

# CONTENTS

Preface . . . . .	xxv
Acknowledgments . . . . .	xxvii
Notation . . . . .	xxix

## **1 INTRODUCTION 1**

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 11**

## **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 . . . . .	30
2.15	Linear CEF . . . . .	31
2.16	Linear CEF with Nonlinear Effects . . . . .	32
2.17	Linear CEF with Dummy Variables . . . . .	33
2.18	Best Linear Predictor . . . . .	35
2.19	Illustrations of Best Linear Predictor . . . . .	39
2.20	Linear Predictor Error Variance . . . . .	41
2.21	Regression Coefficients . . . . .	41
2.22	Regression Subvectors . . . . .	42
2.23	Coefficient Decomposition . . . . .	43
2.24	Omitted Variable Bias . . . . .	44
2.25	Best Linear Approximation . . . . .	45
2.26	Regression to the Mean . . . . .	46
2.27	Reverse Regression . . . . .	47
2.28	Limitations of the Best Linear Projection . . . . .	48
2.29	Random Coefficient Model . . . . .	48
2.30	Causal Effects . . . . .	50
2.31	Existence and Uniqueness of the Conditional Expectation* . . . . .	55
2.32	Identification* . . . . .	56
2.33	Technical Proofs* . . . . .	57
2.34	Exercises . . . . .	59

### 3 THE ALGEBRA OF LEAST SQUARES

3.1	Introduction . . . . .	62
3.2	Samples . . . . .	62
3.3	Moment Estimators . . . . .	63
3.4	Least Squares Estimator . . . . .	64
3.5	Solving for Least Squares with One Regressor . . . . .	65
3.6	Solving for Least Squares with Multiple Regressors . . . . .	66
3.7	Illustration . . . . .	69
3.8	Least Squares Residuals . . . . .	71
3.9	Demeaned Regressors . . . . .	72
3.10	Model in Matrix Notation . . . . .	73
3.11	Projection Matrix . . . . .	74
3.12	Annihilator Matrix . . . . .	76
3.13	Estimation of Error Variance . . . . .	76
3.14	Analysis of Variance . . . . .	77
3.15	Projections . . . . .	78
3.16	Regression Components . . . . .	78
3.17	Regression Components (Alternative Derivation)* . . . . .	80
3.18	Residual Regression . . . . .	81
3.19	Leverage Values . . . . .	83
3.20	Leave-One-Out Regression . . . . .	84
3.21	Influential Observations . . . . .	86
3.22	CPS Dataset . . . . .	88

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	<i>t</i> -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	<i>t</i> -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**

---

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

## **9 HYPOTHESIS TESTING 225**

---

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

9.24	Asymptotic Local Power, Vector Case . . . . .	254
9.25	Exercises . . . . .	255

---

<b>10</b>	<b>RESAMPLING METHODS</b>	<b>262</b>
-----------	---------------------------	------------

---

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

<b>PART III</b>	<b>MULTIPLE EQUATION MODELS</b>	<b>313</b>
-----------------	---------------------------------	------------

<b>11</b>	<b>MULTIVARIATE REGRESSION</b>	<b>315</b>
-----------	--------------------------------	------------

---

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

## 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**

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



---

**28**      **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

---

**29**      **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>B.1</b>	Inequalities for Real Numbers . . . . .	999
<b>B.2</b>	Inequalities for Vectors . . . . .	1000
<b>B.3</b>	Inequalities for Matrices . . . . .	1001
<b>B.4</b>	Probability Inequalities . . . . .	1001
<b>B.5</b>	Proofs* . . . . .	1005
References . . . . .		1021
Index . . . . .		1033

# CHAPTER 1

## INTRODUCTION

### 1.1 WHAT IS ECONOMETRICS?

The term “econometrics” is believed to have been crafted by Ragnar Frisch (1895–1973) of Norway, one of the three principal founders of the Econometric Society, first editor of the journal *Econometrica*, and co-winner of the first Nobel Memorial Prize in Economic Sciences in 1969. It is therefore fitting that we turn to Frisch’s own words in the introduction to the first issue of *Econometrica* to describe the discipline.

A word of explanation regarding the term econometrics may be in order. Its definition is implied in the statement of the scope of the [Econometric] Society, in Section I of the Constitution, which reads: “The Econometric Society is an international society for the advancement of economic theory in its relation to statistics and mathematics. . . . 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. . . .”

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 synonymous with the application of mathematics to economics. Experience has shown that each of these three view-points, 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. (Frisch, 1933, pp. 1–2).

This definition remains valid today, although some terms have evolved somewhat in their usage. Today, we would say that econometrics is the unified study of economic models, mathematical statistics, and economic data.

In the field of econometrics there are subdivisions and specializations. **Econometric theory** concerns the development of tools and methods, and the study of the properties of econometric methods. **Applied econometrics** is a term describing the development of quantitative economic models and the application of econometric methods to these models using economic data.

### 1.2 THE PROBABILITY APPROACH TO ECONOMETRICS

The unifying methodology of modern econometrics was articulated by Trygve Haavelmo (1911–1999) in his seminal paper “The probability approach in econometrics” (1944). Haavelmo argued that quantitative economic models must necessarily be *probability models* (by which today we would mean *stochastic*). Deterministic models are blatantly inconsistent with observed economic quantities, and it is incoherent to apply

deterministic models to nondeterministic data. Economic models should be explicitly designed to incorporate randomness; stochastic errors should not be simply added to deterministic models to make them random. Once we acknowledge that an economic model is a probability model, it follows naturally that an appropriate tool to quantify, estimate, and conduct inferences about the economy is the powerful theory of mathematical statistics. The appropriate method for a quantitative economic analysis follows from the probabilistic construction of the economic model. Haavelmo's probability approach was quickly embraced by the economics profession. Today, no quantitative work in economics shuns its fundamental vision.

While all economists embrace the probability approach, there has been some evolution in its implementation.

The **structural approach** is the closest to Haavelmo's original idea. A probabilistic economic model is specified, and the quantitative analysis performed under the assumption that the economic model is correctly specified. Researchers often describe this as "taking their model seriously." The structural approach typically leads to likelihood-based analysis, including maximum likelihood and Bayesian estimation.

A criticism of the structural approach is that it is misleading to treat an economic model as correctly specified. Instead, it is more accurate to view a model as a useful abstraction or approximation. In this case, how should we interpret structural econometric analysis? The **quasi-structural approach** to inference views a structural economic model as an approximation rather than the truth. This theory has led to the concepts of the pseudo-true value (the parameter value defined by the estimation problem), the quasi-likelihood function, quasi-maximum likelihood estimate (quasi-MLE), and quasi-likelihood inference.

Closely related is the **semiparametric approach**. A probabilistic economic model is partially specified but some features are left unspecified. This approach typically leads to estimation methods, such as least squares and the generalized method of moments. The semiparametric approach dominates contemporary econometrics and is the main focus of this textbook.

Another branch of quantitative structural economics is the **calibration approach**. Similar to the quasi-structural approach, the calibration approach interprets structural models as approximations and hence inherently false. The difference is that the calibrationist literature rejects mathematical statistics (deeming classical theory as inappropriate for approximate models) and instead selects parameters by matching model and data moments using nonstatistical ad hoc<sup>1</sup> methods.

### Trygve Haavelmo

The founding ideas of the field of econometrics are largely due to the Norwegian econometrician Trygve Haavelmo (1911–1999). His advocacy of probability models revolutionized the field, and his use of formal mathematical reasoning laid the foundation for subsequent generations. He was awarded the Nobel Memorial Prize in Economic Sciences in 1989.

## 1.3 ECONOMETRIC TERMS

In a typical application, an econometrician has a set of repeated measurements on a set of variables. For example, in a labor application, the variables could include weekly earnings, educational attainment, age, and other descriptive characteristics. We call this information the **data, dataset, or sample**.

<sup>1</sup>*Ad hoc* means "for this purpose"—a method designed for a specific problem—and not based on a generalizable principle.

We use the term **observations** to refer to distinct repeated measurements on the variables. An individual observation often corresponds to a specific economic unit, such as a person, household, corporation, firm, organization, country, state, city, or other geographical region. An individual observation could also be a measurement at a point in time, such as quarterly gross domestic product (GDP) or a daily interest rate.

Economists typically denote variables by the italicized roman characters  $Y$ ,  $X$ , and/or  $Z$ . The convention in econometrics is to use the character  $Y$  to denote the variable to be explained, while the characters  $X$  and  $Z$  are used to denote the conditioning (explanatory) variables. Following mathematical practice, random variables and vectors are denoted by uppercase roman characters, such as  $Y$  and  $X$ . We make an exception for equation errors, which we typically denote by the lowercase letters  $e$ ,  $u$ , or  $v$ .

Real numbers (elements of the real line  $\mathbb{R}$ , also called **scalars**) are written using lowercase italics, such as  $x$ . Vectors (elements of  $\mathbb{R}^k$ ) are typically also written using lowercase italics, such as  $x$ , or using lowercase bold italics, such as  $\mathbf{x}$ . We use bold in matrix algebraic expressions for compatibility with matrix notation.

Matrices are written using uppercase bold italics, such as  $\mathbf{X}$ . Our notation will not make a distinction between random and nonrandom matrices. Typically we use  $\mathbf{U}$ ,  $\mathbf{V}$ ,  $\mathbf{X}$ ,  $\mathbf{Y}$ ,  $\mathbf{Z}$  to denote random matrices and use  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ ,  $\mathbf{W}$  to denote nonrandom matrices.

We denote the number of observations by the natural number  $n$ , and subscript the variables by the index  $i$  to denote the individual observation (e.g.,  $Y_i$ ). In some contexts, we use indices other than  $i$ , such as in time series applications where the index  $t$  is common. In panel studies, we typically use the double index  $it$  to refer to individual  $i$  at time period  $t$ .

We typically use Greek letters, such as  $\beta$ ,  $\theta$ , and  $\sigma^2$ , to denote unknown parameters (scalar or vectors). Parameter matrices are written using uppercase Latin boldface (e.g.,  $\mathbf{A}$ ). Estimators are typically denoted by putting a hat “^,” tilde “~,” or bar “-” over the corresponding letter (e.g.,  $\hat{\beta}$  and  $\tilde{\beta}$  are estimators of  $\beta$ , and  $\hat{\mathbf{A}}$  is an estimator of  $\mathbf{A}$ ).

The covariance matrix of an econometric estimator will typically be written using the uppercase boldface  $\mathbf{V}$ , often with a subscript to denote the estimator (e.g.,  $\mathbf{V}_{\hat{\beta}} = \text{var}[\hat{\beta}]$  as the covariance matrix for  $\hat{\beta}$ ). Hopefully without causing confusion, we will use the notation  $\mathbf{V}_{\beta} = \text{avar}[\hat{\beta}]$  to denote the asymptotic covariance matrix of  $\sqrt{n}(\hat{\beta} - \beta)$  (the variance of the asymptotic distribution). Covariance matrix estimators will be denoted by appending hats or tildes (e.g.,  $\hat{\mathbf{V}}_{\beta}$  is an estimator of  $\mathbf{V}_{\beta}$ ).

