

# Subject index

## A

abbreviations . . . . . 13  
accelerated failure-time model . . . . .  
    . . . . . 1123–1126  
adjusted  $R^2$  . . . . . 1469  
ado-files . . . . . 30–31, 721  
adoupdate command . . . . . 31  
AIC . . . *see* Akaike information criterion  
Akaike information criterion . . 630, 659,  
    1068, 1193, 1468, 1481  
AME . . . . . *see* marginal effects, average  
    marginal effect  
analysis-of-variance table . . . . . 22  
Anderson–Rubin test . . . . . 354–357  
anova command . . . . . 93  
append command . . . . . 66  
areg command . . . . . 273, 399  
Arellano–Bond estimator . . . . . 428–445  
arithmetic operators . . . . . 13  
ATE . . . . . *see* treatment effects  
    definition . . . . . 1271, 1291  
ATET . . . . . *see* treatment effects  
    definition . . . . . 1272, 1291  
augmented IPW . . . *see* treatment effects  
augmented mean group estimator . . 454  
average  
    marginal effect . . . . . *see* marginal  
        effects  
    structural function . . . . . 1343  
    structural mean . . . . . 1343  
    structural probability . . . . . 1343  
    treatment effect . . . . . 196, *see also*  
        treatment effects, 1272

## B

backfitting algorithm . . . . . 1459  
bagging . . . . . 1499

balanced data, creating . . . . . 1001  
bayes prefix . . . . . 1528  
bayesgraph command . . . . . 1542  
Bayesian information criterion . . . . .  
    . . . 630, 659, 1068, 1193, 1468,  
    1481, 1489, 1505  
Bayesian methods . . . . . 1527–1609  
    Bayes factors . . . . . 1567–1568  
    Bayes rule . . . . . 1533  
    compared with classical approach . .  
        . . . . . 1536  
    credible region . . . . . 1540  
    data augmentation . . . . . 1589–1607  
    diagnostics . . . . . 1542  
    different starting values . . . . . 1546  
    hierarchical prior application . . . . .  
        . . . . . 1562–1566  
    inference . . . . . 1536, 1540  
    linear regression . . . . . 1528–1532,  
        1549–1552  
    marginal effects . . . . . 1574–1576  
    marginal likelihood . . . . . 1533, 1567  
    MCMC . . . *see* Markov chain Monte  
        Carlo methods  
    model selection . . . . . 1567–1569  
    normal i.i.d. example . . . 1538–1549  
    pitfalls . . . . . 1556–1560  
    posterior . . . . . *see* posterior  
        distribution  
    posterior odds ratio . . . . . 1568–1569  
    prediction . . . . . 1569–1572  
    prior . . . . . *see* prior distribution  
    probit application . . . . . 1572–1576,  
        1579–1594  
    random effects . . . . . 1562–1566  
bayesmh command . . . . . 1537, 1581  
    with user-provided evaluator . . 1582

- bayepredict** command.....1570  
**bayesstats**  
   **ess** command.....1554  
   **grubin** command.....1545  
   **ic** command.....1568  
   **model** command.....1569  
   **summary** command.....1541  
**bayestest interval** command...1541  
 before–after comparison.....198  
 BIC..... see Bayesian information criterion  
 binary outcome models.....459–474, 857–899  
   alternative-specific random parameters logit.....884  
   Bayesian example.....1572–1576, 1579–1594  
   Bernoulli trial.....858  
   binomial regression.....897  
   classification test.....872  
   clustered data.....880–881  
   coefficient interpretation..470, 865  
   complementary log–log model..858  
   confusion matrix.....872  
   dynamic logit model.....1162  
   endogenous regressors....887–895  
   FMM logit application..1205–1211  
   generalized linear models.....868  
   generalized logit model.....883  
   goodness-of-fit measures..869–873  
   goodness-of-fit test.....870  
   grouped-data analysis....895–898  
   heteroskedastic probit model..882  
   hypothesis tests.....867–868  
   introduction.....459–474  
   IV probit estimator.....893–895  
   IV probit structural.....888–893  
   lasso.....1492–1493  
   latent-variable model....859–860, 882, 888, 1580  
   linear probability model.....471, 859, 862  
   logit model.....461, 472–474, 858, 864–880  
   binary outcome models, *continued*  
     machine learning estimator.....1517–1518  
     marginal effects.....466–472, 877–880  
     maximum likelihood estimation...860  
     mixed logit.....884  
     model comparison...865–866, 868  
     nonlinear IV estimator.....895  
     nonparametric logit.....885  
     panel-data estimators..1148–1163  
     predicted probabilities...873–876  
     predicted versus actual frequencies.....870  
     predicted versus actual outcome..872  
     probit model.....461–466, 858, 865–866, 868, 873–874  
     pseudo- $R^2$ .....870  
     robust standard errors.....861, 880–881  
     ROC curve.....875  
     selection model.....977  
     shrinkage estimation...1492–1493  
     specification analysis.....883  
     weak instruments.....893  
   binomial regression..... see binary outcome models  
   **biprobit** command.....943  
   BLP market demand model.....937  
   **boost** command.....1506  
   boosting.....1500, 1506  
   bootstrap methods..101–102, 117–120, 537–577  
     asymptotic refinement...538, 545, 555–568  
     asymptotically pivotal statistic...556  
     BCa confidence interval...545, 556  
     bias estimation.....547  
     bias-corrected confidence interval..545  
     bootstrap pairs.....101, 539–540  
     bootstrap pairs resampling....571

