

# INDEX

- Ad valorem tax, 303
- Adulterated liquor, 290, 292, 295
- Aggregate shares, 140
  - for simulation study, 157
- Analytic inference. *See* Model-based inference
- Applied economic research, 175
- ARMA models, 110–111
- Asymptotic
  - design-based variance, 200
  - linearization representation, 217
  - properties of estimators, 214
  - theory for  $M$ -estimation, 146–147
- Asymptotic distribution, 216
  - for matching estimators, 210
- Asymptotic variance, 230
  - estimator, 119
  - formula, 118
- Auditory mode, 22
- Automated clearing house (ACH), 245
- Auxiliary variables, 90
- Average bid auctions (ABA), 80
- Average partial effects, 119
  
- Bahadur-type representation, 217
- Balanced repeated replication (BRR), 96
- Bandwidths, 194–196
  - selection, 186–188
- Bank of Canada, 36, 47, 98, 104
- Bernoulli distributions, 99
- Bernoulli log likelihood, 128
- Bernoulli sampling. *See* Variable probability sampling (VP sampling)
- Bias function, 141
- Bias-adjustment, 215
- “Biased sampling” problems, 147
- “Big data” sets, 37–38
- Binary responses, 128
- Binary variables, 311
- Black market liquor in Colombia, 301–302
  - adulterated sales, 302
  - average marginal effects for contraband and adulterated liquor–levels, 307
  - average marginal effects for MICE, 309
  - coefficients for specifications in salary imputation model, 308
  - contraband goods, 288–289
  - data, 290–292
  - multiple imputation, 297–301, 310–313
  - percentage of adulterated purchases, 304
  - percentage of contraband purchases, 303
  - percentage of illegal purchases, 289
  - residual analysis, 313–314
  - results, 292–296
  - “Blocking” strategy, 247

- Bootstrap, 100
  - conventional, 211
  - sample, 211
  - variance estimation via, 96–98
  - wild bootstrap, 61–84, 211, 219
- bsweights* command, 103
- Business
  - heterogeneity, 240
  - surveys, 239–240
- Calibration
  - equations, 92
  - procedure, 88
- Canadian Internet Usage Survey (CIUS), 98
- Card acceptance, 36, 39, 49
  - estimates, 42–43
  - and usage, 41
- Card payments, regression models of, 43–46
- Card-acceptance model, 45
- Cash on hand* variable, 89, 100–101
- cemplik2* function, 171
- Census region, 15
- Checkable deposits (CHKD), 241
- Choice-based sampling, 162*n*6
- Classical nearest neighbor imputation, 213
- Classical raking estimator, 90–92
- Cluster-means regression (CMR), 74
- Cluster-robust inference, 63
  - WCB, 65–67
- Cluster-robust standard error, 67
- Cluster-robust test statistics, 62
- Cluster-robust variance estimator (CRVE), 64
- Colombian National Administrative Department of Statistics (DANE), 290–291
- Combined inference approach, 176–177
- Commercial banks (CMB), 241
- Complementary log–log (cloglog), 294
- Complex
  - sampling plan, 174
  - survey methods, 122, 176
- Computation, 169–171
- Conditional density, 141
- Conditional mean, 127
  - functions, 130
- Conditional moment restriction models, 138
  - conditional moment equalities, 139–140
  - endogenous and exogenous stratification, 143–146
  - identification, 142–143
  - simulation study, 156–161
  - VP sampling, 140–142
- Contactless (tap-and-go) credit card usage binary variable, 89
- Continuous variables, 89
- Conventional bootstrap, 211
- Conventional jackknife variance estimation, 221
- Conventional replication method, 211
- Converged estimator, 94
- Correction factor, 185
- Correlated random effects approach, 128
- County, 39
  - data, 56
  - effect, 43
- Covariance matrix, 63
- Covariate information, 215
- Cox approach, 110
- Credit cards, 98
- Credit unions (CUS), 241
- Crimes, 264, 272
- Criminal activity, 270
- Criminal Procedure Law, 263