## 1.4 OBSERVATIONAL DATA

A common econometric question is to quantify the causal impact of one set of variables on another variable. For example, a concern in labor economics is the returns to schooling—the change in earnings induced by increasing a worker’s education, holding other variables constant. Another issue of interest is the earnings gap between men and women.

Ideally, we would use **experimental** data to answer these questions. To measure the returns to schooling, an experiment might randomly divide children into groups, mandate different levels of education to the different groups, and then follow the children’s wage path after they mature and enter the labor force. The differences between the groups would be direct measurements of the effects of different levels of education. However, experiments such as this would be widely condemned as immoral! Consequently, in economics, experimental data sets are typically narrow in scope.

Instead, most economic data is **observational**. To continue the above example, through data collection we can record the level of a person’s education and their wage. With such data, we can measure the joint distribution of these variables and assess their joint dependence. But from observational data, it is difficult to

infer **causality** as we are not able to manipulate one variable to see the direct effect on the other. For example, a person's level of education is (at least partially) determined by that person's choices. These factors are likely to be affected by their personal abilities and attitudes toward work. The fact that a person is highly educated suggests a high level of ability, which suggests a high relative wage. This is an alternative explanation for an observed positive correlation between educational levels and wages. High ability individuals do better in school, and therefore choose to attain higher levels of education, and their high ability is the fundamental reason for their high wages. The point is that multiple explanations are consistent with a positive correlation between schooling levels and education. Knowledge of the joint distribution alone may not be able to distinguish between these explanations.

Most economic data sets are observational, not experimental. Thus, all variables must be treated as random and possibly jointly determined.

This discussion shows that it is difficult to infer causality from observational data alone. Causal inference requires identification, which is based on strong assumptions. We will discuss these issues on occasion throughout the text.

## 1.5 STANDARD DATA STRUCTURES

There are five major types of economic data sets: cross-sectional, time series, panel, clustered, and spatial. They are distinguished by the dependence structure across observations.

Cross-sectional data sets have one observation per individual. Surveys and administrative records are a typical source for cross-sectional data. In typical applications, the individuals surveyed are persons, households, firms, or other economic agents. In many contemporary econometric cross-sectional studies, the sample size  $n$  is quite large. It is conventional to assume that cross-sectional observations are mutually independent. Most of this text is devoted to the study of cross-sectional data.

Time series data are indexed by time. Typical examples include macroeconomic aggregates, prices, and interest rates. This type of data is characterized by serial dependence. Most aggregate economic data are only available at a low frequency (annual, quarterly, or monthly), so the sample size is typically much smaller than in cross-sectional studies. An exception is financial data, where data are available at a high frequency (daily, hourly, or by transaction), so sample sizes can be quite large.

Panel data combines elements of cross-sectional and time series. These data sets consist of a set of individuals (typically persons, households, or corporations) measured repeatedly over time. The common modeling assumption is that the individuals are mutually independent of one another, but a given individual's observations are mutually dependent. In some panel data contexts, the number of time series observations  $T$  per individual is small, while the number of individuals  $n$  is large. In other panel data contexts (for example when countries or states are taken as the unit of measurement), the number of individuals  $n$  can be small, while the number of time series observations  $T$  can be moderately large. An important issue in econometric panel data is the treatment of error components.

Clustered samples are becoming increasingly popular in applied economics and are related to panel data. In clustered sampling, the observations are grouped into "clusters" which are treated as mutually independent yet allowed to be dependent within the cluster. The major difference from panel data is that clustered sampling

typically does not explicitly model error component structures, nor the dependence within clusters, but is instead concerned with inference which is robust to arbitrary forms of within-cluster correlation.

Spatial dependence is another model of interdependence. The observations are treated as mutually dependent according to a spatial measure (for example, geographic proximity). Unlike clustering, spatial models allow all observations to be mutually dependent and typically rely on explicit modeling of the dependence relationships. Spatial dependence can also be viewed as a generalization of time series dependence.

### Data Structures

- Cross-section
- Time series
- Panel
- Clustered
- Spatial

As mentioned above, most of this text will be devoted to cross-sectional data under the assumption of mutually independent observations. By mutual independence, we mean that the  $i$ th observation  $(Y_i, X_i)$  is independent of the  $j$ th observation  $(Y_j, X_j)$  for  $i \neq j$ . In this case, we say that the data are **independently distributed**. (Sometimes the label “independent” is misconstrued. It is a statement about the relationship between observations  $i$  and  $j$ , not a statement about the relationship between  $Y_i$  and  $X_i$ .)

Furthermore, if the data are randomly gathered, it is reasonable to model each observation as a draw from the same probability distribution. In this case, we say that the data are **identically distributed**. If the observations are mutually independent and identically distributed, we say that the observations are **independent and identically distributed (i.i.d.)** or a **random sample**. For most of this text, we will assume that our observations come from a random sample.

**Definition 1.1** The variables  $(Y_i, X_i)$  are a **sample** from the distribution  $F$  if they are identically distributed with distribution  $F$ .

**Definition 1.2** The variables  $(Y_i, X_i)$  are a **random sample** if they are mutually independent and identically distributed (i.i.d.) across  $i = 1, \dots, n$ .

In the random sampling framework, we think of an individual observation  $(Y_i, X_i)$  as a realization from a joint probability distribution  $F(y, x)$ , which we call the **population**. This “population” is infinitely large. This abstraction can be a source of confusion, as it does not correspond to a physical population in the real world. It is an abstraction because the distribution  $F$  is unknown, and the goal of statistical inference is to learn about features of  $F$  from the sample. The *assumption* of random sampling provides the mathematical foundation for treating economic statistics with the tools of mathematical statistics.



The random sampling framework was a major intellectual breakthrough of the late nineteenth century, allowing the application of mathematical statistics to the social sciences. Before this conceptual development, methods from mathematical statistics had not been applied to economic data, as the latter were viewed as nonrandom. The random sampling framework enabled economic samples to be treated as random, a necessary precondition for the application of statistical methods.

## 1.6 ECONOMETRIC SOFTWARE

Economists use a variety of econometric, statistical, and programming software. Stata is a powerful statistical program with a broad set of pre-programmed econometric and statistical tools. It is quite popular among economists and is continuously being updated with new methods. It is an excellent package for most econometric analysis but is limited when you want to use new or less-common econometric methods which have not yet been programmed. At many points in this textbook, specific Stata estimation methods and commands are described. These commands are valid for Stata version 16.

MATLAB, GAUSS, and OxMetrics are high-level matrix programming languages with a wide variety of built-in statistical functions. Many econometric methods have been programmed in these languages and are available on the web. The advantage of these packages is that you are in complete control of your analysis, and it is easier to program new methods than it is in Stata. Some disadvantages are that you have to do much of the programming yourself, programming complicated procedures takes significant time, and programming errors are hard to prevent and difficult to detect and eliminate. Of these languages, GAUSS used to be quite popular among econometricians, but currently MATLAB is more popular.

An intermediate choice is R. R has the capabilities of the above high-level matrix programming languages, but it also has many built-in statistical environments which can replicate much of the functionality of Stata. R is the dominant programming language in the statistics field, so methods developed in that arena are most commonly available in R. Uniquely, R is open source, user contributed, and best of all, completely free! A growing group of econometricians are enthusiastic fans of R.

For highly intensive computational tasks, some economists write their programs in a standard programming language, such as Fortran or C. This can lead to major gains in computational speed, at the cost of increased time in programming and debugging.

Many other packages are used by econometricians, including Eviews, Gretl, PcGive, Python, Julia, RATS, and SAS.

As the packages described above have distinct advantages, many empirical economists use multiple packages. As a student of econometrics, you will learn at least one of these packages and probably more than one. My advice is that all students of econometrics should develop a basic level of familiarity with Stata, MATLAB, and R.

## 1.7 REPLICATION

Scientific research needs to be documented and replicable. For social science research using observational data, this requires careful documentation and archiving of the research methods, data manipulations, and coding.

The best practice is as follows. Accompanying each published paper, an author should create a complete replication package (set of data files, documentation, and program code files). This package should contain the source (raw) data used for analysis and code which executes the empirical analysis and other numerical

work reported in the paper. In most cases, this code is a set of programs which may need to be executed sequentially. (For example, there may be an initial program which “cleans” and manipulates the data, and then a second set of programs which estimate the reported models.) The ideal is full documentation and clarity. This package should be posted on the author(s) website and posted on the journal website when that is an option.

A complicating factor is that many current economic data sets have restricted access and cannot be shared without permission. In these cases, the data cannot be posted or shared. The computed code, however, can and should be posted.

Most journals in economics require authors of published papers to make their datasets generally available. For example:

*Econometrica* states:

*Econometrica* has the policy that all empirical, experimental and simulation results must be replicable. Therefore, authors of accepted papers must submit data sets, programs, and information on empirical analysis, experiments and simulations that are needed for replication and some limited sensitivity analysis.

The *American Economic Review* states (on its webpage):

It is the policy of the American Economic Association to publish papers only if the data and code used in the analysis are clearly and precisely documented and access to the data and code is non-exclusive to the authors. Authors of accepted papers that contain empirical work, simulations, or experimental work must provide, prior to acceptance, information about the data, programs, and other details of the computations sufficient to permit replication, as well as information about access to data and programs.

The *Journal of Political Economy* states:

It is the policy of the *Journal of Political Economy* to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication.

If you are interested in using the data from a published paper, first check the journal’s website, as many journals archive data and replication programs online. Second, check the website(s) of the paper’s author(s). Most academic economists maintain webpages, and some make available replication files complete with data and programs. If these investigations fail, email the author(s), politely requesting the data. You may need to be persistent.

As a matter of professional etiquette, all authors absolutely have the obligation to make their data and programs available. Unfortunately, many fail to do so, and typically for poor reasons. The irony of the situation is that it is typically in the best interests of a scholar to make as much of their work (including all data and programs) freely available, as this only increases the likelihood of their work being cited and having an impact.

Keep this in mind as you start your own empirical project. Remember that as part of your end product, you will need (and want) to provide all data and programs to the community of scholars. The greatest form of flattery is to learn that another scholar has read your paper, wants to extend your work, or wants to use your empirical methods. In addition, public openness provides a healthy incentive for transparency and integrity in empirical analysis.

