

Index

- adaptive estimators, 12–13, 17, 31, 74, 76–7, 91, 217–25, 247, 277, 284
- affinity measures, 61, 64–5, 69, 141–50
- ALS estimator, 240–1, 242
- Andrews's MINPIN estimators, 279–80, 381–2
- anticipations, 212
- applications of estimators
 - conditional moment, 94, 134, 150–9
 - density, 5, 71–7
 - derivative, 190–5
 - discrete choice, 299
 - nonparametric regression, 157–9
 - selection-model, 315–16
 - single-equation, 248–53
- approximation
 - to bias and variance, 23–9, 96–106
 - Hermite, 15
 - likelihood, 225–7
 - to MISE, 24–25, 27, 30, 50, 53, 58
 - Nadaraya–Watson estimator, 96–104
 - score, 227–34
 - semiparametric efficiency bound, 246–8
- asset-price models, 134–5
- asymptotic covariance matrix, 192, 212–14, 309, 313, 337–8
- asymptotic properties
 - conditional moment estimator, 108–18
 - density estimator, 6, 21, 32–43, 48, 73
 - for dependent observations, 370–1
 - of derivative estimators, 177–89
 - of kernel estimators, 21, 32–43, 48, 49, 73, 108–16, 178–82, 184–9
 - of series estimators, 49, 126–33, 182
- Atkinson's inequality measure, 156–7
- Bayesian methods, 19
- Bennett's inequality, 36, 366–7
- Berge inequality, 365
- Bernoulli random variable, 274
- Bernstein polynomials, 137
- Berry–Esseen theorem, 367–8
- best instrumental variable estimator, 257, 259, 260, 261–2
- bias
 - approximation to, 23–9, 96–104
 - asymptotic, 73, 118, 123, 128, 184, 188
 - exact, 21–3, 43, 48
 - of kernel estimator, 4, 21–32, 43, 48, 96–106
 - misspecification, 106
 - normalized, 41, 42, 110
 - reduction of, 29–32, 41–2, 66, 106
 - in series estimator, 129, 135
 - small-sample, 31, 54
- binary choices, *see* discrete choice models
- bootstrap procedures, 42, 43, 65, 120, 146, 241–2, 263, 287, 290, 294, 299
- Borel–Cantelli lemma, 36, 353
- Borel measurable functions, 32, 43, 45, 109, 177, 348–50
- Bowden–Turkington method, 267–8
- Cantelli inequality, 366
- Carbon's inequality, 375
- Cauchy–Schwarz inequality, 128, 132, 205
- censored random variable, 324–6
- censored regression estimators, 3, 317–18
 - comparison of, 337–8
 - density-based, 326–9
 - efficiency bounds, 322–4
 - Kaplan–Meier, 324–6
 - least absolute deviation, 330–3
 - least squares, 327–8, 333–6
 - maximum likelihood-type, 329
 - Newey's efficient, 336–7
 - nondensity-based, 329–37
 - parametric, 319–21
 - quantile, 330, 333
 - semiparametric, 322–4, 326–37
- center-free test statistic, 66
- central limit theorems, 40, 63, 69, 110, 115, 116, 143, 187, 200, 202, 203, 257, 298, 357–60, 375–6, 378–9

