

CONTENTS

Preface	xxv
Acknowledgments	xxvii
Notation	xxix

I INTRODUCTION **I**

1.1	What Is Econometrics?	1
1.2	The Probability Approach to Econometrics	1
1.3	Econometric Terms	2
1.4	Observational Data	3
1.5	Standard Data Structures	4
1.6	Econometric Software	6
1.7	Replication	6
1.8	Data Files for Textbook	7
1.9	Reading the Book	9

PART I REGRESSION **II**

2 CONDITIONAL EXPECTATION AND PROJECTION **13**

2.1	Introduction	13
2.2	The Distribution of Wages	13
2.3	Conditional Expectation	15
2.4	Logs and Percentages	17
2.5	Conditional Expectation Function	19
2.6	Continuous Variables	20
2.7	Law of Iterated Expectations	21
2.8	CEF Error	23
2.9	Intercept-Only Model	25
2.10	Regression Variance	25
2.11	Best Predictor	26
2.12	Conditional Variance	27
2.13	Homoskedasticity and Heteroskedasticity	29

2.14	Regression Derivative
2.15	Linear CEF
2.16	Linear CEF with Nonlinear Effects
2.17	Linear CEF with Dummy Variables
2.18	Best Linear Predictor
2.19	Illustrations of Best Linear Predictor
2.20	Linear Predictor Error Variance
2.21	Regression Coefficients
2.22	Regression Subvectors
2.23	Coefficient Decomposition
2.24	Omitted Variable Bias
2.25	Best Linear Approximation
2.26	Regression to the Mean
2.27	Reverse Regression
2.28	Limitations of the Best Linear Projection
2.29	Random Coefficient Model
2.30	Causal Effects
2.31	Existence and Uniqueness of the Conditional Expectation*
2.32	Identification*
2.33	Technical Proofs*
2.34	Exercises

3 THE ALGEBRA OF LEAST SQUARES

3.1	Introduction
3.2	Samples
3.3	Moment Estimators
3.4	Least Squares Estimator
3.5	Solving for Least Squares with One Regressor
3.6	Solving for Least Squares with Multiple Regressors
3.7	Illustration
3.8	Least Squares Residuals
3.9	Demeaned Regressors
3.10	Model in Matrix Notation
3.11	Projection Matrix
3.12	Annihilator Matrix
3.13	Estimation of Error Variance
3.14	Analysis of Variance
3.15	Projections
3.16	Regression Components
3.17	Regression Components (Alternative Derivation)*
3.18	Residual Regression
3.19	Leverage Values
3.20	Leave-One-Out Regression
3.21	Influential Observations
3.22	CPS Dataset

3.23	Numerical Computation	88
3.24	Collinearity Errors	89
3.25	Programming	91
3.26	Exercises	94

4 LEAST SQUARES REGRESSION 98

4.1	Introduction	98
4.2	Random Sampling	98
4.3	Sample Mean	99
4.4	Linear Regression Model	100
4.5	Expectation of Least Squares Estimator	100
4.6	Variance of Least Squares Estimator	102
4.7	Unconditional Moments	103
4.8	Gauss-Markov Theorem	104
4.9	Generalized Least Squares	107
4.10	Residuals	109
4.11	Estimation of Error Variance	110
4.12	Mean-Squared Forecast Error	111
4.13	Covariance Matrix Estimation under Homoskedasticity	113
4.14	Covariance Matrix Estimation under Heteroskedasticity	113
4.15	Standard Errors	117
4.16	Estimation with Sparse Dummy Variables	118
4.17	Computation	120
4.18	Measures of Fit	121
4.19	Empirical Example	122
4.20	Multicollinearity	123
4.21	Clustered Sampling	124
4.22	Inference with Clustered Samples	131
4.23	At What Level to Cluster?	131
4.24	Technical Proofs*	132
4.25	Exercises	134

5 NORMAL REGRESSION 139

5.1	Introduction	139
5.2	The Normal Distribution	139
5.3	Multivariate Normal Distribution	141
5.4	Joint Normality and Linear Regression	142
5.5	Normal Regression Model	143
5.6	Distribution of OLS Coefficient Vector	145
5.7	Distribution of OLS Residual Vector	146
5.8	Distribution of Variance Estimator	146
5.9	t -Statistic	147