## 1.8 DATA FILES FOR TEXTBOOK

On the textbook webpage <https://press.princeton.edu/books/econometrics> are posted files containing data sets which are used in this textbook both for illustration and for end-of-chapter empirical exercises. Most of the datasets have four files: (1) Description (pdf format); (2) Excel data file; (3) Text data file; and (4) Stata data

file. The three data files are identical in content: the observations and variables are listed in the same order in each, and all have variable labels.

For example, the text makes frequent reference to a wage dataset extracted from the Current Population Survey. This dataset is named `cps09mar`, and is represented by the files `cps09mar_description.pdf`, `cps09mar.xlsx`, `cps09mar.txt`, and `cps09mar.dta`.

The datasets currently included are

- `AB1991`  
—Data file from Arellano and Bond (1991)
- `AJR2001`  
—Data file from Acemoglu, Johnson, and Robinson (2001)
- `AK1991`  
—Data file from Angrist and Krueger (1991)
- `AL1999`  
—Data file from Angrist and Lavy (1999)
- `BMN2016`  
—Data file from Bernheim, Meer, and Novarro (2016)
- `cps09mar`  
—household survey data extracted from the March 2009 Current Population Survey
- `Card1995`  
—Data file from Card (1995)
- `CHJ2004`  
—Data file from Cox, Hansen, and Jimenez (2004)
- `CK1994`  
—Data file from Card and Krueger (1994)
- `CMR2008`  
—Date file from Card, Mas, and Rothstein (2008)
- `DDK2011`  
—Data file from Duflo, Dupas, and Kremer (2011)
- `DS2004`  
—Data file from DiTella and Scharfrodsky (2004)
- `FRED-MD` and `FRED-QD`  
—U.S. monthly and quarterly macroeconomic databases from McCracken and Ng (2016, 2021)
- `Invest1993`  
—Data file from B. Hall and R. Hall (1993)
- `LM2007`  
—Data file from Ludwig and Miller (2007) and Cattaneo, Titiunik, and Vazquez-Bare (2017)
- `Kilian2009`  
—Data file from Kilian (2009)

- Koppelman
  - Data file from Forinash and Koppelman (1993), Koppelman and Wen (2000), and Wen and Koppelman (2001)
- MRW1992
  - Data file from Mankiw, Romer, and Weil (1992)
- Nerlove1963
  - Data file from Nerlov (1963)
- PSS2017
  - Data file from Papageorgiou, Saam, and Schulte (2017)
- RR2010
  - Data file from Reinhard and Rogoff (2010)

## 1.9 READING THE BOOK

I have endeavored to use a unified notation and nomenclature. The development of the material is cumulative, with later chapters building on the earlier ones. Nevertheless, every attempt has been made to make each chapter self-contained, so readers can pick and choose topics according to their interests.

To fully understand econometric methods, it is necessary to have a mathematical understanding of its mechanics, and this includes the mathematical proofs of the main results. Consequently, this text is self-contained with nearly all results proved with full mathematical rigor. The mathematical development and proofs aim at brevity and conciseness (sometimes described as mathematical elegance), but also at pedagogy. To understand a mathematical proof, it is not sufficient to simply *read* the proof, you need to follow it and re-create it for yourself.

Nevertheless, many readers will not be interested in each mathematical detail, explanation, or proof. This is okay. To use a method, it may not be necessary to understand the mathematical details. Accordingly, I have placed the more technical mathematical proofs and details in chapter appendices. These appendices and other technical sections are marked with an asterisk (\*). These sections can be skipped without any loss in exposition.

Key concepts in matrix algebra and a set of useful inequalities are reviewed in Appendices A and B. It may be useful to read or review Appendix A.1–A.11 before starting Chapter 3, and review Appendix B before Chapter 6. It is not necessary to understand all the material in the appendices. They are intended to be reference material, and some of the results are not used in this textbook.

# INDEX

- abortion and crime difference in differences, 679–680  
absolute  $t$ -ratio, 183  
acceptance and rejection regions, hypothesis testing, 226–227  
Acemoglu, Daron, 412–414  
addition, matrix, 979  
additively separable models, series regression, 746  
adjusted R-squared, 121  
aggregate supply shock, 544  
Akaike information criterion (AIC), 498, 890; for likelihood, 894–895; for regression, 892–894; relation with likelihood ratio testing, 901; smoothed, 925–927  
algorithmic matrix inversion, 993  
alternative hypotheses, 225–226  
*American Economic Review*, 7  
analysis of variance, 77–78  
Anderson-Hsiao estimator, 656  
Angrist, Joshua D., 414–416, 751–753  
annihilator matrix, 76  
applied econometrics, 1  
approximate factor models, 333–335  
approximate sparsity, 954–955  
Arellano-Bond estimator, 657–659  
asymptotic bias, nonparametric regression, 692–694  
asymptotic critical value, 228  
asymptotic distribution, 167–171, 206–207; binary choice, 836–837; of fixed effects 2SLS estimator, 647–648; fixed effects estimator, 636–637; under general dependence, 493; least squares estimator, 491–492; limited likelihood maximum likelihood (LIML), 366–367;  $m$ -estimators, 783–785; multivariate regression, 319–321; multivariate time series, 531–532; nonlinear least squares (NLS), 794–796; nonparametric regression, 702–704; quantile regression, 815–816; two-stage least squares, 362–363; for unbalanced panels, 637–639  
asymptotic integrated MSE (AIMSE), 695–696; curse of dimensionality, 713  
asymptotic leverage, 192  
asymptotic local power, 251–253; vector case, 254  
asymptotic MSE (AMSE), 695–696  
asymptotic normality, 167–171; series regression, 738–739  
asymptotic null distribution, 227  
asymptotic refinement, 290–292  
asymptotics: bootstrap, 276–279; central limit theorem (CLT), 160–161; continuous mapping theorem (CMT) and delta method, 161–162; convergence of moments, 163–164; introduction, 159; modes of convergence, 159; smooth function model, 162; stochastic order symbols, 162–163; weak law of large numbers (WLLN), 160  
asymptotic selection optimality, 904–906  
asymptotic theory, bootstrap regression, 300–301  
asymptotic theory for least squares: alternative covariance matrix estimators, 177–178; asymptotic leverage, 192; asymptotic normality, 167–171; asymptotic standard errors, 180–182; confidence intervals, 183–185; confidence regions, 188–189; consistency of error variance estimator, 173; consistency of least squares estimator, 165–167; Edgeworth expansion, 189–190; forecast intervals, 187; functions of parameters, 178–180; heteroskedastic covariance matrix estimation, 174–176; homoskedastic covariance matrix estimation, 174; homoskedastic Wald statistic, 188; joint distribution, 171–173; regression intervals, 185–187; summary of covariance matrix notation, 176;  $t$ -statistic, 182–183; uniformly consistent residuals, 191; Wald statistic, 187–188  
asymptotic variance, nonparametric regression, 694–695  
attenuation bias, 342  
autoregressive AR(1) parameter estimation, 572–574; with an intercept, 574–576  
autoregressive AR[p] models with unit root, 576–578  
autoregressive distributed lag models (AR-DL), 501–502, 524; Granger causality, 505–507  
autoregressive-integrated moving-average process (ARIMA), 487  
autoregressive models, estimation of, 490–491  
autoregressive-moving-average process (ARMA), 486–487  
autoregressive Wold representation, 473–474  
average causal effect (ACE), 51–53, 394  
average marginal effect, 839  
average treatment effect (ATE), 763  
backward stepwise regression, 908–909  
bagging, 958–960  
Bahr-Esseen inequality, 1003, 1011–1012  
balanced regression design, 83  
bandwidth, 494; cross-validation bandwidth selection, 701–704; reference, nonparametric regression, 697–698; regression discontinuity, 768–770  
basis transformations, 725  
Basmann's statistic, 387  
Bayesian information criterion, 890, 891–892; relation with likelihood ratio testing, 901; smoothed, 925–927

- Bernstein's inequality, 1003, 1009–1011  
Berry Levinsohn Pakes, 865–867  
best linear approximation, 45–46  
best linear predictor, 35–39; illustrations of, 39–41  
best linear projection, limitations of, 48  
best subset model, 908–909  
between estimator, 630–631  
Beveridge-Nelson decomposition, 563–565  
bias: endogeneity or estimation, 342; sample selection, 881–882; smoothing, 693, 699  
bias-corrected percentile interval, 285–286; bootstrap accelerated, 286–288  
big data, 941  
Billingsley's inequality, 1005, 1019  
binary choice: application, 839–840; asymptotic distribution, 836–837; covariance matrix estimation, 838; IV probit, 841–842; latent variable interpretation, 832–833; likelihood, 833–835; marginal effects, 838–839; models, 829–830; models for response probability, 830–832; panel data, 842–843; pseudo-true values, 835–836; semiparametric, 840–841  
binary variables, 33  
binned means estimator, 687–688  
bivariate regression, 71  
Blanchard and Quah, 551–553  
block bootstrap for time series, 510  
BLP model, 865–867  
blue laws difference in differences, 678–679  
Blundell-Bond estimator, 661–664  
Bonferroni corrections, 248–249  
Bonferroni's inequality, 1002, 1008  
Boole's inequality, 1001, 1008  
bootstrap estimator, 262, 268–270; bias-corrected percentile interval, 285–286; bootstrap accelerated bias-corrected percentile interval, 286–288; bootstrap asymptotics, 276–279; bootstrap distribution, 273–274; bootstrap regression, 298–299; bootstrap regression asymptotic theory, 300–301; for clustered observations, 303–304; consistency of bootstrap estimate of variance, 279–280; criterion-based tests, 295–296; distribution of bootstrap observations, 274–275; distribution of bootstrap sample mean, 275–276; how many replications to use, 297; hypothesis tests, 292–294; panel, 653; parametric, 296–297; percentile interval, 272–273, 283–284; percentile- $t$  asymptotic refinement, 290–292; percentile- $t$  interval, 288–290; setting the bootstrap seed, 298; for time series, 508–510; trimmed estimator of bootstrap variance, 280–282; for two-stage least squares, 369–373; unreliability of untrimmed bootstrap standard errors, 282; variance and standard errors, 270–272; Wald-type tests, 294–295; wild bootstrap, 301–302  
bootstrap for generalized method of moments (GMM), 443–444  
bootstrap overidentification tests, 392  
bootstrap regression, 298–299; asymptotic theory, 300–301  
boundary, estimation at a, 698–700  
Box-Cox regression model, 790, 793  
Brownian Bridge, 570; second-level, 583  
Brownian motion, 572  
C (programming language), 6  
 $c_2$  inequality, 1000, 1006  
calculus, matrix, 993–995  
calibration approach, 2  
Cauchy-Schwarz inequality, 1002, 1009  
causal effects, 50–55; quantile, 820–821  
causality, 4  
censored least absolute deviations (CLAD), 872, 879–880  
censored quantile regression (CQR), 880  
censored regression, 872–874; functions, 874–875; illustrating, 880–881; panel data, 885–886  
censoring: from below, 872; bias of least squares estimation and, 875–876; censored least absolute deviations (CLAD), 872, 879–880; Heckman's model, 882–884  
central limit theorem (CLT), 160–161; asymptotic distribution under general dependence, 493; bootstrap, 277; for correlated observations, 469–471; functional, 565–566; for martingale differences, 467–468  
Chebyshev's inequality, 1003, 1009  
Cholesky decomposition, 990–991  
classical measurement error, 342  
classification and regression trees, 956–958  
clustered dependence, 98, 132; generalized method of moments (GMM), 432–433; instrumental variables, 373; quantile regression, 818–819  
clustered observations: bootstrap for, 303–304; jackknife for, 267–268; nonparametric regression, 709–710; series regression, 743  
clustered sampling, 4–5, 124–130; inference with, 131; levels of, 131–132  
cluster-robust jackknife estimator of variance, 268  
coefficient decomposition, 43  
coefficient heterogeneity, 814  
coefficient of determination, 78  
cointegration, 589–593; regression, 594–597; testing for, in a VECM, 599–602  
collinearity errors, 89–91  
complex matrices, 978–979  
conditional expectation function (CEF), 15–17, 19–20; best linear approximation, 45–46; best linear predictor, 35–41; best linear projection limitations, 48; best predictor, 26–27; causal effects, 50–55; CEF error, 23–25; coefficient decomposition, 43; conditional expectation (conditional mean), 15–17; conditional variance, 27–29; continuous variables, 20–21; continuous variables and, 21; distribution of wages, 13–15; error, 23–25; existence and uniqueness of, 55–56; homoskedasticity and heteroskedasticity, 29–30; identification of, 56–57; intercept-only model, 25; law of iterated expectations, 21–23; linear, 31–32; with dummy variables, 33–35; with non linear effects, 32; linear predictor error variance, 41; logs and percentages, 17–19; nonparametric regression, 705–706; omitted variable bias, 44–45; random coefficient model, 48–49; regression coefficients, 41–42; regression derivative, 30–31; regression discontinuity, 766–767; regression subvectors, 42–43; regression to the mean, 46–47; regression variance, 25–26; reverse regression, 47–48  
conditional expectation inequality, 1002, 1009  
conditional independence assumption (CIA), 53–54

