introduction	183
8.2	Mixed Strategy Nash Equilibrium	183
8.2.1	Coaches' Decision Problem	184
8.2.2	Zero-Sum Game	185
8.2.3	Nash Equilibrium	186
8.2.4	Third and Fourth Down Game	187
8.2.5	Equilibrium Strategies	188
8.2.6	Equilibrium Strategies in R	189
8.2.7	Simulation of Third and Fourth Down Game	189
8.3	Generalized Method of Moments	191
8.3.1	Moments of OLS	191
8.3.2	Simulated Moments OLS	192
8.3.3	GMM OLS Estimator	194
8.3.4	GMM OLS Estimator in R	195
8.3.5	GMM of Returns to Schooling	196
8.4	Estimating the Third Down Game	197
8.4.1	Moment Conditions	198
8.4.2	Third Down GMM Estimator in R	198
8.5	Are NFL Coaches Rational?	199
8.5.1	NFL Data	200
8.5.2	Estimating Third Down Game in R	201
8.5.3	Predicting the Fourth Down Game	203
8.5.4	Testing Rationality of NFL Coaches	205
8.6	Discussion and Further Reading	205
9	Estimating Auction Models	207
9.1	Introduction	207
9.2	Sealed Bid Auctions	208
9.2.1	Sealed Bid Model	209
9.2.2	Sealed Bid Simulation	210
9.2.3	Sealed Bid Estimator	211
9.2.4	Sealed Bid Estimator in R	212
9.3	English Auctions	213
9.3.1	Order Statistics	214
9.3.2	Identifying the Value Distribution	215

9.3.3	English Auction Estimator	216
9.3.4	English Auction Estimator in R	216
9.4	Are Loggers Rational?	218
9.4.1	Timber Data	219
9.4.2	Sealed Bid Auctions	220
9.4.3	English Auctions	221
9.4.4	Comparing Estimates	221
9.5	Are Loggers Colluding?	221
9.5.1	A Test of Collusion	222
9.5.2	“Large” English Auctions	223
9.5.3	Large English Auction Estimator	224
9.5.4	Large English Auction Estimator in R	224
9.5.5	Evidence of Collusion	225
9.5.6	Large Sealed Bid Auction Estimator	226
9.5.7	Large Sealed Bid Auction Estimator in R	227
9.6	Discussion and Further Reading	227

III Repeated Measurement 231

10	Panel Data	233
10.1	Introduction	233
10.2	First Differences	233
10.2.1	First Difference Model	234
10.2.2	Simulated Panel Data	235
10.2.3	OLS Estimation of First Differences	235
10.3	Difference in Difference	236
10.3.1	Difference in Difference Estimator	236
10.3.2	Difference in Difference Estimator in R	237
10.4	Minimum Wage Increase in New Jersey	238
10.4.1	Data from Card and Krueger (1994)	238
10.4.2	Difference in Difference Estimates	239
10.5	Fixed Effects	241
10.5.1	Fixed Effects Estimator	241
10.5.2	Nuisance Parameter	242
10.5.3	Adjusted Fixed Effects Estimator	243
10.5.4	Two Step Fixed Effects Estimator	243
10.5.5	Fixed Effects Estimator in R	244
10.6	Effect of a Federal Minimum Wage Increase	246
10.6.1	NLSY97	247
10.6.2	Fixed Effects Estimators of the Minimum Wage	248
10.6.3	Are Workers Better Off?	249
10.7	Discussion and Further Reading	250