420 **Index**

- characteristic function, 351–2
 Chebychev's inequality, 35, 109, 179, 350, 365–6
 Chebychev's theorem, 355
 Cobb–Douglas approximations, 191–2
 Collomb's inequality, 375
 conditional moment estimators, 2, 15, 78–9, 166
 affinity measurement, 141–50
 applications, 94, 134, 150–9
 asymptotic properties, 108–18, 126–33
 efficiency bounds, 246–8
 finite-sample properties, 95–108
 fixed-design, 88–9
 formal approach, 83–4
 general class of, 84–6
 higher-order, 95
 implementing, 118–22, 133–137
 kernel, 79–95, 108–22
 local linear, 93–4
 with marginal density known, 87–8
 nearest neighbor, 89–91
 nonparametric, 81–3, 84–95, 106–8, 122–3, 141–50
 parametric, 80–1, 106–8, 141–50
 regression, 81–3, 94, 139–40, 157–9
 restrictions, 245, 256
 robust, 122–3
 semiparametric efficiency bound, 246–8
 series, 78, 94, 123–40
 spline smoothing, 91–3
 structure imposed in, 137–40
 window width, 81, 82, 118–22
 confidence intervals, 42–3, 118
 consistency, 34–48, 49, 113, 116–17, 126, 131–2, 133, 178–9, 259, 271, 282, 284, 286–96
 continuity, 10, 356–7
 convergence
 to asymptotic distribution, 65
 bounded, 33
 central limit theorems, 357–60
 in density estimation, 28, 30, 33, 34, 48–9, 51, 52, 58–9, 65
 derivative estimator, 282
 distribution functions, 355–7, 360–1
 laws of large numbers, 354–5, 360–1
 of mean squared error, 58–9, 189
 modes of, 32
 moments, 114, 117, 126, 131–2, 133, 361–5
 to normality, 114
 parametric estimators, 27 n.5, 172–3
 random variables, 352–4
 of sample variance, 52
 Cosslett's estimators, 292–3, 306
 Cramer–Rao bound, 3, 218, 265, 278
 Cramer–Wold theorem, 356
 cross validation, 50–4, 57, 77, 92, 119, 120
 cumulative distribution function, 273–4, 324–6
 density estimation methods, 2, 5–7, 83; *see also* kernel estimator
 affinity measures, 7, 64–5
 applications, 5, 71–7
 for derivatives of a density, 6, 19–20, 23, 56–7, 166
 histogram approach, 6, 7–8
 hypothesis testing, 60–71
 kernel estimator, 6, 7, 9–10, 14, 17, 19, 20–56, 66, 73–5, 167
 likelihood estimators, 15–19, 31
 multivariate, 7, 11–12, 57–60
 nearest neighbor estimator, 11–12
 nonparametric, 6, 7–19, 74
 parametric, 6
 semiparametric, 326–9
 series estimators, 13–15
 symmetric, 10, 29
 window-width, 7, 10, 12–13, 19, 26, 27, 41–2, 43 n.9, 49–54, 56–7, 77
 dependent observations, 43–9, 115–16, 120–1, 133, 235–7, 370–1
 derivative estimators
 applications, 160–1, 190–5
 asymptotic properties, 177–9, 184–9
 average derivatives, 167–70, 172–3, 184–9, 190, 192
 of density, 6, 19–20, 23, 56–7
 for discrete choice models, 294–5
 formulas, 161–3
 higher order, 59, 88, 171, 182–3
 hypothesis testing, 176–7
 implementing, 189–90
 kernel, 164–6, 178–82, 184–9
 local linear, 170–1, 183–4
 partial derivatives, 161–7, 174–5, 177–84, 190–2
 pointwise, 172–3, 190–2
 of regression function, 19
 restricted, 173–6, 182–3
 series, 167, 182
 window widths, 56–7, 189–90
 diagnostic tests, 208–10, 234–5, 242–43
 Dickey–Fuller density estimation, 74–7
 discrete-choice estimators, 3, 82, 272–3
 applications, 299
 binary, 273–86
 derivative-based, 294–5
 with discrete explanatory variables, 295–6
 efficiency bounds, 275–9
 Ichimura's estimator, 280–3
 iterative least squares estimator, 293–4
 Klein–Spady estimator, 283–5
 maximum likelihood, 285–6, 292–3
 maximum score estimators, 286–91
 multinomial, 296–7
 parametric, 273–5
 rank correlation, 291–2
 regression, 172
 seminonparametric, 285–6
 semiparametric, 275–89
 specification tests, 297–9
 discrete explanatory variables, 295–6

