

Contents

Preface	xvii
I Nonparametric Kernel Methods	1
1 Density Estimation	3
1.1 Univariate Density Estimation	4
1.2 Univariate Bandwidth Selection: Rule-of-Thumb and Plug-In Methods	14
1.3 Univariate Bandwidth Selection: Cross-Validation Methods	15
1.3.1 Least Squares Cross-Validation	15
1.3.2 Likelihood Cross-Validation	18
1.3.3 An Illustration of Data-Driven Bandwidth Selection	19
1.4 Univariate CDF Estimation	19
1.5 Univariate CDF Bandwidth Selection: Cross-Validation Methods	23
1.6 Multivariate Density Estimation	24
1.7 Multivariate Bandwidth Selection: Rule-of-Thumb and Plug-In Methods	26
1.8 Multivariate Bandwidth Selection: Cross-Validation Methods	27
1.8.1 Least Squares Cross-Validation	27
1.8.2 Likelihood Cross-Validation	28
1.9 Asymptotic Normality of Density Estimators	28
1.10 Uniform Rates of Convergence	30
1.11 Higher Order Kernel Functions	33
1.12 Proof of Theorem 1.4 (Uniform Almost Sure Convergence)	35
1.13 Applications	40

1.13.1	Female Wage Inequality	41
1.13.2	Unemployment Rates and City Size	43
1.13.3	Adolescent Growth	44
1.13.4	Old Faithful Geyser Data	44
1.13.5	Evolution of Real Income Distribution in Italy, 1951–1998	45
1.14	Exercises	47
2	Regression	57
2.1	Local Constant Kernel Estimation	60
2.1.1	Intuition Underlying the Local Constant Kernel Estimator	64
2.2	Local Constant Bandwidth Selection	66
2.2.1	Rule-of-Thumb and Plug-In Methods	66
2.2.2	Least Squares Cross-Validation	69
2.2.3	AIC_c	72
2.2.4	The Presence of Irrelevant Regressors	73
2.2.5	Some Further Results on Cross-Validation	78
2.3	Uniform Rates of Convergence	78
2.4	Local Linear Kernel Estimation	79
2.4.1	Local Linear Bandwidth Selection: Least Squares Cross-Validation	83
2.5	Local Polynomial Regression (General p th Order)	85
2.5.1	The Univariate Case	85
2.5.2	The Multivariate Case	88
2.5.3	Asymptotic Normality of Local Polynomial Estimators	89
2.6	Applications	92
2.6.1	Prestige Data	92
2.6.2	Adolescent Growth	92
2.6.3	Inflation Forecasting and Money Growth	93
2.7	Proofs	97
2.7.1	Derivation of (2.24)	98
2.7.2	Proof of Theorem 2.7	100
2.7.3	Definitions of $A_{l,p+1}$ and V_l Used in Theorem 2.10	106
2.8	Exercises	108
3	Frequency Estimation with Mixed Data	115
3.1	Probability Function Estimation with Discrete Data	116