- conditional Jensen's inequality, 1002, 1008  
conditional likelihood estimator, 843  
conditional logit, 851–854  
conditional moment equation models, GMM, 444–445  
conditional standard deviation, 27–28  
conditional variance, 27–29; nonparametric regression, 705–706  
conditioning on information sets, 464–465  
conditioning variables, 13  
confidence bands: nonparametric regression, 707; series regression, 743–744  
confidence intervals: asymptotic theory for least squares, 183–185; for error variance, 150; normal-approximation bootstrap, 271; pointwise, 707; for regression coefficients, 148–149; by test inversion, 247–248  
confidence regions, 188–189  
consistent estimation: m-estimators, 780–782; series regression, 736  
constant elasticity of substitution (CES), 790–791  
constrained least squares (CLS) estimator, 200–201; asymptotic distribution, 206–207; efficient minimum distance estimator, 208–209; exclusion restriction, 201–202; finite sample properties, 202–205  
constrained optimization methods, 218–219  
constrained regression, generalized method of moments (GMM), 436  
continuously updated generalized method of moments (GMM), 439  
continuous mapping theorem (CMT), 161–162; bootstrap, 277, 278, 279, 300; consistency of least squares estimator, 165–167  
continuous variables, 20–21  
control function regression, 380–382  
convergence, modes of, 159; bootstrap, 276–277  
convergence rate, 736–737; nonparametric instrumental variables regression (NPIV), 749–750  
convergent series, 460–461  
converging population, 46  
count data, 864–865  
*Course in Econometrics*, A, 124  
covariance matrix estimation, 364–366; alternative, 177–178; binary choice, 838; fixed effects, 628–629; generalized method of moments (GMM), 432; under heteroskedasticity, 113–116, 174–176; heteroskedasticity-robust, 639–641; under homoskedasticity, 113, 174; jackknife, 652–653, 929–930; m-estimators, 785–786; multivariate regression, 321; multivariate time series, 532–533; nonlinear least squares (NLLS), 796–797; notation, 176; quantile regression, 817–818; time series, 493–496  
covariates, 13; regression discontinuity design (RDD) with, 770–771  
Cramér-Rao lower bound, 153–154  
criterion-based bootstrap tests, 295–296  
criterion-based statistic, 236  
critical value, 151; hypothesis testing, 227  
cross-sectional data, 4  
cross-validation, 890; bandwidth selection, 701–702; criterion, 897–898; K-fold, 899–900, 951; series regression, 740–741  
cumulative impulse response function (CIRF), 528  
Current Population Survey (CPS), 88  
curse of dimensionality, 713  
data, 2; big, 941; count, 864–865; cross-sectional, 4; experimental, 3; identically distributed, 5; independently distributed, 5; observational, 3–4; panel (*See* panel data); replication of, 6–7; textbook files, 7–9; time series, 4  
data-generating process, 63  
data sets, 2; CPS, 88; standard structures, 4–6  
decompositions, matrix, 986–987; Cholesky, 990–991; QR, 991  
delta method, 161–162; bootstrap, 278, 300  
demeaned regressors, 72  
demeaned representation, 352  
demeaned values, 76  
demeaning and detrending, nonstationary time series, 569–570  
density discontinuity test, 772–773  
dependent variables, 13  
design matrix, 38  
determinant, 982–983  
Dickey-Fuller coefficient distribution, 573–576  
differenced estimator, 624–626, 669–670  
difference in differences: abortion and crime, 679–680; blue laws and liquor sales, 678–679; do police reduce crime question, 675–677; identification, 672–673; inference, 680–681; minimum wage in New Jersey, 669–672; multiple units, 673–675; trend specification, 677–678  
differences, time series, 456–458  
dimensionality, high, 941  
dissimilarity parameter, 849  
distance test, generalized method of moments (GMM), 438–439  
distributed lag models, 501–502  
distribution: asymptotic (*See* asymptotic distribution); bootstrap, 273–275; convergences in, 159; Dickey-Fuller coefficient, 573–576; GMM estimator, 428–429; joint asymptotic, 171–173; Mammen, 302; OLS coefficient vector, 145; OLS residual vector, 146; standard normal, 139–141; multivariate, 141–142; time series, under homoskedasticity, 492–493; variance estimator, 146–147  
double/debiased machine learning (DML), 967–968  
double selection Lasso, 963–965  
drawing with replacement, 269  
dummy variables, linear CEF with, 33–35  
dummy variables regression, 626–627  
dynamic factor models, 554–555  
dynamic panel models, 653–654; with predetermined regressors, 660–661  
*Econometrica*, 1, 7  
econometrics: defined, 1; probability approach to, 1–2; terms in, 2–3  
econometric software, 6  
econometric theory, 1  
Edgeworth expansion, 189–190  
efficient generalized method of moments (GMM), 429–430  
efficient minimum distance estimator, 208–209  
Eicker-White covariance matrix estimator, 115  
eigenvalue product inequality, 1001  
eigenvalues, 983–984; generalized, 987–988

- elastic net, 955
- endogeneity, 341–342; generalized method of moments (GMM), 441–442; subset tests, 385–386, 441–442
- endogenous regressors, 344–345
- Engle-Granger ADF test, 595–596
- ensembling, 961–962
- equal weighting, 924–925
- equi-correlation model, 618
- ergodicity, time series, 461–463
- ergodic theorem, 463–464
- error-correction representation, 592
- error variance: confidence intervals for, 150; estimation of, 76–77, 110–111; estimators consistency, 173
- estimation, restricted: asymptotic distribution, 206–207; constrained least squares (CLS) estimator, 200–201; efficient minimum distance estimator, 208–209; exclusion restriction, 201–202, 209–211; finite sample properties, 202–205; Hausman equality, 211; inequality restrictions, 219–220; introduction, 199–200; Mankiw, Romer, and Weil example, 212–215; minimum distance, 205–206; misspecification, 216–218; nonlinear constraints, 218–219; variance estimation and standard errors, 208
- estimators, definition of, 3
- Eviews, 6
- existence of conditional expectation, 55–56
- exogenous variables, 345–346
- expectation, 14; errors, regression with, 377–379; of least squares estimator, 100–102; and variance of systems least squares, 318–319
- expectation equality, 1002, 1008
- expectation inequality, 1002, 1009
- experimental data, 3
- explosive process, 481
- exponential inequality, 1004, 1012–1013
- external instrumental variables, 553–554
- extrema of quadratic forms, 988–990
  
- factor-augmented regression, 336–337
- factor models, 331–333; with additional regressors, 335; approximate, 333–335
- familywise error probability, 249
- finite sample properties, CLS estimator, 202–205
- finite sample theory, 369
- first-differencing transformation, 624–625
- first-order autoregressive process, 477–480, 528–529
- fixed design residual bootstrap for time series, 509
- fixed design wild bootstrap for time series, 509–510
- fixed effects: asymptotic distribution of fixed effects 2SLS estimator, 647–648; covariance matrix estimation, 628–629; GMM interpretation of, 634–635; Hausman test for random vs., 641–642; regression, 633
- fixed effects estimation, 623–624, 633–634; asymptotic distribution of fixed effects estimator, 636–637; bias of, 654–655; in Stata, 629–630
- fixed effects model, 620–621; identification in, 636
- focused information criterion (FIC), 906–908
- forecast error decomposition, 542–543
- forecast intervals, asymptotic theory for least squares, 187
  
- Fortran, 6
- forward orthogonal transformation, 664–665
- forward stepwise regression, 909
- Frisch, Ragnar, 1, 82
- Frisch-Waugh-Lovell (FWL) theorem, 82; dummy variable estimator, 627
- F*-tests, 238–239
- full information maximum likelihood (FIML), 357
- functional central limit theorem (FCLT), 562
- functional convergence, nonstationary time series, 561–563
- fuzzy regression discontinuity (FRD), 763, 773–775; estimation of, 774–775
  