5.10	Confidence Intervals for Regression Coefficients	148
5.11	Confidence Intervals for Error Variance	150
5.12	t -Test	150
5.13	Likelihood Ratio Test	152
5.14	Information Bound for Normal Regression	153
5.15	Exercises	154

PART II LARGE SAMPLE METHODS **157**

6 A REVIEW OF LARGE SAMPLE ASYMPTOTICS **159**

6.1	Introduction	159
6.2	Modes of Convergence	159
6.3	Weak Law of Large Numbers	160
6.4	Central Limit Theorem	160
6.5	Continuous Mapping Theorem and Delta Method	161
6.6	Smooth Function Model	162
6.7	Stochastic Order Symbols	162
6.8	Convergence of Moments	163

7 ASYMPTOTIC THEORY FOR LEAST SQUARES **165**

7.1	Introduction	165
7.2	Consistency of Least Squares Estimator	165
7.3	Asymptotic Normality	167
7.4	Joint Distribution	171
7.5	Consistency of Error Variance Estimators	173
7.6	Homoskedastic Covariance Matrix Estimation	174
7.7	Heteroskedastic Covariance Matrix Estimation	174
7.8	Summary of Covariance Matrix Notation	176
7.9	Alternative Covariance Matrix Estimators*	177
7.10	Functions of Parameters	178
7.11	Asymptotic Standard Errors	180
7.12	t -Statistic	182
7.13	Confidence Intervals	183
7.14	Regression Intervals	185
7.15	Forecast Intervals	187
7.16	Wald Statistic	187
7.17	Homoskedastic Wald Statistic	188
7.18	Confidence Regions	188
7.19	Edgeworth Expansion*	189
7.20	Uniformly Consistent Residuals*	191
7.21	Asymptotic Leverage*	192
7.22	Exercises	192

8 RESTRICTED ESTIMATION**199**

8.1	Introduction	199
8.2	Constrained Least Squares	200
8.3	Exclusion Restriction	201
8.4	Finite Sample Properties	202
8.5	Minimum Distance	205
8.6	Asymptotic Distribution	206
8.7	Variance Estimation and Standard Errors	208
8.8	Efficient Minimum Distance Estimator	208
8.9	Exclusion Restriction Revisited	209
8.10	Variance and Standard Error Estimation	211
8.11	Hausman Equality	211
8.12	Example: Mankiw, Romer, and Weil (1992)	212
8.13	Misspecification	216
8.14	Nonlinear Constraints	218
8.15	Inequality Restrictions	219
8.16	Technical Proofs*	220
8.17	Exercises	221

9 HYPOTHESIS TESTING**225**

9.1	Introduction	225
9.2	Hypotheses	225
9.3	Acceptance and Rejection	226
9.4	Type I Error	227
9.5	t -Tests	228
9.6	Type II Error and Power	229
9.7	Statistical Significance	230
9.8	p -Values	231
9.9	t -Ratios and the Abuse of Testing	232
9.10	Wald Tests	233
9.11	Homoskedastic Wald Tests	235
9.12	Criterion-Based Tests	236
9.13	Minimum Distance Tests	236
9.14	Minimum Distance Tests under Homoskedasticity	237
9.15	F Tests	238
9.16	Hausman Tests	239
9.17	Score Tests	240
9.18	Problems with Tests of Nonlinear Hypotheses	242
9.19	Monte Carlo Simulation	245
9.20	Confidence Intervals by Test Inversion	247
9.21	Multiple Tests and Bonferroni Corrections	248
9.22	Power and Test Consistency	250
9.23	Asymptotic Local Power	251