- Cross-validation function, 187
- Cumulative standard normal (CDF), 73
- Current Population Survey (CPS),
  - 6, 110
  - socioeconomic variables in, 11–14
- Data-driven approach, 177, 187
- Data-generating processes (DGPs),
  - 177, 190
  - in Monte Carlo simulations, 188
- Degree of “desirability”, 8–9
- Demographic variables, 11–12
- Dependent variable, 51, 55
- Depository and financial institutions, 240
- Descriptive inference. *See* Design-based inference
- Design-based approach, 180
- Design-based estimators, 174, 176
- Design-based inference, 88, 176
- Design-consistent estimator, 176
- Difference in objective functions, 119
- Difference-in-differences (DiD), 62
- Discounts Asked* indicator, 292, 297
- Discounts Offered* indicator, 292
- Discrete variables, 89
- Drugs and weapon crimes, 264
- Dynamic regression models, 110–111
- “Econ Plus” set, 98
- Econometric methods, 62
- Econometric models, 140
- Economic theory, 174, 260
- Efficiency bounds, 148–152
- E–M algorithm-based imputation method, 244
- Empirical application of 2013 MOP, 98–103
- Empirical distribution, 67
- Empirical likelihood approach, 148, 155
- Empirical loglikelihood (EL), 169
- Empirical researchers, 210
- Endogeneity, 67
- Endogenous sampling, 175
- Endogenous stratification, 143–146, 150–151, 157, 262
  - median bandwidths for WLC and LC under, 194
  - median MSE for WLC, LC, WLS and OLS under, 190
- Estimation approach, 241
- Estrato* (strata) variable, 292, 299
- Exogenous stratification, 111, 126–127, 143–146, 150–151, 157
  - comparing asymptotic variance of LS and GMM estimators under, 166–169
  - median bandwidths for WLC and LC under, 194
  - median MSE for WLC, LC, WLS and OLS under, 191
- Explanatory variables, 41
- Fake goods, 288
- Federal Reserve System, 238, 240
- Finite population correction (FPC), 291–292
- Finite population estimator, 176
- Finite population parameter, 212
- First price auctions (FPA), 79–81
- Four-digit primary industry code, 39
- Fractional responses, 128
- Full-coverage items, 247
- Gamma distributions, 129
- Gaussian kernel, 196
- Gauss–Markov theorem, 163n16
- General M-estimation, 111

- Generalized method of moments
  - estimator (GMM estimator), 147, 150, 151
- Generalized regression estimation (GREG), 89, 93–94
- Geographic information, 97
- Geometric quasi-MLE, 111
- “Group treated” dummy (GTg), 64
- Hajek estimator, 217
- Hausman specification tests, 273
- Health and Retirement Study (HRS), 5–7
  - socioeconomic variables in, 11–14
  - survey outcomes in UAS and, 14–26
- Heckman models, 299
- Heckman two-step estimator, 312
- Heteroskedasticity-robust standard errors, 63–64, 74
- Higher order orthogonal polynomials, 45
- Hit rates test, 262, 267
- Horvitz–Thompson estimator, 218
- Horvitz–Thompson weight (H–T weight), 90
- Hungarian payments system, 38
- Illegal
  - alcohol, 290
  - cigarettes, 290
  - goods, 288
  - liquor, 288
- Imputation models, 238, 239, 297, 299, 310–311
- Indicator
  - function, 179
  - variables, 90, 94
- Industry, 39
  - effect, 43
  - industry-based stratas, 44
- Inference, 138, 146
  - efficiency bounds, 148–152
  - efficient estimation, 152–155
  - modes, 176
  - related literature and contribution, 146–148
  - testing, 155–156
  - see also* Conditional moment restriction models
- Informative sampling, 175
- Internet match high-quality traditional surveys
  - health insurance coverage, 30
  - home ownership, 30
- HRS, 6–7
  - individual earnings, 31
  - methods and outline, 5–6
  - predicted mean health status by age and survey mode, 32
  - predicted probability, 33
  - satisfied with life, 32
  - self-reported health, 30–31
  - socioeconomic variables in HRS, UAS and CPS, 11–14
  - survey outcomes in HRS and UAS, 14–26
  - UAS, 7–11
  - whether retired, 31
- Interview
  - face-to-face, 4, 18, 22
  - mode, 4–5
  - online, 14
  - telephone, 4
- Inverse Mills ratio, 56
- Inverse probability weighted estimator (IPW estimator), 162n12
- ipfraking* command, 103
- Item response rate, 244–245
- Iterative E–M algorithm approach, 241