3.2	Regression with Discrete Regressors	118
3.3	Estimation with Mixed Data: The Frequency Approach	118
3.3.1	Density Estimation with Mixed Data	118
3.3.2	Regression with Mixed Data	119
3.4	Some Cautionary Remarks on Frequency Methods	120
3.5	Proofs	122
3.5.1	Proof of Theorem 3.1	122
3.6	Exercises	123
4	Kernel Estimation with Mixed Data	125
4.1	Smooth Estimation of Joint Distributions with Discrete Data	126
4.2	Smooth Regression with Discrete Data	131
4.3	Kernel Regression with Discrete Regressors: The Irrelevant Regressor Case	134
4.4	Regression with Mixed Data: Relevant Regressors	136
4.4.1	Smooth Estimation with Mixed Data	136
4.4.2	The Cross-Validation Method	138
4.5	Regression with Mixed Data: Irrelevant Regressors	140
4.5.1	Ordered Discrete Variables	144
4.6	Applications	145
4.6.1	Food-Away-from-Home Expenditure	145
4.6.2	Modeling Strike Volume	147
4.7	Exercises	150
5	Conditional Density Estimation	155
5.1	Conditional Density Estimation: Relevant Variables	155
5.2	Conditional Density Bandwidth Selection	157
5.2.1	Least Squares Cross-Validation: Relevant Variables	157
5.2.2	Maximum Likelihood Cross-Validation: Relevant Variables	160
5.3	Conditional Density Estimation: Irrelevant Variables	162
5.4	The Multivariate Dependent Variables Case	164
5.4.1	The General Categorical Data Case	167
5.4.2	Proof of Theorem 5.5	168
5.5	Applications	171
5.5.1	A Nonparametric Analysis of Corruption	171
5.5.2	Extramarital Affairs Data	172
5.5.3	Married Female Labor Force Participation	175

5.5.4	Labor Productivity	177
5.5.5	Multivariate Y Conditional Density Example: GDP Growth and Population Growth Conditional on OECD Status	178
5.6	Exercises	180
6	Conditional CDF and Quantile Estimation	181
6.1	Estimating a Conditional CDF with Continuous Covariates without Smoothing the Dependent Variable	182
6.2	Estimating a Conditional CDF with Continuous Covariates Smoothing the Dependent Variable	184
6.3	Nonparametric Estimation of Conditional Quantile Functions	189
6.4	The Check Function Approach	191
6.5	Conditional CDF and Quantile Estimation with Mixed Discrete and Continuous Covariates	193
6.6	A Small Monte Carlo Simulation Study	196
6.7	Nonparametric Estimation of Hazard Functions	198
6.8	Applications	200
6.8.1	Boston Housing Data	200
6.8.2	Adolescent Growth Charts	202
6.8.3	Conditional Value at Risk	202
6.8.4	Real Income in Italy, 1951–1998	206
6.8.5	Multivariate Y Conditional CDF Example: GDP Growth and Population Growth Conditional on OECD Status	206
6.9	Proofs	209
6.9.1	Proofs of Theorems 6.1, 6.2, and 6.4	209
6.9.2	Proofs of Theorems 6.5 and 6.6 (Mixed Covariates Case)	214
6.10	Exercises	215
II	Semiparametric Methods	219
7	Semiparametric Partially Linear Models	221
7.1	Partially Linear Models	222
7.1.1	Identification of β	222
7.2	Robinson's Estimator	222
7.2.1	Estimation of the Nonparametric Component	228

7.3	Andrews's MINPIN Method	230
7.4	Semiparametric Efficiency Bounds	233
7.4.1	The Conditionally Homoskedastic Error Case	233
7.4.2	The Conditionally Heteroskedastic Error Case	235
7.5	Proofs	238
7.5.1	Proof of Theorem 7.2	238
7.5.2	Verifying Theorem 7.3 for a Partially Linear Model	244
7.6	Exercises	246
8	Semiparametric Single Index Models	249
8.1	Identification Conditions	251
8.2	Estimation	253
8.2.1	Ichimura's Method	253
8.3	Direct Semiparametric Estimators for β	258
8.3.1	Average Derivative Estimators	258
8.3.2	Estimation of $g(\cdot)$	262
8.4	Bandwidth Selection	263
8.4.1	Bandwidth Selection for Ichimura's Method	263
8.4.2	Bandwidth Selection with Direct Estimation Methods	265
8.5	Klein and Spady's Estimator	266
8.6	Lewbel's Estimator	267
8.7	Manski's Maximum Score Estimator	269
8.8	Horowitz's Smoothed Maximum Score Estimator	270
8.9	Han's Maximum Rank Estimator	270
8.10	Multinomial Discrete Choice Models	271
8.11	Ai's Semiparametric Maximum Likelihood Approach	272
8.12	A Sketch of the Proof of Theorem 8.1	275
8.13	Applications	277
8.13.1	Modeling Response to Direct Marketing Catalog Mailings	277
8.14	Exercises	281
9	Additive and Smooth (Varying) Coefficient Semiparametric Models	283
9.1	An Additive Model	283
9.1.1	The Marginal Integration Method	284
9.1.2	A Computationally Efficient Oracle Estimator	286
9.1.3	The Ordinary Backfitting Method	289