11 Synthetic Controls	251
11.1 Introduction	251
11.2 Beyond “Parallel Trends”	251
11.2.1 A General Fixed Effects Model	252
11.2.2 A Slightly Less General Model	252
11.2.3 Synthetic Synthetic Control Data	252
11.2.4 Constructing Synthetic Controls with OLS	254
11.2.5 OLS Weights in R	254
11.2.6 A “Wide” Data Problem	255
11.3 Abadie Estimator	256
11.3.1 Restricting Weights	257
11.3.2 Synthetic Control Estimator in R	258
11.4 Regularization	258
11.4.1 Turducken Estimation	259
11.4.2 LASSO	259
11.4.3 LASSO in R	260
11.5 Factor Models	261
11.5.1 Matrix Factorization	262
11.5.2 Convex Matrix Factorization	262
11.5.3 Synthetic Controls using Factors	264
11.5.4 Estimating the Weights	264
11.5.5 Convex Factor Model Estimator in R	265
11.6 Returning to Minimum Wage Effects	268
11.6.1 NLSY97 Data	268
11.6.2 Synthetic Control Estimates	269
11.6.3 LASSO Estimates	270
11.6.4 Factor Model Estimates	270
11.7 Discussion and Further Reading	274
12 Mixture Models	277
12.1 Introduction	277
12.2 Two-Type Mixture Models	278
12.2.1 Simulation of Two-Type Mixture	278
12.2.2 Knowing the Component Distributions	279
12.2.3 Observing Multiple Signals	280
12.3 Two Signal Mixture Models	283
12.3.1 Model of Two Signal Data	283
12.3.2 Simulation of Two Signal Data	283
12.3.3 Conditional Independence of Signals	284
12.3.4 Two Signal Estimator	284
12.3.5 Two Signal Estimator Algorithm	285
12.3.6 Mixture Model Estimator in R	286
12.4 Twins Reports and Returns to Schooling	288
12.4.1 Mixture Model of Twin Reports	289
12.4.2 Twin Reports Data	290

12.4.3	Mixture Model Approach to Measurement Error	290
12.4.4	Estimating Returns to Schooling from Twins	292
12.5	Revisiting Minimum Wage Effects	294
12.5.1	Restaurant Employment	294
12.5.2	Mixture Model Estimation of Restaurant Type	295
12.5.3	Heterogeneous Minimum Wage Effect	295
12.6	Discussion and Further Reading	299

IV Appendices 301

A	Measuring Uncertainty	303
A.1	Introduction	303
A.2	Classical Statistics	303
A.2.1	A Model of a Sample	304
A.2.2	Simulation of a Sample	304
A.2.3	Many Imaginary Samples	304
A.2.4	Law of Large Numbers	306
A.2.5	Central Limit Theorem	307
A.2.6	Approximation of the Limiting Distribution	308
A.2.7	Simulation of Approximate Distributions	309
A.2.8	Bootstrap	310
A.2.9	Hypothesis Testing	312
A.3	Bayesian Statistics	314
A.3.1	Bayes' Rule	314
A.3.2	Determining the Posterior	314
A.3.3	Determining the Posterior in R	315
A.4	Empirical Bayesian Estimation	316
A.4.1	A Large Number of Samples	316
A.4.2	Solving for the Prior and Posterior	317
A.4.3	Solving for the Prior in R	318
A.4.4	Estimating the Posterior of the Mean	320
A.5	The Sultan of the Small Sample Size	322
A.5.1	Classical or Bayesian?	322
A.5.2	Uniform Prior	323
A.5.3	Estimating the Prior	323
A.5.4	Paciorek's Posterior	325
A.6	Decision Theory	326
A.6.1	Decision Making Under Uncertainty	327
A.6.2	Compound Lotteries	327
A.6.3	What in the World Does Wald Think?	328
A.6.4	Two Types of Uncertainty?	329
A.7	Discussion and Further Reading	329

B Statistical Programming in R	331
B.1 Introduction	331
B.2 Objects in R	331
B.2.1 Vectors	331
B.2.2 Matrices	333
B.2.3 Lists	335
B.2.4 Data Frames	337
B.3 Interacting with Objects	338
B.3.1 Transforming Objects	338
B.3.2 Logical Expressions	339
B.3.3 Retrieving Information from a Position	340
B.3.4 Retrieving the Position from the Information	343
B.4 Statistics	344
B.4.1 Data	345
B.4.2 Missing Values	345
B.4.3 Summary Statistics	346
B.4.4 Regression	347
B.5 Control	347
B.5.1 Loops	347
B.5.2 Looping in R	348
B.5.3 If Else	349
B.6 Optimization	350
B.6.1 Functions	350
B.6.2 <code>optim()</code>	352
B.7 Discussion and Further Reading	353
Note	355
Bibliography	357
Author index	363
Subject index	365