Contents

1	**Introduction**		*page* 1
	1.1	Overview	1
	1.2	Birth of the text	2
	1.3	Who will benefit	3
	1.4	Why this book is relevant	3
	1.5	Examples	4
		1.5.1 CO_2 emissions	4
		1.5.2 Age earnings	5
		1.5.3 Hedonic price function	6
	1.6	Examples in the text	7
		1.6.1 Density	8
		1.6.2 Regression	9
	1.7	Outline of the remainder of the book	9
	1.8	Supplemental materials	11
	1.9	Acknowledgments	12
2	**Univariate density estimation**		15
	2.1	Smoothing preliminaries	16
	2.2	Estimation	19
		2.2.1 A crude estimator	19
		2.2.2 Naïve estimator	22
		2.2.3 Kernel estimator	24
	2.3	Kernel selection	28
	2.4	Kernel efficiency	29
	2.5	Bandwidth selection	30
		2.5.1 Optimal selection	30
		2.5.2 Data-driven methods	33
		2.5.3 Plug-in or cross-validation?	43
	2.6	Density derivatives	45
		2.6.1 Bias and variance	47
		2.6.2 Bandwidth selection	48
		2.6.3 Relative efficiency	50
	2.7	Application	50
		2.7.1 Histograms	51
		2.7.2 Kernel densities	52

3 Multivariate density estimation 59
3.1 Joint densities 59
3.2 Bias, variance, and AMISE 62
3.3 The curse of dimensionality 64
3.4 Bandwidth selection 68
3.4.1 Rule-of-thumb bandwidth selection 70
3.4.2 Cross-validation bandwidth selection 70
3.5 Conditional density estimation 72
3.5.1 Bias, variance, and AMSE 73
3.5.2 Bandwidth selection 74
3.5.3 Inclusion of irrelevant variables 75
3.6 Application 76

4 Inference about the density 83
4.1 Fundamentals 84
4.1.1 Consistent test 86
4.1.2 Distance measures 87
4.1.3 Centering terms 89
4.1.4 Degenerate U-statistics 89
4.1.5 Bootstrap 91
4.2 Equality 92
4.3 Parametric specification 97
4.4 Independence 99
4.5 Symmetry 101
4.6 Silverman test for multimodality 102
4.7 Testing in practice 105
4.7.1 Bootstrap versus asymptotic distribution 106
4.7.2 Role of bandwidth selection on reliability of tests 106
4.8 Application 108
4.8.1 Equality 108
4.8.2 Correct parametric specification 109
4.8.3 Independence 110
4.8.4 Symmetry 111
4.8.5 Modality 112

5 Regression 113
5.1 Smoothing preliminaries 114
5.2 Local-constant estimator 117
5.2.1 Derivation from density estimators 117
5.2.2 An indicator approach 118
5.2.3 Kernel regression on a constant 118
5.3 Bias, variance, and AMISE of the LCLS estimator 120
5.4 Bandwidth selection 121
5.4.1 Univariate digression 121
5.4.2 Optimal bandwidths in higher dimensions 123
5.4.3 Least-squares cross-validation 124

5.4.4 Cross-validation based on Akaike information criteria 125
5.4.5 Interpretation of bandwidths for LCLS 126
5.5 Gradient estimation 127
5.6 Limitations of LCLS 128
5.7 Local-linear estimation 130
5.7.1 Choosing LLLS over LCLS 131
5.7.2 Efficiency of the local-linear estimator 132
5.8 Local-polynomial estimation 133
5.9 Gradient-based bandwidth selection 135
5.10 Standard errors and confidence bounds 137
5.10.1 Pairs bootstrap 137
5.10.2 Residual bootstrap 138
5.10.3 Wild bootstrap 139
5.11 Displaying estimates 139
5.12 Assessing fit 141
5.13 Prediction 141
5.14 Application 142
5.14.1 Data 143
5.14.2 Results 144

6 Testing in regression 159
6.1 Testing preliminaries 160
6.1.1 Goodness-of-fit tests 160
6.1.2 Conditional-moment test 161
6.2 Correct parametric specification 162
6.2.1 Goodness-of-fit test 163
6.2.2 Conditional-moment test 166
6.3 Irrelevant regressors 168
6.3.1 Goodness-of-fit test 168
6.3.2 Conditional-moment test 169
6.4 Heteroskedasticity 171
6.5 Testing in practice 174
6.5.1 Bootstrap versus asymptotic distribution 174
6.5.2 Role of bandwidth selection on reliability of tests 175
6.6 Application 177
6.6.1 Correct functional form 177
6.6.2 Relevance 180
6.6.3 Heteroskedasticity 180
6.6.4 Density tests 182