Index

421

- distributional theory, 39–40
 distribution
 asymptotic, 65, 257
 change in policy as shift in, 84
 conditional, 347–8
 function, 324–6, 345–7, 355–7
 kernel estimator, 113
 dominated convergence theorem, 178, 353
- earnings–age profile, 152–7, 192–4
 Edgeworth approximation, 207–8, 308
 efficiency bounds, 3, 26
 binary discrete choice models, 273, 275–9
 censored regression model, 322–4
 moment-based estimators, 230–4, 245–8
 semiparametric, 26, 219, 246–8, 273, 275–9, 285, 297, 322–4
 single-equation models, 245–8
 efficient estimators, 214–34, 247–8, 264–7, 336–7
 efficient scores, 221–2
 electricity demand/temperature model, 251–2
 empty-space phenomenon, 59–60
 Epanechnikov kernel, 28, 32, 55, 59, 75, 290
 equicontinuity, 377, 379–81
 ergodicity, 371–2
- finite difference approach, 20
 finite-sample properties
 bias and variance, 21–32, 96–104
 combining parametric and nonparametric estimators, 106–8
 conditional moment estimation, 95–108
 density estimation, 20–32
 kernel estimators, 20–32
 local linear regression estimator, 104–6
 Nadaraya–Watson estimator, 96–104
 simultaneous equation models, 267–9
 Fisher’s information matrix, 3, 218
 fixed design estimators, 88–9
 flexible Fourier form, 123–4, 135, 151, 153
 full information model, 260, 271
 functions
 Borel measurable, 32, 43, 45, 109, 177, 348–50
 density, 10, 29, 61–6, 288
 distance, 17, 54
 distribution, 324–6, 345–7
 indicator, 287–8
 influence, 210
 Lipschitz, 377
 moment, 262
 penalty functions, 119–20
 production, 78, 123, 190–2, 194
 regression, 19
 smooth, 287–8
- Gallant–Nychka’s estimator, 310–11, 316
 Gasser–Muller estimator, 89, 92 n.2, 121
 Gaussian kernels, 74, 75, 76–7, 151, 184, 250, 290
 generalized
 additive models, 134, 137–9
 Hajek representation theorem, 217–18
 instrumental variables estimator, 256
 least squares estimator, 123, 196–7, 215, 248–50, 327–8
 method of moments, 231–2, 233–4, 247, 250–1, 256–7, 260–1, 266, 279, 310
 Georgiev–Greblicki inequality, 367
- Han’s maximum rank correlation estimator, 291–2
 Hardle–Stoker estimators, 169, 175, 189
 Helly–Bray theorem, 356
 Hermite polynomials, 14–15, 226
 hetroskedasticity, 103–4, 126, 127, 132–3, 142–3, 145, 156, 206 n.1, 207, 214–17, 235, 237, 248–51, 264, 272, 273, 275, 309
 histogram approaches, 6, 7–8, 52, 59
 Hoeffding inequality, 366
 Holder’s inequality, 350
 homogeneity, 173
 Horowitz’s smoothed maximum score estimator, 287–91, 296
 housing price models, 250–1, 252, 299
 hypothesis testing, 60–71, 176–7
- Ichimura’s estimator, 279, 280–3, 286, 297
 identification issues, 269–70, 275, 276, 305
 independence, 60, 69–71, 203, 233, 264, 281–2, 347–8
 independent observations, 32–43, 108–15, 120–1
 indicator function, 287–8
 inequalities, 23, 35, 109, 128, 132, 156–7, 205, 350–1, 365–8, 375, 377–8
 influence function, 210
 instrumental variables estimators, 256, 257, 260–1, 262, 266–7, 268, 307–8
 integrated conditional moment tests, 147–8
 integrated squared error, 24, 49–51
 intercepts, 198–9, 220–1, 315
 interest rate models, 61
 iterative least squares estimator, 293–4
- jackknife estimators, 31–2, 180–1, 190, 298
 Jensen’s inequality, 23, 351
- Kaplan–Meier estimator, 324–6
 kernel estimators
 adaptive, 250–1
 asymptotic properties, 21, 32–48, 49, 73, 108–18, 178–82, 184–9
 bias and variance, 4, 21–32, 48, 96–106, 180–1, 290
 conditional moment, 79–95, 118–22
 density, 6, 7, 9–10, 14, 17, 19, 20–56, 66, 73–5, 168
 dependent observations, 43–9, 108–15
 derivative, 164–6, 177–82, 184–9