- bootstrap methods, *continued*
- case bootstrap ..... 540
  - cluster pairs ..... 119–120, 540, 543–544, 548
  - cluster-specific indicators ..... 544
  - confidence intervals ..... 544–546, 555–568
  - for 2SLS ..... 363
  - for Hausman test ..... 553–554
  - for nonparametric regression .. 1443
  - for quantile regression ..... 687
  - for two-step estimator .... 551–552
  - jackknife ..... 575–576
  - misuse of bootstrap ..... 539
  - nonparametric bootstrap ..... 540
  - normal-based confidence interval ..  
..... 544
  - number of bootstraps ..... 542
  - parametric bootstrap resampling ..  
..... 571
  - percentile-*t* confidence interval ....  
..... 545, 559
  - percentile-*t* method ..... 555–568
  - residual bootstrap resampling .. 572
  - score-based wild cluster bootstrap  
..... 565–567
  - standard error estimation .....  
. 101–102, 117–120, 537, 541
  - subsampling ..... 575
  - variance matrix ..... 117–120
  - variance of standard error .... 549
  - wild bootstrap ..... 560–568, 573
  - wild bootstrap for IV ..... 363, 567–568
  - wild cluster bootstrap .... 561–568, 1364
  - wild gradient bootstrap ..... 1185
- bootstrap prefix ..... 547
- boottest command ..... 100, 265, 378, 562
- box-and-whisker plot ..... 71
- boxcox command ..... 135
- brain command ..... 1505
- bsample command ..... 569
- bspline command ..... 675
- bsqreg command ..... 687
- by prefix ..... 8, 54
- bysort prefix ..... 54
- C**
- capture command ..... 716
- case sensitivity ..... 13
- cd command ..... 15
- cdeco command ..... 1398
- censored data ..... see tobit models, see count-data models, see duration models
- centile command ..... 88, 688
- cfunction option ..... 1349
- chi-squared goodness-of-fit test .... see specification tests
- chi2gof command ..... 530
- Cholesky decomposition ..... 224
- churdle command ..... 1046
- cii command ..... 515
- ciwidth onemean command ..... 528
- clad command ..... 960
- classification ..... 1500–1501
- clear command ..... 26, 37
- clogit command ..... 922
- cloglog command ..... 861
- clusteff command ..... 264
- cluster analysis ..... 256, 1501
- cluster command ..... 256
- cluster kmeans command ..... 1501
- clustered data ..... 256–286, 632–640
- binary outcome models .. 880–881
  - correlated random-effects model ..  
..... 275, 639
  - count-data models ..... 1089–1090
  - few clusters .... 264–265, 561–567
  - FGLS example ..... 265–269
  - fixed-effects estimator .... 270–271
  - fixed-effects model ..... 269–277
  - hierarchical models ..... 284–285
  - influential clusters ..... 264
  - lasso ..... 1482, 1489, 1505
  - linear mixed models ..... 277–286
  - mixed nonlinear models .....  
..... 1224–1231

- clustered data, *continued*
  - multinomial outcome models.. 942
  - nonlinear example..... 632–640
  - nonlinear FE model example.....
    - ..... 637–640
  - nonlinear RE model example.....
    - ..... 635–637
  - OLS example ..... 256–265
  - population-averaged model
    - example..... 266, 633–635
  - random intercept model..... 280
  - random-effects Bayesian .....
    - ..... 1562–1566
  - random-effects estimator.. 267–269
  - random-effects model .... 267–269,
    - 280
  - survey methods..... 295–301
  - tobit models ..... 961
  - two-way fixed effects..... 276–277,
    - 406
  - two-way random effects ..... 285
- cluster-robust inference
  - cluster pairs bootstrap... 119–120,
    - 540, 543–544, 548
  - design-based ..... 1289
  - dyad-robust ..... 263
  - few clusters .... 249, 264–265, 487,
    - 563–567
  - for DID ..... 1364
  - for FGLS..... 247–249
  - for mixed estimator..... 281–284
  - for OLS ..... 99, 116, 258
  - for panel data ..... 377–378, 1144
  - for RCT..... 1280, 1289
  - limitations ..... 249
  - power ..... 527–528
  - two-way robust..... 261–263
  - variance inflation..... 100, 260
  - wild cluster bootstrap.... 561–567,
    - 1364
- cluster-robust standard errors ..... see
  - variance-covariance matrix
- cmchoiceset command..... 916
- cmclogit command..... 916
- cmdlog command..... 17
- cmmixlogit command..... 935
- cmmprobit command..... 931
- cmrologit command..... 922
- cmroprobit command..... 934
- cmsample command ..... 916
- cmset command..... 915
- cmsummarize command..... 916
- cmtab command..... 916
- coefficient interpretation ..... see also
  - marginal effects
    - in binary outcome models.... 864
    - in multinomial outcome models...
      - ..... 910, 919
    - in Poisson model..... 1033
    - in single-index models.... 470, 614
    - marginal effects ..... 467
- coeflegend option ..... 487
- coefpath command ..... 1485
- collapse command..... 62, 896
- collect command..... 111
- comma-separated values data ... 34, 40
- command prefix..... 8
- comments in Stata commands ..... 18
- common correlated estimator ..... 454
- community-contributed commands.. 30
- complier..... 330, 1357
- compress command ..... 35
- conditional independence assumption..
  - ..... 1290
- conditional mean assumption..... 1291
- condivreg command ..... 358
- confidence intervals..... 485
  - asymmetric ..... 496
  - asymptotic refinement ..... 545,
    - 555–568
  - bootstrap confidence intervals ....
    - ..... 544–546
  - from invert test statistic..... 356,
    - 497–498
  - percentile-*t* confidence interval....
    - ..... 559
  - Wald confidence intervals.....
    - ..... 494–496
  - with desired precision..... 528–529
- constraint command..... 106, 293