9.1.4	The Smoothed Backfitting Method	290
9.1.5	Additive Models with Link Functions	295
9.2	An Additive Partially Linear Model	297
9.2.1	A Simple Two-Step Method	299
9.3	A Semiparametric Varying (Smooth) Coefficient Model	301
9.3.1	A Local Constant Estimator of the Smooth Coefficient Function	302
9.3.2	A Local Linear Estimator of the Smooth Coefficient Function	303
9.3.3	Testing for a Parametric Smooth Coefficient Model	306
9.3.4	Partially Linear Smooth Coefficient Models	308
9.3.5	Proof of Theorem 9.3	310
9.4	Exercises	312
10	Selectivity Models	315
10.1	Semiparametric Type-2 Tobit Models	316
10.2	Estimation of a Semiparametric Type-2 Tobit Model	317
10.2.1	Gallant and Nychka's Estimator	318
10.2.2	Estimation of the Intercept in Selection Models	319
10.3	Semiparametric Type-3 Tobit Models	320
10.3.1	Econometric Preliminaries	320
10.3.2	Alternative Estimation Methods	323
10.4	Das, Newey and Vella's Nonparametric Selection Model	328
10.5	Exercises	330
11	Censored Models	331
11.1	Parametric Censored Models	332
11.2	Semiparametric Censored Regression Models	334
11.3	Semiparametric Censored Regression Models with Nonparametric Heteroskedasticity	336
11.4	The Univariate Kaplan-Meier CDF Estimator	338
11.5	The Multivariate Kaplan-Meier CDF Estimator	341
11.5.1	Nonparametric Regression Models with Random Censoring	343
11.6	Nonparametric Censored Regression	345
11.6.1	Lewbel and Linton's Approach	345
11.6.2	Chen, Dahl and Khan's Approach	346

11.7 Exercises	348
III Consistent Model Specification Tests	349
12 Model Specification Tests	351
12.1 A Simple Consistent Test for Parametric Regression Functional Form	354
12.1.1 A Consistent Test for Correct Parametric Functional Form	355
12.1.2 Mixed Data	360
12.2 Testing for Equality of PDFs	362
12.3 More Tests Related to Regression Functions	365
12.3.1 Härdle and Mammen's Test for a Parametric Regression Model	365
12.3.2 An Adaptive and Rate Optimal Test	367
12.3.3 A Test for a Parametric Single Index Model	369
12.3.4 A Nonparametric Omitted Variables Test	370
12.3.5 Testing the Significance of Categorical Variables	375
12.4 Tests Related to PDFs	378
12.4.1 Testing Independence between Two Random Variables	378
12.4.2 A Test for a Parametric PDF	380
12.4.3 A Kernel Test for Conditional Parametric Distributions	382
12.5 Applications	385
12.5.1 Growth Convergence Clubs	385
12.6 Proofs	388
12.6.1 Proof of Theorem 12.1	388
12.6.2 Proof of Theorem 12.2	389
12.6.3 Proof of Theorem 12.5	389
12.6.4 Proof of Theorem 12.9	391
12.7 Exercises	394
13 Nonsmoothing Tests	397
13.1 Testing for Parametric Regression Functional Form	398
13.2 Testing for Equality of PDFs	401
13.3 A Nonparametric Significance Test	401
13.4 Andrews's Test for Conditional CDFs	402
13.5 Hong's Tests for Serial Dependence	404