422 **Index**

- kernel estimators (*cont.*)
 finite-sample properties, 20–32
 formal derivation, 83–4
 higher order, 29, 42–3, 54–6, 73, 76, 77, 113, 118,
 139, 180–1, 284, 290, 298
 nonparametric, 81–94
 parametric, 66, 80–1, 260–1
 recurrence relationship, 56
 recursive, 86–7
 sampling properties, 42–9, 54
 score approximation, 227–30
 selectivity models, 306–8
 simultaneous equations, 267
 single-equation, 199–200, 227–30
 unbiasedness, 33–4, 43, 48, 49
 variable, 12–13, 21–9, 31
 window-width, 7, 118–22
- Kelejian–Armeyia method, 267–8
 Khinchin’s theorem, 131, 355
 Klein–Spady estimator, 273, 279, 283–5, 286, 297, 299
 Kolmogorov’s inequality, 365
 Kolmogorov–Smirnov test statistic, 5, 64–5
 Kolmogorov’s theorem, 355
 Kullback–Leibler
 distance function, 17, 54
 divergence measure, 71 n.16
 information measure, 61, 238
- large-sample properties, 177–84
 latent variable model, 273
 law of large numbers, 200, 202, 203, 204, 354–5,
 360–1
 least absolute deviation estimator, 239–40, 241, 243,
 330–3
 least squares
 cross validation, 50–2
 estimators, 85, 107, 114, 123, 258–60, 261–2,
 270–1, 293–4, 327–8, 333–6
 leave-one-out estimator, 119, 144, 145, 185, 199,
 268
 Liapunov’s
 central limit theorem, 40, 110, 358–9
 inequality, 350
 likelihood approximation, 225–7
 likelihood cross validation, 53–4
 likelihood estimators, 15–19, 31, 237
 Lindberg–Feller theorem, 41 n.7, 111, 357–8
 Lindberg–Levy theorem, 357
 linear equality restrictions, 174–5
 linear inequality restrictions, 175–6
 linear regression models, 2, 158–9, 198–214, 222,
 234–5
 Linton’s approximation, 208
 Lipschitz function, 186, 271, 377
 local histogram approach, 7–8
 local-linear estimators, 93–4, 95, 104–8, 158–9,
 170–2, 183–4
 local log-likelihood estimators, 17–19, 31
 local regularity, 217
 log-likelihood estimators, 15–19
- M*-estimators, 198, 237–45, 252–3, 266, 286, 291–2,
 329
 Mack–Muller derivative estimator, 169
 macro models, 210–12
 magnitudes, order of, 133–4, 368–9
 Mann–Wald theorem, 183
 Manski’s maximum score estimator, 286–7
 Markov inequality, 39, 350
 Martingale differences, 377–9
 maximum likelihood estimators, 3, 16, 17, 192, 197,
 235, 272, 274–5, 285–6, 292–3, 303–4, 310–14,
 329
 maximum rank correlation estimator, 291–2
 McLeish’s theorem, 375
 mean integrated absolute error (MIAE), 50
 mean integrated squared error (MISE), 24–25, 28, 32,
 50
 method of moments estimators, 85, 231–2, 233–4, 247,
 250–1, 256–7, 260–1, 266, 310
 Minkowski’s inequality, 350
 mixing sequences, 133, 372–6
 mixingales, 377–9
 moment-based estimators, 230–4, 245–8, 361–5;
see also conditional moment estimators;
M-estimators
 monotone convergence theorem, 354
 Monte Carlo analysis, 5, 52, 71, 74, 88, 121, 135,
 136, 147, 170, 190–2, 207, 230, 236, 267,
 297
 Muntz–Satz polynomials, 124, 127, 137
 multi-indices, 124–5, 136
 multinomial discrete choice models, 296–7
 multivariate estimation, 7, 11–12, 57–60, 181, 188
- Nadaraya–Watson estimator, 79, 83, 84, 86, 88, 90, 93,
 95, 96–104, 105, 107–14, 122, 164, 170
 naïve estimator, 8, 11, 12, 15
 near-epoch dependent sequences, 376–7
 nearest neighbor estimators, 11–12, 49, 89–91, 158,
 258, 262–3
 neural networks, 124, 140
 Newey’s estimators, 262–7, 308–10, 311–14, 315, 316,
 336–7
 nonlinear instrumental variable (NLIV), 256
 nonparametric estimators
 conditional moment, 81–3, 84–95, 106–8, 122–3,
 141–50
 density, 6, 7–19, 74
 general class of, 84–6
 kernel, 81–94
 parametric combined with, 106–8
 regression, 157–9, 208–10
 simultaneous equation models, 269–71
 single-equation, 201–2
 two-stage least squares estimator, 270–1