- constraints()** option..... 106  
**continue** command ..... 29  
**contrast** command ..... 1328  
control function estimator .... 325, 891,  
1080–1084, 1356  
    standard errors..... 594–596  
copula models..... 982–986  
**correlate** command ..... 103, 392  
correlation coefficient ..... 103  
    autocorrelations for panel data ...  
        ..... 391  
    intraclass ..... 393  
    intracluster ..... 259  
**count** command ..... 48  
count-data models..... 582–585,  
1021–1098  
    censored Poisson..... 1043  
    clustered data..... 632–640, 1032,  
    1089–1090  
    comparison of four distributions ..  
        ..... 1051  
    endogenous regressors .. 1079–1089  
    equidispersion ..... 1022  
    excess zeros..... 1044  
    finite mixture models .. 1050–1069,  
    1217–1220  
    fitted probabilities ..... 1036, 1056  
    frequency distribution..... 1027  
    generalized NB regression .... 1041  
    GMM estimators..... 1084–1088  
    goodness of fit ..... 1036  
    hurdle model..... 1044–1050  
    interpretation of coefficients.. 1033  
    lasso ..... 1492–1493  
    log-linear model alternative .. 1026  
    machine learning estimator .....  
        ..... 1516–1517  
    main features of count data .. 1022  
    marginal effects.. 1033, 1036, 1048,  
    1071  
    model selection ..... 1067, 1076  
    NB1 model ..... 1024  
    NB2 model ..... 1024, 1034–1041  
    negative binomial distribution ....  
        ..... 1024  
count-data models, *continued*  
    negative binomial model.....  
        ..... 1034–1041  
    nonlinear least squares.. 1042–1043  
    overdispersion ..... 1023  
    overfitting ..... 1068  
    panel-data estimators .. 1172–1184  
    point-mass distribution..... 1074,  
    1220  
    Poisson distribution..... 1022  
    Poisson for continuous nonnegative  
        dependent variable ..... 1026  
    Poisson introduction..... 582–585  
    Poisson mixed effects ... 1226–1231  
    Poisson model..... 1028–1041  
    quantile count regression .....  
        ..... 1090–1096  
    robust standard errors.. 1031–1032  
    shrinkage estimation ... 1492–1493  
    test of endogeneity..... 1083  
    test of overdispersion ..... 1032  
    treatment effects ..... 1353–1356  
    truncated Poisson and NB2... 1044  
    two-part model..... 1044–1050  
    unobserved heterogeneity .... 1023  
    weak instruments ..... 1088  
    zero-inflated models... 1069–1078  
    zero-truncated models ..... 1043  
**counterfactual** command..... 1398  
**countfit** command ..... 1037  
Cox proportional hazards model.... *see*  
    duration models  
**cpoisson** command ..... 1043  
**cquad** commands..... 1163  
**creturn** command..... 19  
critical values ..... 480–485  
    chi-squared versus  $F$  ..... 481  
    computation..... 480–485  
    standard normal versus  $t$  ..... 481,  
    484  
    used in Stata commands ..... 484  
cross-fit partialing-out estimator .....  
    ..... 1514–1515

- cross-validation . . . . . 1471–1476
  - $K$ -fold . . . . . 1472–1475
  - leave-one-out . . . . . 1437, 1475
  - single split . . . . . 1471
  - test sample . . . . . 1471
  - training sample . . . . . 1471
- crossfold** command . . . . . 1474
- cubic splines . . . . . 672
- cumulative hazard function . . . . . 1108
  - Nelson–Aalen estimate . . . . . 1108
- cvplot** command . . . . . 1485
  