	9.24	Asymptotic Local Power, Vector Case	254
	9.25	Exercises	255
10		RESAMPLING METHODS	262
	10.1	Introduction	262
	10.2	Example	262
	10.3	Jackknife Estimation of Variance	263
	10.4	Example	266
	10.5	Jackknife for Clustered Observations	267
	10.6	The Bootstrap Algorithm	268
	10.7	Bootstrap Variance and Standard Errors	270
	10.8	Percentile Interval	272
	10.9	The Bootstrap Distribution	273
	10.10	The Distribution of the Bootstrap Observations	274
	10.11	The Distribution of the Bootstrap Sample Mean	275
	10.12	Bootstrap Asymptotics	276
	10.13	Consistency of the Bootstrap Estimate of Variance	279
	10.14	Trimmed Estimator of Bootstrap Variance	280
	10.15	Unreliability of Untrimmed Bootstrap Standard Errors	282
	10.16	Consistency of the Percentile Interval	283
	10.17	Bias-Corrected Percentile Interval	285
	10.18	BC_a Percentile Interval	286
	10.19	Percentile- t Interval	288
	10.20	Percentile- t Asymptotic Refinement	290
	10.21	Bootstrap Hypothesis Tests	292
	10.22	Wald-Type Bootstrap Tests	294
	10.23	Criterion-Based Bootstrap Tests	295
	10.24	Parametric Bootstrap	296
	10.25	How Many Bootstrap Replications?	297
	10.26	Setting the Bootstrap Seed	298
	10.27	Bootstrap Regression	298
	10.28	Bootstrap Regression Asymptotic Theory	300
	10.29	Wild Bootstrap	301
	10.30	Bootstrap for Clustered Observations	303
	10.31	Technical Proofs*	304
	10.32	Exercises	307
		PART III MULTIPLE EQUATION MODELS	313
11		MULTIVARIATE REGRESSION	315
	11.1	Introduction	315
	11.2	Regression Systems	315

11.3	Least Squares Estimator	316
11.4	Expectation and Variance of Systems Least Squares	318
11.5	Asymptotic Distribution	319
11.6	Covariance Matrix Estimation	321
11.7	Seemingly Unrelated Regression	322
11.8	Equivalence of SUR and Least Squares	324
11.9	Maximum Likelihood Estimator	324
11.10	Restricted Estimation	325
11.11	Reduced Rank Regression	325
11.12	Principal Component Analysis	329
11.13	Factor Models	331
11.14	Approximate Factor Models	333
11.15	Factor Models with Additional Regressors	335
11.16	Factor-Augmented Regression	336
11.17	Multivariate Normal*	337
11.18	Exercises	339

12 INSTRUMENTAL VARIABLES

341

12.1	Introduction	341
12.2	Overview	341
12.3	Examples	342
12.4	Endogenous Regressors	344
12.5	Instruments	345
12.6	Example: College Proximity	346
12.7	Reduced Form	347
12.8	Identification	349
12.9	Instrumental Variables Estimator	350
12.10	Demeaned Representation	352
12.11	Wald Estimator	353
12.12	Two-Stage Least Squares	354
12.13	Limited Information Maximum Likelihood	357
12.14	Split-Sample IV and JIVE	359
12.15	Consistency of 2SLS	361
12.16	Asymptotic Distribution of 2SLS	362
12.17	Determinants of 2SLS Variance	363
12.18	Covariance Matrix Estimation	364
12.19	LIML Asymptotic Distribution	366
12.20	Functions of Parameters	367
12.21	Hypothesis Tests	368
12.22	Finite Sample Theory	369
12.23	Bootstrap for 2SLS	369
12.24	The Peril of Bootstrap 2SLS Standard Errors	372
12.25	Clustered Dependence	373
12.26	Generated Regressors	374
12.27	Regression with Expectation Errors	377
12.28	Control Function Regression	380