7 Smoothing discrete variables 187
7.1 Estimation of a density 188
7.1.1 Kernels for smoothing discrete variables 188
7.1.2 Generalized product kernel 190
7.2 Finite sample properties 191
7.2.1 Discrete-only bias 191

7.2.2 Discrete-only variance 192
7.2.3 Discrete-only MSE 192
7.2.4 Mixed-data bias 193
7.2.5 Mixed-data variance 193
7.2.6 Mixed-data MSE 193
7.3 Bandwidth estimation 194
7.3.1 Discrete-data only 195
7.3.2 Mixed data 196
7.4 Why the faster rate of convergence? 197
7.5 Alternative discrete kernels 198
7.6 Testing 199
7.7 Application 201

8 Regression with discrete covariates 205
8.1 Estimation of the conditional mean 206
8.1.1 Local-constant least-squares 206
8.1.2 Local-linear least-squares 208
8.2 Estimation of gradients 209
8.2.1 Continuous covariates 209
8.2.2 Discrete covariates 210
8.3 Bandwidth selection 212
8.3.1 Automatic bandwidth selection 213
8.3.2 Upper and lower bounds for discrete bandwidths 214
8.4 Testing 215
8.4.1 Correct parametric specification 215
8.4.2 Significance of continuous regressors 216
8.4.3 Significance of discrete regressors 217
8.5 All discrete regressors 220
8.6 Application 222
8.6.1 Bandwidths 222
8.6.2 Elasticities 223
8.6.3 Numerical gradients 223
8.6.4 Testing 225

9 Semiparametric methods 227
9.1 Semiparametric efficiency 228
9.2 Partially linear models 228
9.2.1 Estimation 229
9.2.2 Bandwidth selection 232
9.2.3 Testing 233
9.3 Single-index models 238
9.3.1 Estimation 239
9.3.2 Bandwidth selection 244
9.3.3 Testing 245
9.4 Semiparametric smooth coefficient models 247
9.4.1 Estimation 249

9.4.2 Bandwidth selection 252
9.4.3 Testing 252
9.5 Additive models 254
9.5.1 Estimation 255
9.5.2 Bandwidth selection 258
9.5.3 Testing 259
9.6 Application 261
9.6.1 Bandwidths 261
9.6.2 Plotting estimates 263
9.6.3 Specification testing 264

10 Instrumental variables 267
10.1 The ill-posed inverse problem 268
10.2 Tackling the ill-posed inverse 270
10.3 Local-polynomial estimation of the control-function model 272
10.3.1 Multiple endogenous regressors 274
10.3.2 Bandwidth selection 275
10.3.3 Choice of polynomial order 276
10.3.4 Simulated evidence of the counterfactual simplification 278
10.3.5 A valid bootstrap procedure 279
10.4 Weak instruments 280
10.4.1 Weak identification 282
10.4.2 Estimation in the presence of weak instruments 284
10.4.3 Importance of nonlinearity in the first stage 286
10.5 Discrete endogenous regressor 286
10.6 Testing 287
10.7 Application 288

11 Panel data 293
11.1 Pooled models 294
11.2 Random effects 295
11.2.1 Local-linear weighted least-squares 297
11.2.2 Wang's iterative estimator 298
11.3 Fixed effects 301
11.3.1 Additive individual effects 302
11.3.2 Discrete individual effects 305
11.4 Dynamic panel estimation 306
11.5 Semiparametric estimators 308
11.6 Bandwidth selection 309
11.7 Standard errors 309
11.7.1 Pairs bootstrap 310
11.7.2 Residual bootstrap 310
11.8 Testing 311
11.8.1 Poolability 311
11.8.2 Functional form specification 313
11.8.3 Nonparametric Hausman test 315

	11.9	Application	316
		11.9.1 Bandwidths	317
		11.9.2 Estimation	318
		11.9.3 Testing	318
12	**Constrained estimation and inference**		321
	12.1	Rearrangement	322
		12.1.1 Imposing convexity	324
		12.1.2 Existing literature	325
	12.2	Motivating alternative shape-constrained estimators	326
	12.3	Implementation methods via reweighting	330
		12.3.1 Constraint-weighted bootstrapping	330
		12.3.2 Data sharpening	330
	12.4	Practical issues	331
		12.4.1 Selecting the distance metric	331
		12.4.2 Choice of smoothing parameter	332
		12.4.3 Linear in p implementation issues	333
		12.4.4 Imposing additive separability	336
	12.5	Hypothesis testing on shape constraints	337
	12.6	Further extensions	338
	12.7	Application	339
		12.7.1 Imposing positive marginal product	339
		12.7.2 Imposing constant returns to scale	340
Bibliography			343
Index			359