- Iterative estimators of regression coefficients, 146
- Iterative proportional fitting (IPF), 88
- Jackknife method, 218
- Jackknife variance
  - estimation, 221
  - estimator, 96–97
- Kernel function, 179
  - assumptions for, 230–231
- “Kernel matching” estimators, 211
- Kernel-based derivative estimation, 219
- Kernel-smoothed *p* values, 76
- Knowles, Persico, and Todd test (KPT test), 261, 278–279
- Kullback–Leibler Information Criterion (KLIC), 110, 112
- Labor market duration (LMD), 178, 196
- Lagrange multipliers, 154
- Large-scale surveys, 174
- Least absolute deviations (LAD), 120
- Least squares cross-validation (LSCV), 177, 189, 204–208
- Leave-one-out kernel estimator, 204
- Length biased sampling problem, 147
- Linear regression model, 63, 139, 144–146
- Linearization
  - of estimator, 223
  - methods, 89
  - variance estimation via, 95–96
- Local constant estimation (LC estimation), 188
  - median bandwidths, 194–196
  - median MSE for, 190–194
- Local constant estimator, 175–176, 178–180
- Local empirical loglikelihood, 155
- Log-likelihood functions, 110, 270
  - negative of, 113–114
- Logistic regression analysis, 37
- Logistic regression models, 39–40, 51
  - card acceptance, 51–55
  - card usage, 55–57
- Logit, 311
  - predictive model, 299
- Lognormal distributions, 129
- Lyapunov DoubleArray Central Limit Theorem, 202–203
- Mahalanobis distance, 213
- Marginal density, 141
- Matching
  - discrepancy, 215
  - estimators, 210
  - scalar variable, 221
  - variable, 215
- Maximum likelihood raking estimator, 92–93
- Median absolute deviations (MAD), 190
- Medicare, 17
- Misclassification, 260–261
- Missing at random (MAR), 213, 297
- Missing completely at random (MCAR), 297
- Missing data process, 213
- Missing not at random (MNAR), 297
  - data, 312
  - see also* Planned missing data design
- Missing-by-design approaches, 239
- Mode effects, 21–26
- Model-assisted estimator, 176
  - local constant estimator, 180
- Model-assisted nonparametric regression estimator, 178
  - asymptotic properties, 180–186