Index

423

- normality, 39–42, 48, 49, 66, 117–18, 126–33, 178–9, 182, 187–8, 259, 284
- options pricing, 160–1
- ordinary least squares estimators, 152, 197, 199–200, 223, 245, 250–1
- orthogonal series estimator, 13–14, 49, 137
- Owen's inequality, 367
- parametric estimators
 binary discrete choice, 273–5
 censored regression, 319–21
 conditional moment, 80–1, 106–8, 141–50
 convergence of, 27 n.5, 172–3
 density, 6
 kernel, 66, 80–1, 258–9
 nonparametric combined with, 106–8
 selectivity, 301–4
 simultaneous equations, 256–8, 260–1
 single-equation, 200–1, 256–8
 submodels, 217–22, 265, 276
- partial derivative estimators, 161–7, 177–84
- penalized likelihood estimators, 15–17, 237
- penalty functions, 119–20
- Pitman alternatives, 64, 65, 148
- plug-in methods, 26, 50, 53, 104, 120, 190, 291, 332
- polynomial estimators, 94, 106–7, 123–4, 171, 267, 268; *see also* series estimators
- Powell's estimators, 306–8, 316, 330–6
- price elasticity, 194–5
- Priestly–Chao estimator, 91, 121
- probability
 concepts, 342–52
 density functions, 60
 inequalities, 365–8, 377–8
 as set function P , 343–5
 weights, 85
- product kernels, 59, 70
- product-limit estimator, 324–6
- production functions, 78, 123, 190–2, 194
- projection pursuit regression, 139–40
- propensity score, 308
- quantile estimators, 60, 78, 93, 158, 198, 239–40, 241, 242, 252–3, 333
- random variables, 122–3, 180, 197, 200, 207, 224, 272, 273–4, 276, 282, 324–6, 345–7, 352–4
- rank correlation estimator, 291–2
- recurrence relationship, 56
- recursive estimators, 13, 86–7, 117
- regression models, *see also* censored regression models
 conditional moment, 15, 78, 81–3, 104–8, 120, 134, 139–50, 157–9
 diagnostic tests after, 208–10
 discrete choice, 274–5, 283
 efficient estimators, 225–34
 linear part of model, 198–214
 local linear, 95, 104–8
 nonlinear, 274–5, 283
 nonparametric, 157–9, 208–10
 parameters with unknown error density, 225–34
 projection pursuit, 139–40
 residuals, 66, 67–8
 semiparametric, 198–214, 225–34
 restrictions, 48, 173–6, 182–3, 198–9, 200, 209–10, 217, 229, 234, 245, 247–8, 256, 264, 266, 270, 282
 reverse causality (feedback), 254
 Rilstone's estimators, 189, 258–60
 robust estimators, 26, 122–3, 287
 Rosenblatt–Parzen estimator, 9–10, 18
 Rosenblatt's measure of dependence, 379
 Rotterdam system, 78
- sample selection, 126, 321
- sample size, 2, 267–8, 284; *see also* large sample properties; small sample properties
- sample space, 167, 342–3
- sampling properties, 43–9, 243–4
- saturation ratio, 135
- scale parameters, 137, 223–4, 234
- scores, 3, 6, 19, 21, 167, 218, 221–2, 225, 227–34, 235, 265, 277, 286–91, 308
- seemingly unrelated regressions (SUR) estimator, 247
- selectivity estimators, 3, 126
 applications, 315–16
 intercept estimation, 315
 kernel, 306–8
 maximum likelihood, 310–14
 parametric, 301–4
 sequential semiparametric, 304–10
- semiparametric methods, 15, 124, 196, 285
- semiparametric estimators, 19, 32, 196
 adaptive, 217–25, 235–7
 binary discrete choice, 275–86
 censored regression, 322–4, 326–37
 conditional moment, 125–6
 density-based, 326–9
 with dependent observations, 235–7
 efficiency bounds, 26, 219, 246–8, 273, 275–9, 285, 297, 322–4
 linear regression, 198–214, 225–35
 simultaneous equations, 258–60, 261–7
 single-equation, 198–225, 245–53
 three-stage least squares, 261–2
- separability, 134, 135
- sequences, near-epoch dependent, 376–7
- sequential estimators, 200, 243–5, 304–10
- series estimators, 263, 268
 applications, 315
 asymptotic properties, 126–33, 182
 conditional moment, 78, 119, 123–6