13.6	More on Nonsmoothing Tests	408
13.7	Proofs	409
13.7.1	Proof of Theorem 13.1	409
13.8	Exercises	410
IV Nonparametric Nearest Neighbor and Series Methods		413
14	<i>K</i>-Nearest Neighbor Methods	415
14.1	Density Estimation: The Univariate Case	415
14.2	Regression Function Estimation	419
14.3	A Local Linear k -nn Estimator	421
14.4	Cross-Validation with Local Constant k -nn Estimation	422
14.5	Cross-Validation with Local Linear k -nn Estimation	425
14.6	Estimation of Semiparametric Models with k -nn Methods	427
14.7	Model Specification Tests with k -nn Methods	428
14.7.1	A Bootstrap Test	431
14.8	Using Different k for Different Components of x	432
14.9	Proofs	432
14.9.1	Proof of Theorem 14.1	435
14.9.2	Proof of Theorem 14.5	435
14.9.3	Proof of Theorem 14.10	440
14.10	Exercises	444
15	Nonparametric Series Methods	445
15.1	Estimating Regression Functions	446
15.1.1	Convergence Rates	449
15.2	Selection of the Series Term K	451
15.2.1	Asymptotic Normality	453
15.3	A Partially Linear Model	454
15.3.1	An Additive Partially Linear Model	455
15.3.2	Selection of Nonlinear Additive Components	461
15.3.3	Estimating an Additive Model with a Known Link Function	463
15.4	Estimation of Partially Linear Varying Coefficient Models	466
15.4.1	Testing for Correct Parametric Regression Functional Form	471

15.4.2 A Consistent Test for an Additive Partially Linear Model	474
15.5 Other Series-Based Tests	479
15.6 Proofs	480
15.6.1 Proof of Theorem 15.1	480
15.6.2 Proof of Theorem 15.3	484
15.6.3 Proof of Theorem 15.6	488
15.6.4 Proof of Theorem 15.9	492
15.6.5 Proof of Theorem 15.10	497
15.7 Exercises	502
V Time Series, Simultaneous Equation, and Panel Data Models	503
16 Instrumental Variables and Efficient Estimation of Semiparametric Models	505
16.1 A Partially Linear Model with Endogenous Regressors in the Parametric Part	505
16.2 A Varying Coefficient Model with Endogenous Regressors in the Parametric Part	509
16.3 Ai and Chen's Efficient Estimator with Conditional Moment Restrictions	511
16.3.1 Estimation Procedures	511
16.3.2 Asymptotic Normality for $\hat{\theta}$	513
16.3.3 A Partially Linear Model with the Endogenous Regressors in the Nonparametric Part	515
16.4 Proof of Equation (16.16)	517
16.5 Exercises	520
17 Endogeneity in Nonparametric Regression Models	521
17.1 A Nonparametric Model	521
17.2 A Triangular Simultaneous Equation Model	522
17.3 Newey-Powell Series-Based Estimator	527
17.4 Hall and Horowitz's Kernel-Based Estimator	529
17.5 Darolles, Florens and Renault's Estimator	532
17.6 Exercises	533
18 Weakly Dependent Data	535
18.1 Density Estimation with Dependent Data	537