- GAUSS, 6
- Gauss, Carl Friedrich, 144–145
- Gaussian distribution, 139–141
- Gaussian elimination, 144
- Gaussian tail inequality, 1003, 1009
- Gauss-Markov theorem, 104–107, 144; differenced estimator, 626; panel data, 619
- Gauss-Newton algorithm, 144
- generalized eigenvalues, 987–988
- generalized least squares (GLS), 107–108; differenced estimator, 626; feasible, panel data, 632–633; forward orthogonal transformation, 665; panel data, 618–619
- generalized method of moments (GMM): Blundell-Bond, 661–664; bootstrap for, 443–444; clustered dependence, 432–433; conditional moment equation models, 444–445; constrained regression, 436; continuously updated, 439; covariance matrix estimation, 432; distance test, 438–439; distribution of GMM estimator, 428–429; efficient, 429–430; efficient GMM versus 2SLS, 430; endogeneity tests, 441–442; estimation of efficient weight matrix, 430–431; GMM estimator, 427–428; interpretation of fixed effects, 634–635; iterated, 431; linear, 648; linear moment models, 427; method of moments estimators, 425–426; moment equation models, 424; multivariate regression, 436–437; nonlinear, 442–443; nonlinear restricted, 435–436; overidentification tests, 439–441; overidentified moment equations, 426–427; restricted, 434–435; subset overidentification tests, 440–441; Wald test, 433–434
- general multinomial probit, 861–862
- generated regressors, 374–377
- geometric mean inequality, 999, 1005
- Golderberger, Arthur S., 124
- Granger, Clive W. J., 507
- Granger causality, 505–507
- Granger-Ramanathan averaging, 930–931
- Granger representation theorem, 589, 597
- Gretl, 6
- group James-Stein, 919–920
- growth rates, time series, 456–458
  
- Haavelmo, Trygve, 1–2
- Hall, Peter Gavin, 292
- hat matrix, 74
- Hausman equality, 211
- Hausman-Taylor model, 650–652
- Hausman tests, 239–240; for random vs. fixed effects, 641–642



- Heckman's model, 882–884  
Hermite polynomial, 727  
heteroskedasticity, 29–30; covariance matrix estimation under, 113–116, 174–176; joint asymptotic distribution, 171–173  
heteroskedasticity and autocorrelation consistent (HAC)  
  covariance matrix estimators, 494–496  
heteroskedasticity and autocorrelation robust (HAR) covariance matrix estimators, 494–496  
heteroskedasticity-robust covariance matrix estimation, 639–641  
heteroskedasticity-robust estimation-unbalanced case, 641  
high dimensionality, 941  
high-dimensional regression, 942  
Hölder's inequality, 1000, 1002, 1006, 1009  
hold-out criterion, 896–897  
homoskedasticity, 29–30; covariance matrix estimation under, 113, 174; joint asymptotic distribution, 171–173; linear regression model, 100; minimum distance tests under, 237–238; time series distribution under, 492–493  
homoskedastic martingale difference sequence, 466–467  
homoskedastic Wald statistic, 188, 235–236  
hypothesis testing, 150–152; acceptance and rejection regions, 226–227; alternative hypothesis, 225–226; asymptotic local power, 251–253; vector case, 254; bootstrap, 292–296; confidence intervals by test inversion, 247–248; criterion-based tests, 236, 295–296; endogeneity, 382–385; *F*-tests, 238–239; Hausman tests, 239–240; homoskedastic Wald tests, 235–236; instrumental variables, 368; KPSS stationarity test, 581–584; likelihood ratio test, 152–153; minimum distance tests, 236–237; minimum distance tests under homoskedasticity, 237–238; Monte Carlo simulation, 245–247; multiple tests and Bonferroni corrections, 248–249; nonlinear components, nonlinear least squares (NLLS), 802–804; of no serial correlation, 496; null hypothesis, 150–152, 225–226; power and test consistency, 250–251; problems with tests of nonlinear hypotheses, 242–245; *p*-values, 231–232; score tests, 240–242; statistical significance, 230–231; *t*-ratios and abuse of testing, 232–233; *t*-tests, 228–229; type I error, 227–228; type II error and power, 229; unit root, 578–581; Wald tests, 233–235
- idempotent matrices, 74, 985  
identically distributed data, 5  
identification: of conditional expectation, 56–57; difference in differences, 672–673; failure in, 396–397; in fixed effects model, 636; with instrumental variables, 646; *m*-estimators, 780; nonlinear least squares (NLLS), 792; nonparametric instrumental variables regression (NPIV), 748–749; parameter, instrumental variables, 349; of recursive VARs, 543–544; regression discontinuity, 764–765; of structural VARs, 549–550; time series model parameters, 488–490; Tobit regression, 877–879  
ill-posed inverse problem, 748  
impulse response function (IRF), 485–486; multivariate time series, 527–528, 535–537; orthogonalized impulse response function (OIRF), 540–542  
inclusion-exclusion principle, 1001, 1008  
independence of irrelevant alternatives, multiple choice, 854–855  
independent and identically distributed (i.i.d.) data, 5  
independently distributed data, 5  
index models, 831  
index series model, 831  
indicator variables, 33  
indirect least squares (ILS), 350  
inequalities: for matrices, 1001; probability, 1001–1005; for real numbers, 999–1000; for vectors, 1000  
inequality restrictions, restricted estimation, 219–220  
inference: after model selection, 911–913; difference in differences, 680–681; regression discontinuity, 767–768  
infinite-order moving average process, 476–477  
influential observations, 86–88  
information bound for normal regression, 153–154  
instrumental variables: Acemoglu, Johnson, and Robinson example, 412–414; Angrist and Krueger example, 414–416; asymptotic distribution of 2SLS, 361–362; bootstrap for 2SLS, 369–373; bootstrap overidentification tests, 392; clustered dependence, 373; college proximity example, 346–347; control function regression, 380–382; covariance matrix estimation, 364–366; demeaned representation, 352; determinants of 2SLS variance, 363–364; endogeneity, 341–342, 382–386; endogenous regressors, 344–345; estimator, 350–352; examples, 342–344; finite sample theory, 369; functions of parameters, 367–368; generated regressors, 374–377; hypothesis tests, 368; identification failure, 396–397; identification with, 646; instruments, 345–346; Lasso, 962–963; limited likelihood maximum likelihood (LIML), 357–359, 366–367; local average treatment effects, 392–395; many instruments, 400–404; nonparametric instrumental variables regression (NPIV), 746–750; overidentification tests, 386–392; overview, 341–342; panel data, 645–646; parameter identification, 349; probit model, 841–842; programming, 416–417; quantile regression, 824–825; reduced form, 347–348; regression with expectation errors, 377–379; split-sample IV and JIVE, 359–361; subset endogeneity tests, 385–386; subset overidentification tests, 389–392; two-stage least squares, 354–364, 369–373; Wald estimator, 353–354; weak instruments, 397–412  
instruments, 345–346  
integrated mean-squared error (IMSE), 701–702  
integrated squared error (ISE), 736  
interaction effect, 32  
intercept-only model, 25  
interdependence, 5  
inverse Mills ratio, 834, 875  
invertibility and identification, 38  
iterated generalized method of moments (GMM), 431
- jackknife covariance matrix estimation, 652–653  
jackknife covariance matrix model averaging, 929–930  
jackknife estimates, 262; for clustered observations, 267–268; example, 266–267; of variance, 263–266  
James-Stein shrinkage, 889, 915–916; empirical illustration, 920–923; group, 919–920; interpretation of the Stein effect, 916; positive part estimator, 916–917; shrinkage toward restrictions, 918–919  
Jensen's inequality, 999, 1002, 1005, 1008  
Johnson, Simon, 412–414  
joint asymptotic distribution, 171–173  
joint normality and linear regression, 142  
joint probability distribution, 5

*Journal of Political Economy*, 7  
Julia, 6

K-class estimator, 358  
kernel regression, 689–690; local nature of, 707  
K-fold cross-validation, 899–900, 951  
Khintchine's inequality, 1004, 1013–1015  
knots, 727  
Kolmogorov's inequality, 1005, 1019  
Kronecker products, 995  
Krueger, Alan B., 414–416  
Kwiatkowski, Phillips, Schmidt, and Shin stationarity test, 581–584

lag length, VAR, 533  
lag operator, time series, 473  
lag truncation number, 494–495  
Laguerre polynomial, 727  
large sample asymptotics: central limit theorem (CLT), 160–161; continuous mapping theorem (CMT) and delta method, 161–162; convergence of moments, 163–164; modes of convergence, 159; smooth function model, 162; stochastic order symbols, 162–163; weak law of large numbers (WLLN), 160  
Lasso estimation, 948–951; asymptotic theory for, 952–954; computation, 952; double selection, 963–965; elastic net, 955; instrumental variables, 962–963; penalty selection, 951; post-Lasso, 956; post-regularization, 965–967  
latent variable interpretation, 832–833  
Lavy, Victor, 751–753  
law of iterated expectations, 21–23, 104  
least absolute deviations, 779; quantile regression, 809–810  
least squares, asymptotic theory for: alternative covariance matrix estimators, 177–178; asymptotic leverage, 192; asymptotic normality, 167–171; asymptotic standard errors, 180–182; confidence intervals, 183–185; confidence regions, 188–189; consistency of error variance estimator, 173; consistency of least squares estimator, 165–167; Edgeworth expansion, 189–190; forecast intervals, 187; functions of parameters, 178–180; heteroskedastic covariance matrix estimation, 174–176; homoskedastic covariance matrix estimation, 174; homoskedastic Wald statistic, 188; joint distribution, 171–173; regression intervals, 185–187; summary of covariance matrix notation, 176; *t*-statistic, 182–183; uniformly consistent residuals, 191; Wald statistic, 187–188  
least squares estimator, 64–65; analysis of variance, 77–78; annihilator matrix, 76; bias of, 875–876; collinearity errors, 89–91; consistency of, 165–167; constrained, 200–201; CPS dataset, 88; demeaned regressors, 72; equivalence of SUR and, 324; estimation of error variance, 76–77; expectation of, 100–102; illustration of, 69–71; influential observations, 86–88; leave-one-out regression, 84–86; leverage values, 83–84; model in matrix notation, 73–74; multivariate regression, 316–318; numerical computation, 88–89; Phillips curve, 505; programming, 91–93; projection matrix, 74–75; projections, 78; regression components, 78–81; residual regression, 81–82; samples of, 62–63; solving for: with multiple regressors, 66–69; with one regressor, 65–66; two-stage, 354–357; variance of, 102–103