424 **Index**

- series estimators (*cont.*)
 density, 13–15, 49
 with dependent observations, 49, 133
 derivative, 167, 182
 implementing, 133–7
 with independent observations, 126–33
 selectivity models, 308–9
- Shannon's entropy, 17
- Sheather–Jones method, 53
- Shibata's criterion, 119
- simultaneous equation models, 3, 254–5
 efficient distribution-free estimators, 264–7
 finite-sample properties, 267–9
 identification of structural function, 269–70
 Newey's estimators, 262–7
 nonparametric, 269–71
 parametric, 256–8, 260–1
 Rilstone's estimator, 258–60
 semiparametric, 258–60, 261–7
 single-equation, 255–60
 systems of equations, 260–7
 three-stage least squares, 261–2
 two-stage least squares, 258–60, 270–1
- single-equation models, 2–3, 196–8, 254
 adaptive, 217–25, 235–7, 250–1
 applications, 248–53
 asymptotic covariance matrix, 212–14
 with dependent observations, 235–7
 diagnostic tests, 208–10, 234–5, 242–3
 efficiency bounds, 245–8
 efficient, 225–34
 heteroskedasticity and, 214–17, 248–50
 kernel-based, 199–200, 227–30
 likelihood approximation, 225–7
 linear regression, 198–214, 234–5
M-estimators, 237–45
 macro models, 210–12
 moment-based, 230–4, 245–8
 nonparametric, 201–2
 parametric, 200–1, 256–8
 regression parameters with unknown error density, 225–34
 Rilstone's estimator, 258–60
 of scale parameters, 234
 score approximation, 227–34
 semiparametric, 198–217, 245–53
 simultaneous equation models, 255–60
 two-stage least squares, 258–60
- single-index models, 126, 172, 274, 276, 295, 298, 299
- small sample properties, 31, 42–3, 47–8, 54, 65, 167
- smooth function, 287–8
- specification tests, 297–9
- spline smoothing, 91–3, 329
- stationarity, 370–1
- statistical methods
 asymptotic theory for dependent observations, 370–82
 convergence, 352–65
 order of magnitudes, 368–9
 probability concepts, 342–52
 probability inequalities, 365–8
- stochastic
 differentiability condition, 241
 equicontinuity, 202, 203, 207, 282, 379–81
- stock market returns, 71–4, 158
- strike duration model, 150–2, 153
- structural equation, 269–70
- structure on conditional moments, 137–40
- symmetric
 censored least squares estimators, 333–6
 density function, 10, 29
 kernel, 83–4, 189
- symmetry, 67–8, 173, 194–5, 233, 264, 277, 336–7
- Theil's entropy measure, 156–7
- three stage least squares estimator, 261
- Trans-Log approximation, 191–2
- travel expenditure/income model, 248–51
- trimming, 122, 148, 168, 229, 230, 259, 282–3
- twicing procedure, 118–19
- two-stage least squares estimators, 256, 258–60, 267, 270–1
- U*-statistics, 63, 69, 143, 185, 186, 359–60
- unbiasedness, 33–4, 43, 48, 49
- Vinod–Ullah estimator, 169
- volatility measure, 210–11, 252
- wages of married women, 315–16
- Wald statistic, 176, 182, 238
- window-width, 7
 conditional moment methods, 81, 82, 118–22
 cross validation, 50–4, 57, 77
 density estimation, 41–2, 43 n.9, 49–54, 56–7
 for derivatives of density, 56–7, 189–90
 plug-in methods, 50, 53
 selection procedures, 49–57, 77, 118–22
 variable, 12–13