- D**
- data augmentation . . . . . 1589–1595, 1597–1607
  - for multiple imputation . . . . . 1597–1607
- data frames . . . . . 61
- data management . . . . . 33–67
  - appending datasets . . . . . 66
  - collapsing and expanding data . . 62
  - common pitfalls . . . . . 42
  - data frames . . . . . 61
  - data types . . . . . 33–36
  - demeaned variables . . . . . 56
  - dictionary file . . . . . 42
  - exporting data . . . . . 57
  - importing data . . . . . 39
  - imputing missing values . . . . . 51
  - indicator variables . . . . . 54–56
  - inputting data . . . . . 36–43
  - interacted variables . . . . . 9–13, 56
  - labeling variables . . . . . 46
  - long form . . . . . 63, 414–418
  - manipulating datasets . . . . . 60–67
  - merging datasets . . . . . 63–66
  - missing values . . . . . 49
  - naming variables . . . . . 46
  - ordering data . . . . . 60
  - outputting data . . . . . 43, 57
  - preserving and restoring data . . 60
  - PSID example . . . . . 43
  - saving data . . . . . 57
  - selecting sample . . . . . 58
  - special data formats . . . . . 39
- data management, *continued*
  - transforming data . . . . . 51–57
  - wide form . . . . . 63, 414–418
- data summary examples . . . . . 87
  - for cross-sectional data . . . . . 85–94
  - for duration data . . . . . 1100–1109
  - for multinomial data . . . . . 905–908
  - for panel data . . . . . 379–387
- data transformations . . . . 51–57, 94–96, 102
  - regression in logs . . . . . 102
  - retransformation . . . . . 102
- data-generating process . . . . . 207
- dataset description
  - Acemoglu–Johnson–Robinson data
    - for lasso example . . . . . 1519
  - American Community Survey earnings Bayesian example . . 1528
  - bootstrap example . . . . . 541
  - cigarette consumption panel in
    - wide form . . . . . 415
  - cigarette sales long panel example . . . . . 446
  - cigarette sales synthetic control example . . . . . 1365
  - CPS few clusters example . . . . . 563
  - fishing-mode multinomial data . . . . . 905
  - homicide spatial data example . . . . . 1407
  - HRS private health insurance
    - binary data . . . . . 863
  - Medical Expenditure Panel Survey
    - ambulatory for tobit . . . . . 953
    - doctor-visits count data . . . . 1026
    - doctor-visits young . . . . . 459
    - drug expenditures for IV . . . . 317
    - emergency room visits . . . . . 1069
    - for count QR example . . . . . 1092
    - for GSEM example . . . . . 1259
    - for SUR example . . . . . 288
    - medical expenditures . . . . . 85
  - Medicare multivalued treatment-effects example . . . . . 1324

- dataset description, *continued*
- Medicine in Australia:
    - Balancing Employment and Life doctor earnings for FMM example . . . . . 650
    - Balancing Employment and Life earnings for regression decomposition . . . . . 188
    - Balancing Employment and Life panel attrition example . . . . . 996
  - NSW Panel Survey of Income Dynamics earnings diff-in-diff . . . . . 196
  - OHIE treatment-effects example . . . . . 1307
  - Panel Study of Income Dynamics short panel wage data . . . . . 379
  - PSID earnings data management example . . . . . 43
  - Rand HIE nonlinear panel example . . . . . 1145
  - Senate elections for regression discontinuity design example . . . . . 1376
  - the second National Health and Nutrition Examination Survey survey design data . . . . . 296
  - unemployment duration example . . . . . 1100
  - Vietnam clustered village data . . . . . 257
- date()** function . . . . . 60
- decode** command . . . . . 53
- defier** . . . . . 1357
- delimit** command . . . . . 18
- delta** method . . . . . 491–497
- describe** command . . . . . 45, 86
- descriptive statistics . . . . . 87–89
- design-based inference . . . . . 1289
- destring** command . . . . . 36
- deviance residual . . . . . 631, 1492
- deviance statistic . . . . . 631
- didregress** command . . . . . 202, 1363
- difference in differences . . . . . 195–204, 1363–1364
- average treatment effect on the treated . . . . . 201
  - before–after comparison . . . . . 198
  - OLS regression computation . . . . . 200
  - parallel trends assumption . . . . . 196, 1364
  - treatment–control comparison . . . . . 199
- difference in means test . . . . . 122
- dimension reduction principal components . . . . . 1494–1496
- discrete-time hazards model . . . . . 1132–1137
- discrim** command . . . . . 1501
- discriminant analysis . . . . . 1500
- display** command . . . . . 19
- do** command . . . . . 15
- do-files . . . . . 15–19
- template do-file . . . . . 30
- double selection estimator . . . . . 1515–1516
- doubly robust ATE estimator . . . . . 1295
- doubly robust methods . . see treatment effects
- drawnorm** command . . . . . 223
- drop** command . . . . . 59
- dslogit** command . . . . . 1517
- dspoisson** command . . . . . 1516
- dsregress** command . . . . . 1515
- duration data . . . . see duration models
- duration models . . . . . 1099–1138
- accelerated failure-time models . . . . . 1123–1126
  - censoring . . . . . 1099
  - clustered data . . . . . 1136
  - competing risks . . . . . 1116–1118
  - complete spells . . . . . 1099
  - Cox proportional hazards model . . . . . 1110–1116, 1133
  - cumulative hazard function . . . . . 1108
  - data summary . . . . . 1100–1104
  - diagnostics . . . . . 1114–1116
  - discrete-time hazards model . . . . . 1132–1137