- local constant estimator, 178–180
- model-assisted local constant estimator, 180
- Model-based inference, 176, 180
- Model-selection statistic, 122
- Model-selection tests
  - for complex survey samples, 110
  - examples, 128–129
  - exogenous stratification, 126–127
  - nonnested competing models and null hypothesis, 111–114
  - with panel data, 124–126
  - rejection frequencies, 131–132
  - small simulation study, 129–133
  - statistics under multistage sampling, 122–124
  - under stratified sampling, 114–122
- Modern imputation methods, 239
- Modified kernel density estimator, 204
- Modified plug-in method, 186
- Money market deposits (MMDA), 241
- Monte Carlo simulations, 188, 212
  - bandwidths, 194–196
  - DGPs in, 188, 190
  - median bandwidths for WLC and LC, 194–196
  - median MSE for WLC, LC, WLS and OLS, 190–193
  - sample mean squared error, 189–194
  - strata borders, 189
- Multiple imputation, 242, 297–301
  - See also* Nearest neighbor imputation
- Multiple Imputation by Chained Equations (MICE), 310
  - average marginal effects for, 309
  - binary variables, 311
  - imputation models, 310–311
- MNAR data, 312
  - ordered variables, 311–312
  - Rubin's rules, 312–313
- Multiple-matrix sampling, 239
- Multiplicative nonresponse models, 99
- Multistage sampling, tests statistics
  - under, 122–124
- Multivariate extensions, 139
- Multivariate normal distribution, 310
- National statistical institutes (NSIs), 89
- National Survey of Families and Households (NSFH), 110, 122
- Nearest neighbor imputation, 210
  - basic setup, 212–214
- Nearest neighbor imputation estimator, 210–213, 217–218, 222, 230
  - proofs of theorems, 228–234
  - replication variance estimation, 218–220
  - results, 214–217
  - simulation study, 221–224
- Nearest neighbor ratio imputation, 225
- New York City Police Department (NYPD), 277*n*1
- Non-negligible sampling fraction, 224–225
- Non-sampling errors, 260
- Non-smoothness, 211, 212
- Nonlinear least squares estimators, 129
- Nonlinear regression models, 139
- Nonmandatory survey, 239
- Nonnested competing models, 111–114
- Nonparametric kernel regression, 174
  - application, 196–197
  - bandwidth selection, 186–188
  - model-assisted nonparametric regression estimator, 178–186
  - Monte Carlo simulations, 188–196
  - proofs, 199–209

- Nonparametric maximum likelihood estimators (NPMLE), 147
- Nonparametric methods for estimating conditional mean functions, 174
- Nonprobability Internet panels, 8
- Nonrandom weight for stratum-cluster pair, 122
- Nonresponse, 100
  - bias, 240
  - linearization method, 101–102
- Normalized asymptotic distribution, 95
- Novel bootstrap procedure, 211
- Null hypothesis, 111–114, 117, 126
- OLS, 189
  - median MSE for, 191–193
- Online cash register (OCR), 49–50
- Organisation for Economic Co-operation and Development (OECD), 288
- Pairs cluster bootstrap, 65
- Panel data, model-selection tests with, 124–126
- Panel Survey of Income Dynamics, 110
- Payment card acceptance and usage
  - data description, 38–39
  - estimates of card acceptance, 42–43
  - estimating card acceptance, 41
  - Hungarian payments landscape, 38
  - key variables, 39–40
  - logistic regression models of card acceptance and usage, 51–57
  - methodology, 41
  - models of card acceptance and usage, 41
  - regression models of card payments, 43–46
  - review of literature, 49–51
  - “Period treated” dummy, 64
- Planned missing data design, 239, 242, 244
  - full-coverage and partial-coverage items, 248
  - original 2016 sample counts, 251
  - problem of dividing survey form, 248–249
  - sampling scheme, 247
  - survey for subsample, 244–245
  - survey form, 245–246, 250
  - survey stratification, counts and rates, 246
- Plug-in method for bandwidth, 186–187
- Poisson distribution, 132
- Poisson quasi-MLE, 111
- Polychotomous choice, 262
- Pooled estimation methods, 125
- Pooled nonlinear least squares, 125
- Pooled objective functions, 126
- Population minimization problem, 112
- Post-stratification
  - factors, 15
  - weights, 10
- Post-stratification weights, 9
- Power series estimators, 219, 231
- Predictor variables, 188
- Primacy effect, 21–22
- Primary sampling units (PSUs), 88, 290
- Probability
  - density function, 124
  - distributions, 141
  - models of illegal sales, 294
  - probability-based Internet surveys, 6
  - probability-based panels, 8