12.29	Endogeneity Tests	382
12.30	Subset Endogeneity Tests	385
12.31	Overidentification Tests	386
12.32	Subset Overidentification Tests	389
12.33	Bootstrap Overidentification Tests	392
12.34	Local Average Treatment Effects	392
12.35	Identification Failure	396
12.36	Weak Instruments	397
12.37	Many Instruments	400
12.38	Testing for Weak Instruments	404
12.39	Weak Instruments with $k_2 > 1$	410
12.40	Example: Acemoglu, Johnson, and Robinson (2001)	412
12.41	Example: Angrist and Krueger (1991)	414
12.42	Programming	416
12.43	Exercises	418
13	GENERALIZED METHOD OF MOMENTS	424
13.1	Introduction	424
13.2	Moment Equation Models	424
13.3	Method of Moments Estimators	425
13.4	Overidentified Moment Equations	426
13.5	Linear Moment Models	427
13.6	GMM Estimator	427
13.7	Distribution of GMM Estimator	428
13.8	Efficient GMM	429
13.9	Efficient GMM versus 2SLS	430
13.10	Estimation of the Efficient Weight Matrix	430
13.11	Iterated GMM	431
13.12	Covariance Matrix Estimation	432
13.13	Clustered Dependence	432
13.14	Wald Test	433
13.15	Restricted GMM	434
13.16	Nonlinear Restricted GMM	435
13.17	Constrained Regression	436
13.18	Multivariate Regression	436
13.19	Distance Test	438
13.20	Continuously Updated GMM	439
13.21	Overidentification Test	439
13.22	Subset Overidentification Tests	440
13.23	Endogeneity Test	441
13.24	Subset Endogeneity Test	441
13.25	Nonlinear GMM	442
13.26	Bootstrap for GMM	443
13.27	Conditional Moment Equation Models	444
13.28	Technical Proofs*	445
13.29	Exercises	447

PART IV DEPENDENT AND PANEL DATA**453****14 TIME SERIES****455**

14.1	Introduction	455
14.2	Examples	455
14.3	Differences and Growth Rates	456
14.4	Stationarity	458
14.5	Transformations of Stationary Processes	460
14.6	Convergent Series	460
14.7	Ergodicity	461
14.8	Ergodic Theorem	463
14.9	Conditioning on Information Sets	464
14.10	Martingale Difference Sequences	465
14.11	CLT for Martingale Differences	467
14.12	Mixing	468
14.13	CLT for Correlated Observations	469
14.14	Linear Projection	471
14.15	White Noise	471
14.16	The Wold Decomposition	472
14.17	Lag Operator	473
14.18	Autoregressive Wold Representation	473
14.19	Linear Models	474
14.20	Moving Average Process	475
14.21	Infinite-Order Moving Average Process	476
14.22	First-Order Autoregressive Process	477
14.23	Unit Root and Explosive AR(1) Processes	480
14.24	Second-Order Autoregressive Process	481
14.25	AR(p) Process	484
14.26	Impulse Response Function	485
14.27	ARMA and ARIMA Processes	486
14.28	Mixing Properties of Linear Processes	487
14.29	Identification	488
14.30	Estimation of Autoregressive Models	490
14.31	Asymptotic Distribution of Least Squares Estimator	491
14.32	Distribution under Homoskedasticity	492
14.33	Asymptotic Distribution under General Dependence	493
14.34	Covariance Matrix Estimation	493
14.35	Covariance Matrix Estimation under General Dependence	494
14.36	Testing the Hypothesis of No Serial Correlation	496
14.37	Testing for Omitted Serial Correlation	496
14.38	Model Selection	498
14.39	Illustrations	498
14.40	Time Series Regression Models	500
14.41	Static, Distributed Lag, and Autoregressive Distributed Lag Models	501
14.42	Time Trends	502
14.43	Illustration	505

14.44	Granger Causality	505
14.45	Testing for Serial Correlation in Regression Models	507
14.46	Bootstrap for Time Series	508
14.47	Technical Proofs*	510
14.48	Exercises	519

15 MULTIVARIATE TIME SERIES 524

15.1	Introduction	524
15.2	Multiple Equation Time Series Models	524
15.3	Linear Projection	525
15.4	Multivariate Wold Decomposition	525
15.5	Impulse Response	527
15.6	VAR(1) Model	528
15.7	VAR(p) Model	529
15.8	Regression Notation	529
15.9	Estimation	530
15.10	Asymptotic Distribution	531
15.11	Covariance Matrix Estimation	532
15.12	Selection of Lag Length in a VAR	533
15.13	Illustration	533
15.14	Predictive Regressions	533
15.15	Impulse Response Estimation	535
15.16	Local Projection Estimator	537
15.17	Regression on Residuals	537
15.18	Orthogonalized Shocks	539
15.19	Orthogonalized Impulse Response Function	540
15.20	Orthogonalized Impulse Response Estimation	540
15.21	Illustration	541
15.22	Forecast Error Decomposition	542
15.23	Identification of Recursive VARs	543
15.24	Oil Price Shocks	544
15.25	Structural VARs	546
15.26	Identification of Structural VARs	549
15.27	Long-Run Restrictions	550
15.28	Blanchard and Quah (1989) Illustration	551
15.29	External Instruments	553
15.30	Dynamic Factor Models	554
15.31	Technical Proofs*	556
15.32	Exercises	557