- duration models, *continued*
- exponential model . . . . 1099, 1122
  - frailty in models . . . . . 1126–1129
  - fully parametric models . . . . .  
     . . . . . 1118–1128
  - generalized gamma model . . . . 1122
  - Gompertz model . . . . . 1122
  - hazard function . . . . . 1107
  - Kaplan–Meier estimate . . . . . 1106
  - log–log survivor curve . . . . . 1114
  - log-integrated hazard curve . . 1114
  - loglogistic model . . . . . 1124
  - lognormal model . . . . . 1124
  - marginal effects . . 1122, 1128, 1135
  - multiple-records data . . 1129–1136
  - Nelson–Aalen estimate . . . . . 1108
  - prediction . . . . . 1112, 1121
  - proportional hazards model . . 1110,  
     1122–1124
  - Schoenfeld residual . . . . . 1115
  - survivor function . . . . . 1105
  - time-varying regressors . . . . . 1135
  - Weibull FMM application . . . . .  
     . . . . . 1220–1223
  - Weibull model . . . . . 1119
  - dyad–robust standard errors . . . . . 263
- E**
- e-class commands . . . . . 21, 719
  - egen** command . . . . . 52
  - eintreg** command . . . . 990, 1019, 1304,  
     1342
  - elastic net . . . . . 1480–1481, 1489–1491
  - elasticities
    - in linear regression model . . . . 182
    - in nonlinear models . . . . . 625–628
  - elasticnet** command . . . . . 1482, 1489
  - encode** command . . . . . 53
  - endogeneity test . . . . . *see* specification tests
  - endogenous regressors . . . . . *see also* instrumental variables
    - definition . . . . . 311
    - dynamic panel model . . . . 428–445
    - dynamic panel systems . . . . . 445
  - endogenous regressors, *continued*
    - endogenous-switching regression
      - model . . . . . 990
    - ERM commands . . . . . 1261–1266,  
     1341–1348
    - ET commands . . . . . 1348–1356
    - fixed-effects model . . . . . 375, 397
    - Hausman test . . . . . 531, 553–554
    - in binary outcome models . . . . .  
     . . . . . 887–895
    - in count-data models . . 1079–1089
    - in tobit model . . . . . 989–990
    - instrumental-variables regression . .  
     . . . . . 313
    - nonlinear instrumental variables . .  
     . . . . . 851
    - panel data . . . . . 1187–1188
    - panel IV . . . . . 421–445
    - quantile treatment effects . . . . .  
     . . . . . 1388–1394
    - simulation example . . . . . 241
    - structural equation model . . . . .  
     . . . . . 1256–1259
    - treatment effects . . . . . 1337–1403
  - endogenous-switching regression model  
     . . . . . 990
  - eoprobit** command . . 1261, 1304, 1342
  - eprobit** command . . . . 891, 1261, 1304,  
     1342
  - eregress** command . . . . 327, 364, 976,  
     990, 1019, 1261, 1304, 1341
  - ereturn**
    - command . . . . . 715
    - display** command . . . . . 146
    - list** command . . . . . 21
  - ERM commands . . . . . 1261–1266
    - for endogenous treatment . . . . .  
     . . . . . 1341–1348
    - for exogenous treatment . . . . .  
     . . . . . 1304–1305
    - for panel data . . . . . 1187
  - error messages . . . . . 14, 724
  - estat**
    - abond** command . . . . . 435
    - alternatives** command . . . . . 920



- estat, continued**
- bootstrap command..... 546
  - classification command... 872
  - commands ..... 105, 503
  - correlation command..... 934
  - covariance command ..... 934
  - endogenous command ..... 323
  - firststage command ..... 344
  - gof command..... 530, 871
  - hettest command .. 136, 254, 973
  - ic command..... 631
  - impact command ..... 1423
  - imtest command ..... 138, 530
  - lcmean command ..... 655, 1197
  - lcprob command ..... 658, 1197
  - mfxc command..... 920
  - moran command..... 1414
  - overid command ..... 326, 531
  - ovtest command ..... 134
  - phtest command ..... 1115
  - sargan command ..... 436
  - teffects command... 1305, 1343
  - vce command..... 487
- estimates**
- selected command..... 107
  - stats command ..... 107
  - store command ..... 107
  - table command ..... 107
- estimation commands**
- linear panel summary ..... 379
  - multinomial summary ..... 904
  - nonlinear cross-sectional summary  
..... 579
  - nonlinear panel summary .... 1144
- estout** command..... 109
- eststo** command..... 109
- esttab** command..... 109
- ET** commands for endogenous treatment  
..... 1348–1356
- etable** command..... 109
- eteffects** command..... 1356
- etpoisson** command..... 1353
- etregress** command ..... 328, 1348
- event study models**..... 1364
- expand** command..... 62
- exponential model**..... 1122
- export** command ..... 43
- extended regression models** ... see ERM  
commands
- extreat()** option ..... 1304
- F**
- factor variables**.. 9–13, 55–56, 111–116,  
622–625
- factor-variable notation**..... 114
- marginal effects ..... 179
- failure data** ..... see duration models
- feasible generalized least squares** ... see  
generalized least squares
- femlogit** command ..... 942
- finite mixture models**..... 643–665,  
1192–1224
- computational method .. 650, 1194
  - count-data regression... 1050–1069
  - definition ..... 646, 1192
  - gamma regression ..... 1195–1205
  - linear regression example.. 647–648
  - log-linear regression example .....  
..... 650–665
  - logit regression ..... 1205–1211
  - marginal effects ..... 657, 1198
  - mixture of normals .. 644–645, 649
  - model selection ..... 659–661
  - modeling considerations ..... 648
  - multimodality..... 652
  - multinomial logit ..... 1211–1214
  - multiple response example.....  
..... 1259–1261
  - point-mass distribution..... 1074,  
1220
  - Poisson regression ..... 1217–1220
  - predict component means ..... 654
  - predicted class posterior probabili-  
ties ..... 658,  
1201
  - predicted class probabilities .. 658,  
1197
  - predicted component means.. 1199
  - test of coefficient equality ..... 661
  - tobit regression..... 1215–1217