least squares regression: clustered sampling, 124–130; cluster levels, 131–132; computation, 120–121; covariance matrix estimation under heteroskedasticity, 113–116; covariance matrix estimation under homoskedasticity, 113; empirical example, 122–123; estimation of error variance, 110–111; estimation with sparse dummy variables, 118–119; expectation of least squares estimator, 100–102; Gauss-Markov theorem, 104–107; generalized least squares, 107–108; inference with clustered samples, 131; linear regression model, 100; mean-squared forecast error, 111–112; measures of fit, 121–122; multicollinearity, 123–124; random sampling, 98; residuals, 109–110; sample mean, 99; standard errors, 117–118; unconditional moments, 103–104; variance of least squares estimator, 102–103  
least squares residuals, 71–72  
leave-one-out cross validation, 122  
leave-one-out regression, 84–86  
Legendre, Adrien-Marie, 69  
Legendre polynomial, 726  
leverage point, 87  
leverage values, 83–84  
likelihood function, 143; Akaike information criterion (AIC) for, 894–895; probit and logit models, 833–835  
likelihood ratio test, 152–153; AID and BIC selection and, 901  
limited dependent variables, 829  
limited likelihood maximum likelihood (LIML), 357–359; asymptotic distribution, 366–367  
linear conditional expectation function (CEF), 31–32; with dummy variables, 33–35; with nonlinear effects, 32  
linear generalized method of moments, 648  
linear index function, 831  
linear models, time series, 474–475  
linear moment models, 427  
linear predictor error variance, 41  
linear probability model, 830–831  
linear projection, 37–38, 62–63; multivariate time series, 525; time series, 471  
linear projection coefficient, 36  
linear quantile functions, 813–814  
linear regression model, 100; joint normality and, 142  
linear series model, 831  
linear series regression model, 724  
local average treatment effects (LATE), 392–395  
local linear estimator, nonparametric regression, 690–692; estimation at a boundary, 698–700  
local mean, nonstationary time series, 567–569  
local polynomial estimator, nonparametric regression, 692  
local projection estimator, multivariate time series, 537  
local-to-zero parameter, 398  
Loève's  $c_r$  inequality, 999–1000, 1005  
log concave functions, 834  
logit model, 831; application, 839–840; asymptotic distribution, 836–837; binary panel data, 842–843; conditional, 851–854; covariance matrix estimation, 838; latent variable interpretation, 832–833; likelihood, 833–835; marginal effects, 838–839; mixed, 858–859; models, 829–830; models for response probability, 830–832; multinomial, 848–850; nested, 855–858; pseudo-true values, 835–836; semiparametric, 840–841

- log-likelihood function, 143, 152–153
- logs and percentages, 17–19
- long regression, 44–45
- long-run covariance matrix, 566
- long-run impulse response matrix, 528
- long-run multiplier, 502
- long-run restrictions, structural VARs, 550–551
- Lyapunov's inequality, 1003, 1009
  
- machine learning: approximate sparsity, 954–955; bagging, 958–960; big data, high dimensionality, and, 941–942; defined, 941; double/debiased, 967–968; elastic net, 955; ensemble, 961–962; high-dimensional regression, 942; Lasso estimation, 948–956; Lasso IV, 962–963; p-norms, 943; post-Lasso, 956; post-regularization Lasso, 965–967; random forests, 960–961; regression trees, 956–958; ridge regression, 944–951
- Mallows criterion, 895–896, 905–906; model averaging, 927–929
- Mammen distribution, 302
- Mankiw, Romer, Weil, 212–215
- Marcinkiewicz WLLN, 304
- Marcinkiewicz-Zygmund inequality, 1004, 1015–1016
- marginal effects, 829; binary choice, 838–839
- marginal significance level, 231
- Markov's inequality, 1003, 1009
- martingale difference sequence (MDS), 465–467; asymptotic distribution of least squares estimator, 491–492; functional CLT, 565–566; stochastic integral, 571–572; white noise, 471–472
- MATLAB, 6; clustered standard errors, 130; collinearity errors, 89–90; *F*-tests, 239; least squares regression, 93; nonlinear least squares (NLLS) computation, 804; normal distribution, 140–141; restricted estimation, 212, 215; setting the bootstrap seed in, 298; standard errors computation, 120–121; Wald statistic, 234
- matrices, 3, 977–978; addition, 979; algorithmic inversion, 993; calculus, 993–995; Cholesky decomposition, 990–991; complex, 978–979; decompositions, 986–987; determinant, 982–983; eigenvalues, 983–984; extrema of quadratic forms, 988–990; generalized eigenvalues, 987–988; idempotent, 74, 985; inequalities for, 1001; Kronecker products and vec operator, 995; multiplication, 979–980; norms, 996–998; notation, 73–74, 977–978; orthogonal and orthonormal, 982; positive definite, 984–985; QR decomposition, 991; rank and inverse, 981–982; singular values, 986; solving linear systems, 991–993; trace, 980; vector norms, 996
- matrix Cauchy-Schwarz inequality, 1002, 1009
- matrix convergence, series regression, 734–735
- maximal Khintchine inequality, 1004, 1015
- maximum likelihood estimator (MLE), 143–144, 779; factor models, 331–333; Heckman's model, 884; probit and logit models, 835
- mean, 14; nonstationary time series, 567–569; regression to the, 46–47
- mean squared error (MSE), 890; focused information criterion (FIC), 906–907; of model selection estimators, 909–911
- mean-squared forecast error (MSFE), 111–112
- measurement error bias, 342
- measurement error in the regressor, 342, 346
- measures of central tendency, 14–15
- measures of fit, least squares regression, 121–122
- median, 14
- median regression, 807–809
- medium regression, 44–45
- m-estimators: asymptotic distribution, 783–785; consistency, 780–782; covariance matrix estimation, 785–786; identification and estimation, 780; uniform law of large numbers (ULLN), 782–783
- method of moments estimators (MME), 425–426
- minimization, 39
- minimum distance, 205–206; tests for, 236–237; tests under homoskedasticity, 237–238
- minimum wage in New Jersey difference in differences, 669–672
- Minkowski's inequality, 1000, 1002, 1006–1007, 1009
- misspecification, restricted estimation, 216–218
- mixed logit, 858–859
- mixing properties of linear processes, 487–488
- model averaging, 889, 923–925; empirical illustration, 931; ensemble, 961–962; Granger-Ramanathan, 930–931; jack-knife [CV], 929–930; Mallows, 927–929; smoothed BIC and AIC, 925–927
- model selection, 889–891; Akaike information criterion (AIC), 498, 890, 892–895, 925–927; asymptotic selection optimality, 904–906; Bayesian information criterion, 890, 891–892, 925–927; best subset, 908–909; consistent, 902–904; cross-validation criterion, 897–898; empirical illustration, 913–914; focused information criterion (FIC), 906–908; hold-out criterion, 896–897; inference after, 911–913; K-fold cross-validation, 899–900; Mallows criterion, 895–896, 905–906, 927–929; model averaging and, 889, 923–931; MSE of, 909–911; relation with likelihood ratio testing, 901; similar criteria for, 900–901; stepwise regression, 908–909
- moment equation models, 424; overidentified, 426–427
- moment estimators, 64
- moments, convergence of, 163–164
- monotone probability inequality, 1001, 1008
- monotonicity, 821; norm, 1000, 1005
- Monte Carlo simulation, 245–247; mixed logit, 859
- moving average process, 475–476; infinite-order, 476–477
- multicollinearity, 123–124
- multinomial logit, 848–850
- multinomial response, 847–848
- multiple choice: BLP demand model, 865–867; conditional logit, 851–854; count data, 864–865; general multinomial probit, 861–862; independence of irrelevant alternatives, 854–855; mixed logit, 858–859; multinomial logit, 848–850; multinomial response, 847–848; nested logit, 855–858; ordered response, 862–864; simple multinomial probit, 860–861
- multiple comparisons problem, 248–249
- multiple regression, 71; nonparametric regression, 712–713
- multiple testing problem, 248–249
- multiplication, matrix, 979–980
- multivariate Lindeberg-Lévy Central Limit Theorem (CLT), 160–161
- multivariate matrix normal, 337–338
- multivariate normal distribution, 141–142

- multivariate regression: approximate factor models, 333–335; asymptotic distribution, 319–321; covariance matrix estimation, 321; equivalence of SUR and least squares, 324; expectation and variance of systems least squares, 318–319; factor-augmented regression, 336–337; factor models, 331–337; generalized method of moments (GMM), 436–437; least squares estimator, 316–318; maximum likelihood estimator, 324–325; multivariate matrix normal, 337–338; principal component analysis (PCA), 329–331; reduced rank regression, 325–328; regression systems, 315–316; restricted estimation, 325; seemingly unrelated regression, 322–323
- multivariate time series: asymptotic distribution, 531–532; Blanchard and Quah illustration, 551–553; covariance matrix estimation, 532–533; dynamic factor models, 554–555; estimation, 530–531; external instruments, 553–554; first-order vector autoregressive process, 528–529; forecast error decomposition, 542–543; illustrations, 533, 541–542, 551–553; impulse response function (IRF), 527–528, 535–537; linear projection, 525; local projection estimator, 537; long-run restrictions, 550–551; multiple equation time series models, 524–525; multivariate Wold decomposition, 525–527; oil price shocks, 544–545; orthogonalized impulse response function (OIRF), 540–542; orthogonalized shocks, 539–540; predictive regressions, 533–535;  $p$ th-order vector autoregressive process, 529; regression notation, 529–530; regression on residuals, 537–538; selection of lag length in VAR, 533; structural VARs, 546–550. *See also* time series
- multivariate Wold decomposition, 525–527
- Nadaraya-Watson (NW) estimator, 689
- near multicollinearity, 123
- nested logit, 855–858
- nonlinear constraints, restricted estimation, 218–219
- nonlinear effects, linear CEF with, 32
- nonlinear generalized method of moments (GMM), 442–443
- nonlinear hypotheses, 242–245
- nonlinear least squares (NLLS), 779; asymptotic distribution, 794–796; computation, 804; covariance matrix estimation, 796–797; estimation, 792–794; identification, 792; panel data, 797–798; testing for nonlinear components, 802–804; threshold models, 798–802
- nonlinear restricted generalized method of moments (GMM), 435–436
- nonparametric identification, 57; parametric vs., 750–751
- nonparametric instrumental variables regression (NPIV), 746–750; convergence rate, 749–750; identification, 748–749
- nonparametric quantile regression, 822–823
- nonparametric regression: application to test scores, 710–711; application to wage regression, 707–709; asymptotic bias, 692–694; asymptotic distribution, 702–704; asymptotic MSE (AMSE), 695–696; asymptotic variance, 694–695; binned means estimator, 687–688; clustered observations, 709–710; computation, 715; conditional variance estimation, 705–706; confidence bands, 707; cross-validation bandwidth selection, 701–702; curse of dimensionality, 713; estimation at a boundary, 698–700; kernel regression, 689–690, 707; local linear estimator, 690–692; local polynomial estimator, 692; multiple regressors, 712–713; nonparametric residuals and prediction errors, 700–701; partially linear regression, 714–715; reference bandwidth, 697–698; undersmoothing, 704–705; variance estimation and standard errors, 706
- nonparametric selection, 884–885
- nonrandom matrices, 3
- nonstationary time series: AR(1) estimation with an intercept, 574–576; AR[ $p$ ] models with unit root, 576–578; Beveridge-Nelson decomposition, 563–565; cointegrating regression, 594–597; cointegration, 589–593; demeaning and detrending, 569–570; Dickey-Fuller coefficient distribution, 573–576; estimation of AR(1) parameter, 572–574; functional CLT, 565–566; KPSS stationarity test, 581–584; means, local means, and trends, 567–569; nonstationary VARs, 588–594; orders of integration, 566–567; partial sum process and functional convergence, 561–563; role of intercept and trend, 593–594; sample covariances of integrated and stationary processes, 576; spurious regression, 584–588; stochastic integrals, 570–572; testing for a unit root, 578–581; testing for cointegration in VECM, 599–602; VECM estimation, 597–599. *See also* time series
- normal-approximation bootstrap confidence interval, 271
- normal distribution, 139–141; multivariate, 141–142
- normal regression: confidence intervals for regression coefficients, 148–149; distribution of OLS coefficient vector, 145; distribution of OLS residual vector, 146; distribution of variance estimator, 146–147; information bound for, 153–154; joint normality and linear regression, 142; likelihood ratio test, 152–153; multivariate normal distribution, 141–142; normal distribution and, 139–141; normal regression model, 143–144;  $t$ -statistic, 147–148;  $t$ -test, 150–152
- norm equivalence, 1001, 1008
- norm monotonicity, 1000, 1005
- no serial correlation, testing hypothesis of, 496
- notation: matrix algebra, 977–978; panel data, 615
- null hypothesis, 150–152, 225–226
- numerical calculation, least squares, 88–89
- observational data, 3–4
- observations, 3; bootstrap, 274–275, 303–304; clustered, 267–268, 303–304, 709–710, 743; influential, 86–88
- oil price shocks, 544–545
- omitted serial correlation, testing for, 496–498
- omitted variable bias, 44–45
- one standard error (1se) rule, 899–900
- one-way error component model, 617
- ordered response, 862–864
- orders of integration, nonstationary time series, 566–567
- ordinary least squares (OLS) coefficient vector, distribution of, 145
- ordinary least squares (OLS) estimator, 65–66, 69, 779; Frisch-Waugh-Lovell (FWL), 82
- ordinary least squares (OLS) residual vector, distribution of, 146
- orthogonal decomposition, 77–78
- orthogonalized impulse response function (OIRF), 540–542
- orthogonal matrices, 982
- orthogonal polynomials, 726–727
- orthogonalized shocks, 539–540
- orthonormal matrices, 982