16 NONSTATIONARY TIME SERIES 561

16.1	Introduction	561
16.2	Partial Sum Process and Functional Convergence	561

16.3	Beveridge-Nelson Decomposition	563
16.4	Functional CLT	565
16.5	Orders of Integration	566
16.6	Means, Local Means, and Trends	567
16.7	Demeaning and Detrending	569
16.8	Stochastic Integrals	570
16.9	Estimation of an AR(1)	572
16.10	AR(1) Estimation with an Intercept	574
16.11	Sample Covariances of Integrated and Stationary Processes	576
16.12	AR(p) Models with a Unit Root	576
16.13	Testing for a Unit Root	578
16.14	KPSS Stationarity Test	581
16.15	Spurious Regression	584
16.16	NonStationary VARs	588
16.17	Cointegration	589
16.18	Role of Intercept and Trend	593
16.19	Cointegrating Regression	594
16.20	VECM Estimation	597
16.21	Testing for Cointegration in a VECM	599
16.22	Technical Proofs*	603
16.23	Exercises	610
19.26	Technical Proofs*	715
19.27	Exercises	717

17**PANEL DATA****613**

17.1	Introduction	613
17.2	Time Indexing and Unbalanced Panels	614
17.3	Notation	615
17.4	Pooled Regression	615
17.5	One-Way Error Component Model	617
17.6	Random Effects	617
17.7	Fixed Effects Model	620
17.8	Within Transformation	621
17.9	Fixed Effects Estimator	623
17.10	Differenced Estimator	624
17.11	Dummy Variables Regression	626
17.12	Fixed Effects Covariance Matrix Estimation	628
17.13	Fixed Effects Estimation in Stata	629
17.14	Between Estimator	630
17.15	Feasible GLS	632
17.16	Intercept in Fixed Effects Regression	633
17.17	Estimation of Fixed Effects	633
17.18	GMM Interpretation of Fixed Effects	634
17.19	Identification in the Fixed Effects Model	636
17.20	Asymptotic Distribution of Fixed Effects Estimator	636
17.21	Asymptotic Distribution for Unbalanced Panels	637
17.22	Heteroskedasticity-Robust Covariance Matrix Estimation	639
17.23	Heteroskedasticity-Robust Estimation—Unbalanced Case	641

17.24	Hausman Test for Random vs. Fixed Effects	641
17.25	Random Effects or Fixed Effects?	642
17.26	Time Trends	642
17.27	Two-Way Error Components	643
17.28	Instrumental Variables	645
17.29	Identification with Instrumental Variables	646
17.30	Asymptotic Distribution of Fixed Effects 2SLS Estimator	647
17.31	Linear GMM	648
17.32	Estimation with Time-Invariant Regressors	648
17.33	Hausman-Taylor Model	650
17.34	Jackknife Covariance Matrix Estimation	652
17.35	Panel Bootstrap	653
17.36	Dynamic Panel Models	653
17.37	The Bias of Fixed Effects Estimation	654
17.38	Anderson-Hsiao Estimator	656
17.39	Arellano-Bond Estimator	657
17.40	Weak Instruments	659
17.41	Dynamic Panels with Predetermined Regressors	660
17.42	Blundell-Bond Estimator	661
17.43	Forward Orthogonal Transformation	664
17.44	Empirical Illustration	665
17.45	Exercises	666

18**DIFFERENCE IN DIFFERENCES****669**

18.1	Introduction	669
18.2	Minimum Wage in New Jersey	669
18.3	Identification	672
18.4	Multiple Units	673
18.5	Do Police Reduce Crime?	675
18.6	Trend Specification	677
18.7	Do Blue Laws Affect Liquor Sales?	678
18.8	Check Your Code: Does Abortion Impact Crime?	679
18.9	Inference	680
18.10	Exercises	682