- finite mixture models, *continued*  
     varying mixture probabilities . . . 663, 1193  
     Weibull regression . . . . . 1220–1223
- fixed effects . . . . see clustered data, see panel data
- floating-point data . . . . . 34
- fmm** prefix . . . . . 649, 1053, 1194
- foreach** command . . . . . 27, 621, 969
- format** command . . . . . 48
- formats to display data . . . . . 36, 48
- forvalues** command . . . . . 28, 219, 392
- fp** prefix . . . . . 667
- fracreg logit** command . . . . . 897
- fracreg probit** command . . . . . 897
- fractional response data regression . . . . . 862, 897
- frame** command . . . . . 61
- frmtable** command . . . . . 109
- fsum** command . . . . . 89
- G**
- gam** command . . . . . 674, 1459
- gamma regression model . . . . . 1195–1196  
     FMM application . . . . . 1195–1205
- Gauss–Hermite quadrature . . . . . 227, 635
- generalized additive model . . . . . 681, 1458–1461
- generalized estimating equations . . . . . 634–635, 1141, 1150
- generalized gamma model . . . . . 1122
- generalized least squares . . . . . 245–303  
     cluster–robust variance matrix . . . . . 248  
     definition . . . . . 247  
     efficient estimation . . . . . 247  
     for heteroskedastic errors . . . . . 246, 254  
     for SUR model . . . . . 287–293  
     leading examples . . . . . 249  
     robust standard errors . . . . . 247–249  
     robust variance matrix . . . . . 248  
     seemingly unrelated regressions . . . . . 286–295
- generalized least squares, *continued*  
     WLS estimator . . . . . 247, 256  
     working matrix . . . . . 247
- generalized linear models . . . . . 589–591  
     definition . . . . . 589  
     link function . . . . . 590  
     logit . . . . . 868  
     mixed effects . . . . . 1225–1231  
     Poisson example . . . . . 590  
     probit . . . . . 868
- generalized method of moments . . . . see also instrumental variables  
     count-data example . . . . . 1084–1089  
     definition . . . . . 591–596  
     for two-step estimators . . . . . 594–596  
     nonlinear example . . . . . 593, 851–854  
     panel Arellano–Bond estimator . . . . . 430
- generalized SEM . . . . . 1251–1261
- generate** command . . . . . 52
- geospatial data . . . . . 1407–1411
- gettoken** command . . . . . 720
- Gibbs sampler simulation example . . . . . 225
- Gibbs sampling algorithm . . . . see Markov chain Monte Carlo methods
- GLM . . . . see generalized linear models
- glm** command . . . . . 590, 868
- global** command . . . . . 24
- global macros . . . . . 24
- gmm** command . . . . . 592, 1087
- gnbreg** command . . . . . 1041
- gologit2** command . . . . . 942
- Gompertz model . . . . . 1122
- goodness-of-fit measures . . . . . 629–630
- gradient methods . . . . . see iterative methods
- graph**  
     **box** command . . . . . 71  
     **combine** command . . . . . 69, 93, 253  
     **export** command . . . . . 69  
     **matrix** command . . . . . 82  
     **save** command . . . . . 69  
     **twoway** command . . . . . 67

- graphs ..... 67–82
- Bayesian diagnostics ..... 1542
  - binary outcome plot ..... 874
  - box-and-whisker plot ..... 71
  - combining graphs ..... 69, 253
  - exporting ..... 69
  - graph formats ..... 70
  - graphing commands ..... 67
  - histogram ..... 72
  - kernel density plot .. 73–76, 93, 482
  - line plot ..... 76
  - multiple scatterplots ..... 82
  - nonparametric regression ... 77–82, 678, 1447
  - of known density function .... 482
  - of predictive margins ..... 169
  - outlier detection ..... 125
  - panel-data plots ..... 386
  - quantile plot ..... 689
  - residual plots ..... 124, 253
  - scatterplot ..... 76
  - Stata Graph Editor ..... 70
- grmap** command ..... 1409
- grouped data binary outcome models ..  
..... 895–898
- grqreg** command ..... 698
- gsem** command ..... 1253
- hurdle model ..... 1049
- gvselect** command ..... 1476
- H**
- halton()** Mata function ..... 231
- Halton sequences ..... 230
- Hammersley sequences ..... 230
- hausman** command ..... 323, 409, 531
- Hausman test ... see specification tests
- Hausman–Taylor estimator ... 425–427
- hazard function ..... 1107
- heckman** command ..... 976
- eregress** command equivalent ...  
..... 976, 1263
- heckmancopula** command ..... 983
- heckprobit** command ..... 977
- heckpoisson** command ..... 977
- heckprobit** command ..... 977
- help**
- command ..... 3
  - contents** command ..... 4
  - mata** command ..... 729
- heteroskedastic errors, feasible GLS ...  
..... 246
- heteroskedastic–robust ..... see  
variance–covariance matrix
- heteroskedasticity test ... 136, 253, 695
- hetoprobit** command ..... 941
- hetprobit** command ..... 882
- hierarchical models ..... 284–285,  
1562–1566
- histogram** command ..... 72
- hnblogit** command ..... 1047
- hplogit** command ..... 1046
- hurdle model ... see count-data models
- GSEM estimation ..... 1049
- hypothesis tests ..... 479–535
- asymptotic refinement ..... 538,  
555–568
  - auxiliary regression computation ..  
..... 503–505
  - Benjamini–Hochberg procedure ...  
..... 511
  - Bonferroni correction ..... 506
  - bootstrap methods ..... 537–577
  - chi-squared versus  $F$  ..... 481, 484
  - cross-equation restrictions ..... 291
  - delta method ..... 491–497
  - effect size ..... 519
  - false discovery proportion ..... 511
  - false discovery rate ..... 511
  - familywise error rate ..... 506–508
  - few clusters ..... 264–265, 487
  - Hochberg’s step-up procedure .. 507
  - Holm’s step-down procedure .. 506
  - invariance under transformation ..  
..... 498
  - invert to give confidence interval ..  
..... 356, 497–498
  - Lagrange multiplier tests .. 502–505
  - likelihood-ratio tests ..... 498–502
  - linear hypotheses ..... 485–486
  - linear regression model ..... 106