18.1.1	Uniform Almost Sure Rate of Convergence	541
18.2	Regression Models with Dependent Data	541
18.2.1	The Martingale Difference Error Case	541
18.2.2	The Autocorrelated Error Case	544
18.2.3	One-Step-Ahead Forecasting	546
18.2.4	d -Step-Ahead Forecasting	547
18.2.5	Estimation of Nonparametric Impulse Response Functions	548
18.3	Semiparametric Models with Dependent Data	551
18.3.1	A Partially Linear Model with Dependent Data	551
18.3.2	Additive Regression Models	552
18.3.3	Varying Coefficient Models with Dependent Data	553
18.4	Testing for Serial Correlation in Semiparametric Models	554
18.4.1	The Test Statistic and Its Asymptotic Distribution	554
18.4.2	Testing Zero First Order Serial Correlation	555
18.5	Model Specification Tests with Dependent Data	556
18.5.1	A Kernel Test for Correct Parametric Regression Functional Form	556
18.5.2	Nonparametric Significance Tests	557
18.6	Nonsmoothing Tests for Regression Functional Form	558
18.7	Testing Parametric Predictive Models	559
18.7.1	In-Sample Testing of Conditional CDFs	559
18.7.2	Out-of-Sample Testing of Conditional CDFs	562
18.8	Applications	564
18.8.1	Forecasting Short-Term Interest Rates	564
18.9	Nonparametric Estimation with Nonstationary Data	566
18.10	Proofs	567
18.10.1	Proof of Equation (18.9)	567
18.10.2	Proof of Theorem 18.2	569
18.11	Exercises	572
19	Panel Data Models	575
19.1	Nonparametric Estimation of Panel Data Models: Ignoring the Variance Structure	576
19.2	Wang's Efficient Nonparametric Panel Data Estimator	578
19.3	A Partially Linear Model with Random Effects	584
19.4	Nonparametric Panel Data Models with Fixed Effects	586

19.4.1	Error Variance Structure Is Known	587
19.4.2	The Error Variance Structure Is Unknown	590
19.5	A Partially Linear Model with Fixed Effects	592
19.6	Semiparametric Instrumental Variable Estimators	594
19.6.1	An Infeasible Estimator	594
19.6.2	The Choice of Instruments	595
19.6.3	A Feasible Estimator	597
19.7	Testing for Serial Correlation and for Individual Effects in Semiparametric Models	599
19.8	Series Estimation of Panel Data Models	602
19.8.1	Additive Effects	602
19.8.2	Alternative Formulation of Fixed Effects	604
19.9	Nonlinear Panel Data Models	606
19.9.1	Censored Panel Data Models	607
19.9.2	Discrete Choice Panel Data Models	614
19.10	Proofs	618
19.10.1	Proof of Theorem 19.1	618
19.10.2	Leading MSE Calculation of Wang's Estimator	621
19.11	Exercises	624
20	Topics in Applied Nonparametric Estimation	627
20.1	Nonparametric Methods in Continuous-Time Models	627
20.1.1	Nonparametric Estimation of Continuous-Time Models	627
20.1.2	Nonparametric Tests for Continuous-Time Models	632
20.1.3	Ait-Sahalia's Test	632
20.1.4	Hong and Li's Test	633
20.1.5	Proofs	636
20.2	Nonparametric Estimation of Average Treatment Effects	639
20.2.1	The Model	640
20.2.2	An Application: Assessing the Efficacy of Right Heart Catheterization	642
20.3	Nonparametric Estimation of Auction Models	645
20.3.1	Estimation of First Price Auction Models	645
20.3.2	Conditionally Independent Private Information Auctions	648
20.4	Copula-Based Semiparametric Estimation of Multivariate Distributions	651

20.4.1	Some Background on Copula Functions	651
20.4.2	Semiparametric Copula-Based Multivariate Distributions	652
20.4.3	A Two-Step Estimation Procedure	653
20.4.4	A One-Step Efficient Estimation Procedure	655
20.4.5	Testing Parametric Functional Forms of a Copula	657
20.5	A Semiparametric Transformation Model	659
20.6	Exercises	662
A	Background Statistical Concepts	663
1.1	Probability, Measure, and Measurable Space	663
1.2	Metric, Norm, and Functional Spaces	672
1.3	Limits and Modes of Convergence	680
1.3.1	Limit Supremum and Limit Infimum	680
1.3.2	Modes of Convergence	681
1.4	Inequalities, Laws of Large Numbers, and Central Limit Theorems	688
1.5	Exercises	694
	Bibliography	697
	Author Index	737
	Subject Index	744