PART V NONPARAMETRIC METHODS**685****19****NONPARAMETRIC REGRESSION****687**

19.1	Introduction	687
19.2	Binned Means Estimator	687
19.3	Kernel Regression	689
19.4	Local Linear Estimator	690

19.5	Local Polynomial Estimator	692
19.6	Asymptotic Bias	692
19.7	Asymptotic Variance	694
19.8	AIMSE	695
19.9	Reference Bandwidth	697
19.10	Estimation at a Boundary	698
19.11	Nonparametric Residuals and Prediction Errors	700
19.12	Cross-Validation Bandwidth Selection	701
19.13	Asymptotic Distribution	702
19.14	Undersmoothing	704
19.15	Conditional Variance Estimation	705
19.16	Variance Estimation and Standard Errors	706
19.17	Confidence Bands	707
19.18	The Local Nature of Kernel Regression	707
19.19	Application to Wage Regression	707
19.20	Clustered Observations	709
19.21	Application to Test Scores	710
19.22	Multiple Regressors	712
19.23	Curse of Dimensionality	713
19.24	Partially Linear Regression	714
19.25	Computation	715
19.26	Technical Proofs*	715
19.27	Exercises	720

20 SERIES REGRESSION

723

20.1	Introduction	723
20.2	Polynomial Regression	724
20.3	Illustrating Polynomial Regression	725
20.4	Orthogonal Polynomials	726
20.5	Splines	727
20.6	Illustrating Spline Regression	728
20.7	The Global/Local Nature of Series Regression	729
20.8	Stone-Weierstrass and Jackson Approximation Theory	731
20.9	Regressor Bounds	733
20.10	Matrix Convergence	734
20.11	Consistent Estimation	736
20.12	Convergence Rate	736
20.13	Asymptotic Normality	738
20.14	Regression Estimation	739
20.15	Undersmoothing	739
20.16	Residuals and Regression Fit	740
20.17	Cross-Validation Model Selection	740
20.18	Variance and Standard Error Estimation	742
20.19	Clustered Observations	743
20.20	Confidence Bands	743

20.21	Uniform Approximations	744
20.22	Partially Linear Model	745
20.23	Panel Fixed Effects	745
20.24	Multiple Regressors	746
20.25	Additively Separable Models	746
20.26	Nonparametric Instrumental Variables Regression	746
20.27	NPIV Identification	748
20.28	NPIV Convergence Rate	749
20.29	Nonparametric vs. Parametric Identification	750
20.30	Example: Angrist and Lavy (1999)	751
20.31	Technical Proofs*	754
20.32	Exercises	759
21	REGRESSION DISCONTINUITY	763
21.1	Introduction	763
21.2	Sharp Regression Discontinuity	763
21.3	Identification	764
21.4	Estimation	766
21.5	Inference	767
21.6	Bandwidth Selection	768
21.7	RDD with Covariates	770
21.8	A Simple RDD Estimator	771
21.9	Density Discontinuity Test	772
21.10	Fuzzy Regression Discontinuity	773
21.11	Estimation of FRD	774
21.12	Exercises	775
PART VI NONLINEAR METHODS		777
22	M-ESTIMATORS	779
22.1	Introduction	779
22.2	Examples	779
22.3	Identification and Estimation	780
22.4	Consistency	780
22.5	Uniform Law of Large Numbers	782
22.6	Asymptotic Distribution	783
22.7	Asymptotic Distribution under Broader Conditions*	784
22.8	Covariance Matrix Estimation	785
22.9	Technical Proofs*	786
22.10	Exercises	788