- hypothesis tests, *continued*
- minimum effect size . . . . . 520
  - minimum sample size . . . . . 520
  - multiple outcomes . . . . . 510
  - multiple testing . . . . . 505–512
  - nonlinear hypotheses . . . . . 491–497
  - on difference in means . . . . . 122
  - on population mean . . . . . 93, 121
  - one-sided tests . . . . . 490
  - percentile-*t* method . . . . . 555–568
  - permutation tests . . . 532–534, 1274, 1369
  - placebo test . . . . . 515
  - power . . . . . 238, 516–528
  - power in more than one direction . . . . . 138–140
  - power with clustering . . . . . 527–528
  - pretest bias . . . . . 493
  - randomization tests . . . . . 534, 562
  - score tests . . . . . 502–505
  - Šidák correction . . . . . 507
  - size . . . . . 237, 513–515
  - standard normal versus *t* . . . . . 481, 484
  - subgroup analysis . . . . . 508–510
  - tests at the boundary . . . . . 502
  - Wald test examples . . . 488–490, 492
  - Wald tests . . . . . 485–498
  - with weak instruments . . . . . 353–363
- I**
- `idcluster()` option . . . . . 544
  - identification and nonconvergence . . . 826
  - `if` qualifier . . . . . 6
  - ignorability . . . . . see unconfoundedness
  - `import` command . . . . . 39
  - imputation
    - inverse-probability weighting . . . . . 1004–1011
    - multiple imputation . . . . . 1595–1607
    - regression-based . . . . . 1596
    - sample selection . . . . . 1011–1018
    - summary of methods . . . . . 995
  - `in` qualifier . . . . . 6
  - indicator variables . . . . . 54–56
  - `infile` command . . . . . 41
  - `infix` command . . . . . 41
  - influential observations . . . 126, 264, 362
  - information criteria . . . . . 630–631, 659, 1068, 1193, 1468, 1481
  - information matrix test . . . . . see specification tests
  - `input` command . . . . . 38
  - instrumental variables . . . . . 305–371
    - 2SLS bias . . . . . 338
    - 2SLS estimator . . . 314, 317, 330–363, 430
    - 2SLS example . . . . . 319
    - 3SLS estimator . . . . . 367–368
    - 3SLS example . . . . . 367
    - alternative estimators . . . . . 363–367
    - Anderson–Rubin test . . . . . 354–358, 568
    - basic theory . . . . . 311–316
    - binary endogenous regressor . . . . . 328–330
    - bootstrap . . . . . 363
    - concentration parameter . . . . . 338
    - conditional LR tests . . . . . 358–359
    - control function estimator . . . . . 325, 594–596, 891, 1080–1084, 1356
    - Cragg–Donald minimum eigenvalue statistic . . . . . 341
    - effective *F* statistic . . . . . 343
    - example . . . . . 316–330
    - finite-sample properties . . . 332–339, 362–363
    - first-stage *F* statistic . . . . . 337–338, 345
    - first-stage equation . . . . . 312
    - GMM estimator . . . . . 314
    - Hansen test . . . . . 325
    - in binary outcome models . . . . . 893–895
    - in count-data models . . . 1079–1089
    - in spatial regression models . . . 1419, 1427–1428
    - inconsistency of OLS . . . . . 311
    - IV estimator . . . . . 313
    - jackknife IV estimator . . . . . 365