- outliers, 86
- overidentification tests, 386–392; bootstrap, 392; generalized method of moments (GMM), 439–441; subset, 389–392, 440–441
- overidentified moment equations, 426–427
- OxMetrics, 6
  
- pairs bootstrap, 299; for time series, 509
- panel data, 4; Anderson-Hsiao estimator, 656; Arellano-Bond estimator, 657–659; asymptotic distribution for unbalanced panels, 637–639; asymptotic distribution of fixed effects 2SLS estimator, 647–648; asymptotic distribution of fixed effects estimator, 636–637; bias of fixed effects estimation, 654–655; binary, 842–843; Blundell-Bond estimator, 661–664; bootstrap, 653; censored regression, 885–886; differenced estimator, 624–626; dummy variables regression, 626–627; dynamic panel models, 653–654; dynamic panels with predetermined regressors, 660–661; empirical illustration, 665–666; estimation of fixed effects, 633–634; estimation of time-invariant regressors, 648–650; estimation with time-invariant regressors, 648–650; between estimator, 630–631; feasible GLS, 632–633; fixed effects covariance matrix estimation, 628–629; fixed effects estimation in Stata, 629–630; fixed effects estimator, 623–624; fixed effects model, 620–621, 636; forward orthogonal transformation, 664–665; GMM interpretation of fixed effects, 634–635; Hausman-Taylor model, 650–652; Hausman test for random vs. fixed effects, 641–642; heteroskedasticity-robust covariance matrix estimation, 639–641; heteroskedasticity-robust estimation-unbalanced case, 641; identification with instrumental variables, 646; instrumental variables, 645–646; intercept in fixed effects regression, 633; jackknife covariance matrix estimation, 652–653; linear GMM, 648; nonlinear least squares (NLS), 797–798; notation, 615; one-way error component model, 617; pooled regression, 615–617; quantile regression, 823–824; random effects, 617–619; time indexing and unbalanced panels, 614; time trends, 642–643; within transformation, 621–623; two-way error components, 643–645; weak instruments, 659
- panel nonparametric bootstrap, 653
- parallel quantile functions, 814
- parameters functions, asymptotic theory for least squares, 178–180
- parametric bootstrap, 296–297
- parametric identification, 57; nonparametric vs., 750–751
- partially linear regression, 714–715; series regression, 745
- partial sum process, 561–563
- partitioned matrix, 978
- partitioned matrix inversion, 80–81
- PcGive, 6
- percentile interval, bootstrap, 272–273, 283–284; bias-corrected, 285–286; percentile- $t$ , 288–290
- percentiles, 15
- percentile- $t$  asymptotic refinement, 290–292
- percentile- $t$  interval, 288–290
- percentages and logs, 17–19
- Phillips curve, 505
- plug-in estimators, 64; smooth function model, 162
- p-norms, 943
  
- pointwise confidence intervals, 707
- Poisson regression, 864–865
- police presence and crime difference in differences, 675–677
- polynomial regression, 724–725
- pooled regression, 615–617
- population, 5
- positive definite matrices, 984–985
- post-Lasso estimation, 956
- post model-selection (PMS) estimator, 910
- post-regularization Lasso, 965–967
- potential outcomes framework, 50
- power, asymptotic local, 251–253; vector case, 254
- power and test consistency, hypothesis testing, 250–251
- power function, hypothesis testing, 229
- prediction errors, nonparametric regression, 700–701
- prediction standard error, 85
- predictive regressions, multivariate time series, 533–535
- principal component analysis (PCA), 329–331
- probability approach to econometrics, 1–2
- probability convergence, 159
- probability density function, 13–14
- probability distribution, 13
- probability inequalities, 1001–1005
- probit model, 831; application, 839–840; asymptotic distribution, 836–837; binary panel data, 842–843; covariance matrix estimation, 838; general multinomial, 861–862; instrumental variables, 841–842; latent variable interpretation, 832–833; likelihood, 833–835; marginal effects, 838–839; models, 829–830; models for response probability, 830–832; pseudo-true values, 835–836; semiparametric, 840–841; simple multinomial, 860–861
- projection matrix, least squares estimator, 74–75
- proxy SVARs, 553–554
- pseudo-true values, binary choice, 835–836
- pth-order autoregressive process, 484–485, 529
- p-values, 231–232
- Python, 6
  
- QR decomposition, 991
- quadratic forms, extrema of, 988–990
- quadratic inequality, 1001, 1008
- Quadratic Inequality, 84
- quantile causal effects, 820–821
- quantile crossings, 819–820
- quantile regression, 779, 810–813; asymptotic distribution, 815–816; clustered dependence, 818–819; covariance matrix estimation, 817–818; estimation, 814–815; example quantile shapes, 813–814; instrumental variables, 824–825; least absolute deviations, 809–810; median regression, 807–809; nonparametric, 822–823; panel data, 823–824; quantile causal effects, 820–821; quantile crossings, 819–820; random coefficient representation, 821–822
- quantiles, 15; normal distribution, 140–141
- quantile treatment effects, 820
- quasi-structural approach, 2
  
- R (programming language), 6; clustered standard errors, 129–130; least squares regression, 92–93; nonlinear least

- R (programming language) (cont.)  
squares (NLLS) computation, 804; normal distribution, 140–141; restricted estimation, 212, 214; setting the bootstrap seed in, 298; standard errors computation, 120
- Rademacher random variables, 302, 1003
- random coefficient model, 48–49; quantile regression, 821–822
- random effects, panel data, 617–619; binary, 842–843; between estimator, 630–631; Hausman test for fixed vs., 641–642
- random forests, 960–961
- random matrices, 3
- random sampling, 5–6; least squares regression, 98
- random walk, 480
- range space, 74
- rank and inverse, matrix, 981–982
- RATS, 6
- R-bar-squared, 121
- real numbers, 3; inequalities for, 999–1000
- recursive bootstrap for time series, 508–509
- recursive VARs, 543–544
- reduced form, instrumental variables, 347–348
- reduced rank regression, 325–328
- reference bandwidth, nonparametric regression, 697–698
- regression, 13; Akaike information criterion (AIC) for, 892–894; bivariate, 71; bootstrap, 298–301; Box-Cox, 790, 793; censored, 872–875; cointegrating, 594–597; control function, 380–382; discontinuity (*See* regression discontinuity); dummy variables, 626–627; with expectation errors, 377–379; factor-augmented, 336–337; fixed effects, 633; high-dimensional, 942; intervals, asymptotic theory for least squares, 185–187; kernel, 689–690, 707; kink model, 791–792; least squares (*See* least squares regression); least squares estimator, 78–81; leave-one-out, 84–86; linear, 100, 142; long, 44–45; to the mean, 46–47; median, 807–809; medium, 44–45; multiple, 71; multivariate (*See* multivariate regression); nonparametric (*See* nonparametric regression); nonparametric instrumental variables regression (NPIV), 746–750; normal (*See* normal regression); notation, multivariate time series, 529–530; partially linear, 714–715; Poisson, 864–865; polynomial, 724–725; pooled, 615–617; predictive, multivariate time series, 533–535; quantile (*See* quantile regression); reduced rank, 325–328; regressor bounds, 733–734; residual, 81–82; on residuals, multivariate time series, 537–538; reverse, 47–48; ridge, 944–951; series (*See* series regression); short, 44–45; simple, 71; spline, 728–729; spurious, 584–588; stepwise, 908–909; testing for serial correlation in, 507–508; time series, 500–501; Tobit, 873, 876–879
- regression coefficients, 41–42
- regression derivative, 30–31
- regression discontinuity design (RDD): bandwidth selection, 768–770; with covariates, 770–771; density discontinuity test, 772–773; estimation, 766–767; fuzzy, 763, 773–775; identification, 764–765; inference, 767–768; sharp, 763–764; simple estimator, 771–772
- regression fallacy, 46–47
- regression subvectors, 42–43
- regression trees, 956–958; random forests, 960–961
- regression variance, 25–26
- regressors, 13
- replication, 6–7; bootstrap, 297
- resampling methods: bootstrap, 262, 268–304; jackknife, 263–268
- residual bootstrap, 302
- residuals: least squares, 71–72; regression, 81–82, 109–110; uniformly consistent, 191; nonparametric, 700–701; regression on, 537–538; series regression, 740
- response probability, 829; models for, 830–832; multinomial response, 847–848
- restricted estimation: asymptotic distribution, 206–207; constrained least squares (CLS) estimator, 200–201; efficient minimum distance estimator, 208–209; exclusion restriction, 201–202, 209–211; finite sample properties, 202–205; Hausman equality, 211; inequality restrictions, 219–220; Mankiw, Romer, and Weil example, 212–215; minimum distance, 205–206; misspecification, 216–218; multivariate regression, 325; nonlinear constraints, 218–219; variance estimation and standard errors, 208
- restricted generalized method of moments (GMM), 434–435
- restricted wild bootstrap, 302
- restricted wild cluster bootstrap, 303
- reverse regression, 47–48
- ridge regression, 944–946; illustrating, 948; Lasso, 948–951; statistical properties, 947–948
- Robinson, James A., 412–414
- Rosenthal's inequality, 1005, 1017–1019
- R-squared, 78
- Rubin causal model, 50
- Rule-of-Thumb (ROT) bandwidth: nonparametric regression, 697–698; regression discontinuity, 768–770
- running variable, 763
- sample mean, 99; distribution of bootstrap, 275–276; squared prediction error, 85
- sample selection bias, 881–882
- sampling, 2; clustered, 4–5, 124–130; random, 5–6. *See also* resampling methods
- Sargan, Denis, 389
- Sargan test, 387–389, 390
- SAS, 6
- scalars, 3, 977
- Schwarz criterion, 891–892
- Schwarz inequality, 1000, 1006
- Schwarz matrix inequality, 1001, 1007
- score tests, 240–242
- second-level Brownian Bridge, 583
- second-order autoregressive process, 481–484
- seemingly unrelated regression (SUR), 322–323; equivalence of least squares and, 324
- selection: model, 889–901; nonparametric, 884–885
- selection bias, 881–882
- semiparametric approach, 2; binary choice, 840–841
- serial correlation, testing for, 507–508
- series regression: additively separable models, 746; Angrist and Lavy example, 751–753; asymptotic normality, 738–739; clustered observations, 743; confidence bands, 743–744; consistent estimation, 736; convergence rate, 736–737; cross-validation model selection, 740–741; global/local nature of, 729–731; matrix convergence, 734–735; multiple regressors, 746; nonparametric instrumental variables regression (NPIV), 746–750;