- Product kernel
  - for continuous variables, 179
  - function for vector of discrete variables, 179
- Propensity score matching, 223
- Pseudo-empirical likelihood objective function, 93
- Pseudo-true values, 112
- Pure design-based inference, 180
- Qin's treatment, 148
- Quantile estimation, 217, 220
- Quasi-MLEs (QMLE), 111, 130
- Quasi-true values, 112
- Racial discrimination, 260, 261, 262, 272, 273, 275
- Rademacher distribution, 76
- Raking ratio estimator, 88–89
  - classical raking estimator, 90–92
  - GREG, 93–94
  - maximum likelihood raking estimator, 92–93
- RAND American Life Panel, 8
- RAND-HRS indicator of retirement, 18
- Randomization inference (RI), 62, 67
  - problem of interval  $p$  values, 69–71
- Randomized planned missing items, 239
- Randomness, 88
- Re-randomizations, 69
  - re-randomized test statistics, 73
- “Realized” population, 138
- Recency effects, 21–22, 24, 27
- Regression analysis, 288–289
- Regression models of card payments, 43
  - acceptance, 43–45
  - usage, 45–46
- Replication method, 233
- Replication variance
  - estimation, 211–212, 218–220
  - estimator, 218
- Resampling method, 89
- Residual analysis, 313–314
- Respondents, 22
  - feedback, 242
- Response linearization method, 101–102
- Response quality improvement
  - business surveys, 239–240
  - challenge, 242–244
  - mean number of item responses, 256
  - planned missing data design, 239, 244–251
  - response counts, 252
  - response rates by stratum, 254
  - standard approach, 240–242
  - survey design, 251
  - 2016 survey stratification, counts and rates, 255
  - unit response rate, 253
  - unit responses, 238
- Rubin's rules, 312–313
- Rule-of-thumb method, 162–163n 14, 177
- Salary, 298
  - imputation model, 308
  - ratio of missing values, 298
- Sample mean squared error, 190–194
- Sample selection
  - correction, 262
  - procedure, 241
- Savings institutions (SAV), 241
- Scalar matching variable, 211, 228
- Scalar population parameter, 96
- Secondary sampling units (SSUs), 290
- Selection bias, 138, 147

- Selection correction, 271
  - model, 273
- Selective crime reporting
  - characterization of equilibrium, 269
  - coefficient estimation, 283
  - empirical strategy, 269–272
  - estimates of hit rates test on summons, 284
  - existence of equilibrium, 268
  - hit rates test, 262
  - misclassification, 260–261
  - Pedestrians, 267–268
  - police officers, 268
  - reason for Stop and Pedestrian characteristics, 285
  - results, 272–275
  - robustness checks for sample correction estimates, 286
  - Stop-and-Frisk program, 263–267
  - theory, 267
- Self-reported health with CPS limited to HH respondents, 30
- Self-reported mammography, 277
- Self-reported Transaction Price (STP)*, 292, 294–295, 301–302
- Sequential importance sampling (SIS), 8
- Shrinkage empirical-likelihood-based estimator, 94
- Sieves estimation, 231–233
- Sieves estimators, 219
- Simple random sampling (SRS), 100–101, 174, 210
  - median bandwidths for WLC and LC under, 194
  - median MSE for WLC, LC, WLS and OLS under, 191
- Simulation study, 156, 160–161, 221–222
  - aggregate shares for, 157
  - design, 156–157
  - inconsistency of LS estimator, 158, 159
- Size, 292
  - key variables, 39
  - size-based stratas, 44
- Small simulation study, 129–133
- Smoothed empirical likelihood estimator (SEL estimator), 152, 153, 154, 159, 162–164, 169
- Smoothing parameters, 186
- “Snob effect”, 288
- Social desirability effects, 22, 24–25
- Socioeconomic class, 297
- Socioeconomic variables in HRS, UAS and CPS, 11–14
- Split questionnaire survey design, 239
- Standard likelihood ratio testing principle, 110
- Standard ratio estimators, 241
- Standard stratified sampling (SS sampling), 111, 114–121
- Stata, 128
  - implementation under, 103
- Statistical inference, 138
- Stop-and-Frisk program, 260–262, 263–267
- Strata borders, 189
- Strategic money counterfeiting model, 290
- Stratification, 44–45, 122, 241
  - approach, 42–43
  - in terms of individual income, 121
- Stratified random samples, 41
- Stratified sampling, 114, 153
  - SS sampling, 114–121
  - testing under, 114–122
  - VP sampling, 121–122
- STRATVAR, 241, 244