23 NONLINEAR LEAST SQUARES 790

23.1 Introduction 790

23.2 Identification 792

23.3 Estimation 792

23.4 Asymptotic Distribution 794

23.5 Covariance Matrix Estimation 796

23.6 Panel Data 797

23.7 Threshold Models 798

23.8 Testing for Nonlinear Components 802

23.9 Computation 804

23.10 Technical Proofs* 804

23.11 Exercises 805

24 QUANTILE REGRESSION 807

24.1 Introduction 807

24.2 Median Regression 807

24.3 Least Absolute Deviations 809

24.4 Quantile Regression 810

24.5 Example Quantile Shapes 813

24.6 Estimation 814

24.7 Asymptotic Distribution 815

24.8 Covariance Matrix Estimation 817

24.9 Clustered Dependence 818

24.10 Quantile Crossings 819

24.11 Quantile Causal Effects 820

24.12 Random Coefficient Representation 821

24.13 Nonparametric Quantile Regression 822

24.14 Panel Data 823

24.15 IV Quantile Regression 824

24.16 Technical Proofs* 825

24.17 Exercises 827

25 BINARY CHOICE 829

25.1 Introduction 829

25.2 Binary Choice Models 829

25.3 Models for the Response Probability 830

25.4 Latent Variable Interpretation 832

25.5 Likelihood 833

25.6 Pseudo-True Values 835

25.7 Asymptotic Distribution 836

25.8	Covariance Matrix Estimation	838
25.9	Marginal Effects	838
25.10	Application	839
25.11	Semiparametric Binary Choice	840
25.12	IV Probit	841
25.13	Binary Panel Data	842
25.14	Technical Proofs*	844
25.15	Exercises	845
26	MULTIPLE CHOICE	847
26.1	Introduction	847
26.2	Multinomial Response	847
26.3	Multinomial Logit	848
26.4	Conditional Logit	851
26.5	Independence of Irrelevant Alternatives	854
26.6	Nested Logit	855
26.7	Mixed Logit	858
26.8	Simple Multinomial Probit	860
26.9	General Multinomial Probit	861
26.10	Ordered Response	862
26.11	Count Data	864
26.12	BLP Demand Model	865
26.13	Technical Proofs*	868
26.14	Exercises	870
27	CENSORING AND SELECTION	872
27.1	Introduction	872
27.2	Censoring	872
27.3	Censored Regression Functions	874
27.4	The Bias of Least Squares Estimation	875
27.5	Tobit Estimator	876
27.6	Identification in Tobit Regression	877
27.7	CLAD and CQR Estimators	879
27.8	Illustrating Censored Regression	880
27.9	Sample Selection Bias	881
27.10	Heckman's Model	882
27.11	Nonparametric Selection	884
27.12	Panel Data*	885
27.13	Exercises	886

MODEL SELECTION, STEIN SHRINKAGE, AND MODEL AVERAGING **889**

28.1	Introduction	889
28.2	Model Selection	889
28.3	Bayesian Information Criterion	891
28.4	Akaike Information Criterion for Regression	892
28.5	Akaike Information Criterion for Likelihood	894
28.6	Mallows Criterion	895
28.7	Hold-Out Criterion	896
28.8	Cross-Validation Criterion	897
28.9	K-Fold Cross-Validation	899
28.10	Many Selection Criteria Are Similar	900
28.11	Relation with Likelihood Ratio Testing	901
28.12	Consistent Selection	902
28.13	Asymptotic Selection Optimality	904
28.14	Focused Information Criterion	906
28.15	Best Subset and Stepwise Regression	908
28.16	The MSE of Model Selection Estimators	909
28.17	Inference after Model Selection	911
28.18	Empirical Illustration	913
28.19	Shrinkage Methods	914
28.20	James-Stein Shrinkage Estimator	915
28.21	Interpretation of the Stein Effect	916
28.22	Positive Part Estimator	916
28.23	Shrinkage Toward Restrictions	918
28.24	Group James-Stein	919
28.25	Empirical Illustrations	920
28.26	Model Averaging	923
28.27	Smoothed BIC and AIC	925
28.28	Mallows Model Averaging	927
28.29	Jackknife (CV) Model Averaging	929
28.30	Granger-Ramanathan Averaging	930
28.31	Empirical Illustration	931
28.32	Technical Proofs*	932
28.33	Exercises	939

MACHINE LEARNING **941**

29.1	Introduction	941
29.2	Big Data, High Dimensionality, and Machine Learning	941
29.3	High-Dimensional Regression	942
29.4	p-norms	943
29.5	Ridge Regression	944
29.6	Statistical Properties of Ridge Regression	947
29.7	Illustrating Ridge Regression	948