- nonparametric vs. parametric identification, 750–751; orthogonal polynomials, 726–727; panel fixed effects, 745–746; partially linear model, 745; polynomial regression, 724–725; regression estimation, 739; regressor bounds, 733–734; residuals and regression fit, 740; spline regression, 728–729; splines, 727–728; Stone-Weierstrass and Jackson approximation theory, 731–733; undersmoothing, 739–740; uniform approximations, 744–745; variance and standard error estimation, 742
- sharp regression discontinuity, 763–764
- shocks: aggregate supply, 544; oil price, 544–545; orthogonalized, 539–540
- short regression, 44–45
- shrinkage methods, 914–915; James-Stein shrinkage, 915–923; toward restrictions, 918–919
- significance level, 151, 228
- simple conditional multinomial probit, 860–861
- simple law, 21–23
- simple multinomial probit, 860–861
- simple regression, 71
- simulated maximum likelihood, 861
- simultaneous equations bias, 344
- single index model, 831
- singular values, matrix, 986
- smoothed BIC and AIC, 925–927
- smooth function model, 162
- smoothing bias, 693, 699
- software, econometric, 6; least squares programming, 91–93
- sparse dummy variables, 118–119
- sparsity, approximate, 954–955
- spatial dependence, 5
- spectral norm, series regression, 735
- splines, 727–728; regression model, 728–729
- split-sample IV and JIVE, 359–361
- spurious regression, 584–588
- stacking, 929–930, 961
- standard data structures, 4–6
- standard errors: asymptotic, 180–182; bootstrap, 270–272, 282, 372–373; constrained least squares (CLS), 208; least squares regression, 117–118; nonparametric regression, 706; series regression, 742
- standard normal distribution, 139–141; multivariate, 141–142
- Stata, 6; asymptotic standard errors computation, 182; bootstrap standard error, 271–272, 301; clustered standard errors, 129; fixed effects estimation in, 629–630; instrumental variables, 416–417; jackknife standard errors, 267, 268; least squares regression, 92; multivariate probit, 862; nonlinear least squares (NLLS) computation, 804; nonparametric regression, 715; normal distribution, 140–141; quantile regression estimation, 815; random effects probit, 843; random effects regression, 619; restricted estimation, 212–213; setting the bootstrap seed in, 298; standard errors computation, 120; within transformation, 623; Wald statistic, 234–235
- static models, 501–502
- stationarity, time series, 458–460; KPSS test, 581–584
- statistical significance, hypothesis testing, 230–231
- stepwise regression, 908–909
- stochastic integrals, nonstationary time series, 570–572
- stochastic order symbols, 162–163
- Stone-Weierstrass theorem, 731–733
- strict mean independence, 615–616
- strict multicollinearity, 123
- strong Schwarz matrix inequality, 1001, 1008
- structural approach, 2
- structural effects, 51
- structural vector autoregressive models (SVARs), 546–548; external instruments, 553–554; identification of, 549–550; long-run restrictions, 550–551
- studentized statistic, 182–183
- subsampling, 262
- subset endogeneity tests, 385–386; generalized method of moments (GMM), 441–442
- subset overidentification tests, 389–392; generalized method of moments (GMM), 440–441
- sum of squared errors (SSE), 65, 66–68, 78–79
- switchers, 843
- symmetrization inequality, 1004
  
- terminology, econometrics, 2–3
- test inversion, confidence intervals by, 247–248
- test scores, nonparametric regression application to, 710–711
- test statistic, 226–227
- Theil, Henri, 122
- threshold models, nonlinear least squares (NLLS), 798–802
- tied-down Brownian motion, 570
- time indexing, 614
- time series, 4; Akaike information criterion (AIC), 498; ARMA and ARIMA processes, 486–487; AR[P] process, 484–485; asymptotic distribution of least squares estimator, 491–492; asymptotic distribution under general dependence, 493; autoregressive Wold representation, 473–474; bootstrap for, 508–510; CLT for correlated observations, 469–471; CLT for martingale differences, 467–468; conditioning on information sets, 464–465; convergent series, 460–461; covariance matrix estimation, 493–494; covariance matrix estimation under general dependence, 494–496; differences and growth rates, 456–458; distribution under homoskedasticity, 492–493; ergodicity, 461–463; ergodic theorem, 463–464; estimation of autoregressive models, 490–491; examples, 455–456; first-order autoregressive process, 477–480; Granger causality, 505–507; identification of parameters, 488–490; illustrations, 498–500, 505; impulse response function, 485–486; infinite-order moving average process, 476–477; lag operator, 473; linear models, 474–475; linear projection, 471; martingale difference sequences (MDS), 465–467; mixing, 468–469; mixing properties of linear processes, 487–488; model selection, 498; moving average process, 475–476; multivariate (*See* multivariate time series); nonstationary (*See* nonstationary time series); regression models, 500–501; second-order autoregressive process, 481–484; static, distributed lag, and autoregressive distributed lag models, 501–502; stationarity, 458–460; testing for omitted serial correlation, 496–498; testing for serial correlation in regression models, 507–508; testing the hypothesis of no serial correlation, 496; time trends, 502–505; transformations of stationary processes, 460; unit root and explosive AR[1] processes, 480–481; white noise, 471–472; Wold decomposition, 472–473

- time trends, 502–505, 642–643  
Tobin, James, 877  
Tobit estimator, 876–877  
Tobit regression, 873, 876–879; identification in, 877–879; panel data, 885–886  
trace, matrix, 980  
trace inequality, 1001, 1007–1008  
transformations of stationary processes, 460  
transpose of a matrix, 978  
trees, regression, 956–958; random forests, 960–961  
trend specification, difference in differences, 677–678  
triangle inequality, 999, 1000, 1001, 1005, 1007  
trimmed estimator of bootstrap variance, 280–282  
*Triumph of Mediocrity in Business, The*, 47  
trivial power, 229  
 $t$ -statistic/ $t$ -ratio, 147–148, 182–183; Edgeworth expansion, 189–190; hypothesis testing, 232–233  
 $t$ -test, 150–152, 228–229  
two-stage least squares (2SLS), 354–357; asymptotic distribution of, 362–363; asymptotic distribution of fixed effects, 647–648; bootstrap for, 369–373; consistency of, 361–362; determinants of variance, 363–364; efficient GMM versus, 430; standard errors, 372–373  
two-step GMM estimator, 430–431  
two-way error components, panel data, 643–645  
two-way within estimator, 644  
type I error, 227–228  
type II error, 229
- unbalanced panels, 614  
unbalanced regression design, 83  
unbiased estimator, 99, 105  
unconditional moments, 103–104  
unconditional variance, 27  
undersmoothing: nonparametric regression, 704–705; series regression, 739–740  
uniform approximations, series regression, 744–745  
uniform law of large numbers (ULLN), 782–783  
unit root process, 480–481; AR[p] models with, 576–578; testing for, 578–581  
univariate normal distribution, 140  
untrimmed bootstrap standard errors, 282
- variables, 3; best linear predictor, 62–63; continuous, 20–21; estimation with sparse dummy, 118–119; instrumental (See instrumental variables)
- variance estimators: bootstrap (See bootstrap estimator); consistency of, 173, 279–280; constrained least squares (CLS), 208; distribution of, 146–147; jackknife, 263–266; nonparametric regression, 706; series regression, 742  
variance of least squares estimator, 102–103  
vector autoregressive (VAR) model, 524–525; forecast error decomposition, 542–543; identification of recursive, 543–544; long-run restrictions, 550–551; nonstationary, 588–589; selection of lag length in, 533; structural, 546–550  
vector error correction model (VECM), 592–593; estimation, 597–599; role of intercept and trend in, 593–594; testing for cointegration in, 599–602  
vector norms, 996  
vectors, 3, 977; inequalities for, 1000
- wage regression, nonparametric regression application to, 707–709  
Wald, Abraham, 235  
Wald estimator, 353–354  
Wald test, 187–188; F statistic, 239; generalized method of moments (GMM), 433–434; Hausman statistic, 240; homoskedastic Wald tests, 188, 235–236; hypothesis testing, 233–235; score tests, 241–242  
Wald-type bootstrap tests, 294–295  
weak instruments, 397–400; with  $K_2$  greater than 1, 410–412; panel data, 659; testing for, 404–410  
weak law of large numbers (WLLN), 160; bootstrap, 277, 279, 300; consistency of least squares estimator, 165–167; Marcinkiewicz, 304  
weighted mean-squared error (WMSE), 889–890; shrinkage methods, 914–915  
White, Halbert L., 116–117  
White covariance matrix estimator, 115  
white noise, 471–472  
Whittle's inequalities, 1004, 1016–1017  
Wiener process, 563  
wild bootstrap, 301–302; for time series, 509–510  
within-cluster correlation, 5  
within transformation, panel data, 621–623  
Wold decomposition, 472–473; multivariate, 525–527
- Zellner, Arnold, 323  
 $z$ -statistic, 182–183