- Superpopulation
  - model, 215
  - nonparametric regression model, 175
  - parameters, 181
- Survey
  - of banks, 240
  - business, 239–240
  - cross-validation, 187
  - face-to-face, 4–5
  - form, 245
  - Internet-based, 5
  - large-scale, 174
  - modes, 6
  - nonmandatory, 239
  - online, 4–5
  - statistics, 37–38
  - telephone surveys, 4–5
  - web, 5
  - See also* 2013 methods-of-payment survey (2013 MOP survey); Internet match high-quality traditional surveys
- Survey of Labour and Income Dynamics (SLID), 196
- Survey outcomes in HRS and UAS, 14
  - health economics literature, 20
  - health insurance coverage, 17
  - home ownership, 16
  - individual earnings, 18
  - mode effects, 21–26
  - post-stratification factors, 15
  - self-reported health, 19
  - whether retired, 17
- “Synthetic controls” method, 63
- TEÁOR variable, 52
- Temporal attributes of store, 52
- Temporal data, 56
- Time series problems, 110–111
- Tng\_credit\_year* variable, 100
- Tourist test, 49
- Trade policy, 288
- Traditional hit rates test, 269
- Traditional survey modes, 4
- Transaction size distribution, 45–46
- Transaction value, 45, 55
- Transaction-level data, 50–51
  - aggregated data, 39
- “Trimmed” SEL objective function, 155
- Trimming indicator, 155
- “True” variance, 100
- Two-step method, 270, 271
- 2013 methods-of-payment survey (2013 MOP survey), 89, 98
  - empirical application, 98–103
  - raking ratio estimator, 89–94
  - variance estimation, 94–98
- UF-250 form, 264
- Unconditional empirical probabilities, 170
- Unconditional summary statistics, 261
- Underrepresents high-income households, 14
- Understanding America Study (UAS), 5–6, 7
  - sampling and weighting procedures, 8–11
  - socioeconomic variables in, 11–14
  - survey outcomes in HRS and, 14–26
- “Unit nonresponse”, 297
- Unit response rate, 244–245, 254
- Unordered multinomial logit model, 270
- Unplanned missing items, 239
- Unweighted HRS race/ethnicity distribution, 13
- Unweighted life-satisfaction distribution, 21

- Unweighted objective function, 119
- Unweighted statistic, 131
- Unweighted tests, 133
- Urban Area to ZIP Code Tabulation  
Area Relationship File of US  
Census Bureau, 9
- Value categories, 51
- Value-added tax, 303
- Variable probability sampling (VP  
sampling), 111, 121–122, 138,  
140–142
- Variance estimation, 88, 94, 211
  - via bootstrap, 96–98
  - via linearization, 95–96
- Violent crimes, 264
- Vuong's approach, 110, 111
- Vuong's statistic, weighted version of,  
129
- Weighted local constant estimation  
(WLC estimation), 189, 191
  - estimator, 194, 196
  - median bandwidths, 194–196
  - median MSE for, 191–193
- Weighted/weighting, 6
  - life-satisfaction distribution, 21
  - objective function, 122
  - procedure, 88
  - tests, 133
- Wild bootstrap method, 211, 219
- Wild bootstrap randomization
  - inference (WBRI), 63, 71
  - alternative procedures, 72–74
  - cluster-robust inference, 63–67
  - empirical example, 79–82
  - RI, 62, 67–71
  - simulation experiments, 74–79
- Wild cluster bootstrap (WCB), 62,  
65–67
- WLS, 190, 192
  - median MSE for, 191–193
- Zero mean function, 202