29.8	Lasso	948
29.9	Lasso Penalty Selection	951
29.10	Lasso Computation	952
29.11	Asymptotic Theory for the Lasso	952
29.12	Approximate Sparsity	954
29.13	Elastic Net	955
29.14	Post-Lasso	956
29.15	Regression Trees	956
29.16	Bagging	958
29.17	Random Forests	960
29.18	Ensembling	961
29.19	Lasso IV	962
29.20	Double Selection Lasso	963
29.21	Post-Regularization Lasso	965
29.22	Double/Debiased Machine Learning	967
29.23	Technical Proofs*	968
29.24	Exercises	974

APPENDIXES

975

APPENDIX A: MATRIX ALGEBRA

977

A.1	Notation	977
A.2	Complex Matrices	978
A.3	Matrix Addition	979
A.4	Matrix Multiplication	979
A.5	Trace	980
A.6	Rank and Inverse	981
A.7	Orthogonal and Orthonormal Matrices	982
A.8	Determinant	982
A.9	Eigenvalues	983
A.10	Positive Definite Matrices	984
A.11	Idempotent Matrices	985
A.12	Singular Values	986
A.13	Matrix Decompositions	986
A.14	Generalized Eigenvalues	987
A.15	Extrema of Quadratic Forms	988
A.16	Cholesky Decomposition	990
A.17	QR Decomposition	991
A.18	Solving Linear Systems	991
A.19	Algorithmic Matrix Inversion	993
A.20	Matrix Calculus	993
A.21	Kronecker Products and the Vec Operator	995
A.22	Vector Norms	996
A.23	Matrix Norms	996

APPENDIX B: USEFUL INEQUALITIES

999

B.1	Inequalities for Real Numbers	999
B.2	Inequalities for Vectors	1000
B.3	Inequalities for Matrices	1001
B.4	Probability Inequalities	1001
B.5	Proofs*	1005

References	1021
-----------------------------	------

Index	1033
------------------------	------

Econometrics. The sequence is:

1. *Probability and Statistics for Economists* (first volume).
2. *Econometrics* (this volume).

I assume that students have background in multivariate calculus, probability theory, linear algebra, and mathematical statistics. A prior course in undergraduate econometrics would be helpful but is not required. Two excellent undergraduate textbooks are Stock and Watson (2019) and Wooldridge (2020). The background in probability theory and mathematical statistics is provided in *Probability and Statistics for Economists*.

In addition, the basic tools of matrix algebra and probability inequalities are reviewed in Appendix A.

This textbook contains more material than can be covered in a one-semester course. This is intended to provide maximum flexibility concerning which topics to cover, which to cover in depth, and which to cover briefly. Some material is suitable for second-year Ph.D. instruction. At the University of Wisconsin, where this manual was developed, in the first half of the fall semester, we cover *Probability and Statistics for Economists*; in the second half of the fall semester, we cover Chapters 1–9 of *Econometrics*. In the first half of the spring semester, we cover Chapters 10–17, with some chapters covered briefly. In the second half of the spring semester, we cover Chapters 18–29, with many details only covered briefly. We revisit much of the latter part of the second-year curriculum, with greater focus on the econometric theory.

For students wishing to deepen their knowledge of matrix algebra in relation to econometrics, I recommend *Matrix Algebra* by Abadir and Magnus (2005).

For the study in econometrics beyond this text, I recommend Davidson (1994) for asymptotic theory, Wooldridge (2009) and Kilian and Lutkepohl (2017) for time series methods, Cameron and Trivedi (2005) and Wooldridge (2010) for panel data and discrete response models, and Li and Racine (2007) for nonparametric and semiparametric econometrics. Beyond these texts, the *Handbook of Econometric Series* provides advanced treatments of contemporary econometric methods and theory.

Advanced Ph.D.-level econometrics textbooks include Theil (1971), Amemiya (1985), Judge, Griffiths, Hill, Lutkepohl, and Lee (1985), Goldberger (1991), Davidson and MacKinnon (1993, 2004), Johnston and Dinardo (1997), Davidson (2000), Hayashi (2000), Raud (2000), Greene (2018), and Magnus (2017). For a survey of empirical issues, see Angrist and Pischke (2009) and Cunningham (2021).

Some of the chapter exercises are important parts of the text and are meant to help teach students of