- instrumental variables, *continued*
- just-identified model . . . . 313, 319, 344
  - $k$ -class estimator . . . . . 364
  - lasso . . . . . 1518–1521
  - LATE . . . . . see LATE
  - minimum distance-based tests . . . . . 360–363
  - nonlinear IV . . . . 593, 851–854, 895
  - optimal GMM estimator . . 314, 317
  - optimal GMM example . . . . . 321
  - overidentified model . . . . 313, 321, 346
  - panel IV estimator . . . . . 421–445
  - panel systems IV estimator . . . 445
  - partial  $R^2$  . . . . . 340, 345
  - quantile regression . . . . 1388–1394
  - reduced form . . . . . 307
  - relative asymptotic bias test . . 343, 346
  - relative bias test . . . . . 341
  - robust standard errors . . . . . 315
  - Sargan test . . . . . 325, 436
  - sensitivity to instrument choice . . . . . 351
  - simple example . . . . . 312
  - simultaneous equations model . . . . . 306–310
  - structural equation . . . . . 312
  - structural model approach . . . . . 306–310, 888–893, 1079–1089
  - test of endogeneity . . . . . 322–325
  - test of overidentifying restrictions . . . . . 325–327
  - test of underidentification . . . . 341
  - test size distortion . . . . . 339
  - two-sample 2SLS estimator . . . . 366
  - underidentified model . . . . . 313
  - valid instruments . . . . . 315
  - Wald size-distortion test . . . . . 342
- instrumental variables, *continued*
- weak instruments . . . 315, 330–363, 893, 1088
    - asymptotics . . . . . 357–359
    - diagnostics . . . . . 339–340
    - inference . . . . . 353–363
    - simulation . . . . . 333–337
    - tests . . . . . 341–351
  - wild bootstrap for IV . . . . . 567–568
- integral computation . . . . . 227–232
- intention-to-treat effect . . . . 1307, 1360
- interacted regressors . . . . . 111–116, 623–625
- interacted variables . . . . . 9–13, 56
- interactive use of Mata . . . . . 728
- interactive use of Stata . . . . . 1–2
- interval regression . . . . see tobit models
- intraclass correlation . . . . . 393
- intracluster correlation coefficient . . 259
- intreg** command . . . . . 960
- inverse-probability weighting . . . . . 142, 993–994, 1004–1011, 1293–1294
- invnormal()** function . . . . . 210
- ipolate** command . . . . . 51
- IPW regression adjustment . . . . . see treatment effects
- iqreg** command . . . . . 687
- iterative methods . . . . . 819–856
  - checking analytical derivatives . . . . . 839
  - checking data . . . . . 839
  - checking parameter estimates . . 841
  - checking program . . . . . 837–843
  - checking standard errors . . . . . 842
  - constraints on parameters . . . . . 829
  - derivative methods . . . . . 830
  - evaluator types . . . . 741, 751, 829, 844
  - GMM example in Mata . . . 852–854
  - gradient methods . . . . . 824–829
  - linear program simplex method . . . . . 686
  - linear-form methods . . . . . 830
  - Mata example . . . . . 821–824

- iterative methods, *continued*
- Mata optimization ..... 741
  - maximization options ..... 824
  - messages during iterations .... 826
  - ml command methods ..... 832, 844–851
  - moptimize() Mata function .....  
..... 741–751
  - multicollinearity ..... 840
  - multiple optimums ..... 827
  - Newton–Raphson method .....  
..... 819–824
  - NLS example ..... 836
  - nonconvergence ..... 826
  - nonlinear IV example in Mata ....  
..... 852–854
  - not identified ..... 826
  - NR example in Mata ..... 821–824
  - numerical derivatives ..... 828
  - optimization in Mata .... 741–754,  
851–854
  - optimization techniques .. 825, 831
  - optimize() Mata function .....  
..... 751–754
  - overview of optimization tools ....  
..... 829–831
  - Poisson example in Mata .. 752–754
  - Poisson gradient and Hessian. 820
  - quadratic form methods ..... 830
  - single-index model example .....  
..... 833–835
  - step-size adjustment ..... 825
  - stopping criteria ..... 826
  - two-index model example .....  
..... 835–836
- ivpoisson gmm command... 1085–1087
- ivprobit command ..... 889
- ivqte command ..... 1389
- ivreg2 command ..... 347
- ivregress command ..... 316, 894
- ivtobit command ..... 989
- J**
- jackknife ..... see variance–covariance  
matrix
- jackknife IV estimator ..... 365
- jackknife prefix ..... 576
- javacall command ..... 29
- jive command ..... 365
- K**
- k*-class estimator ..... 364
- K*-fold cross-validation ..... 1472–1475
- k*-means clustering ..... 1501
- Kaplan–Meier estimate ..... 1106, 1108
- kdensity command ..... 73, 93
- keep command ..... 58
- kernel density plot ..... 73–76
- kernel regression .... see nonparametric  
methods
- kmatch command ..... 1334
- knnreg command ..... 676
- kurtosis measure ..... 88
- L**
- label command ..... 46
- Lagrange multiplier test ..... see  
hypothesis tests
- lars command ..... 1480
- lasso ..... see also machine learning,  
1479–1480
- adaptive lasso ..... 1488–1489
  - application ..... 1482–1493, 1503
  - clustered data .... 1482, 1489, 1505
  - distribution of estimators .....  
..... 1481–1482
  - elastic net .. 1480–1481, 1489–1491
  - inference ..... 1507–1524
  - linear example ..... 1483–1491
  - logit ..... 1492–1493
  - oracle property ..... 1481
  - Poisson ..... 1492–1493
  - postlasso OLS ..... 1482, 1487
  - probit ..... 1492–1493
  - selection(bic) ..... 1489, 1505
  - sparsity assumption ..... 1508

- lasso
  - linear command . . . . . 1482
  - logit command . . . . . 1492
  - poisson command . . . . . 1492
  - probit command . . . . . 1492
- lassocoeef command . . . . . 1486
- lassogof command . . . . . 1487
- lassoinfo command . . . . . 1486
- lassoknots command . . . . . 1484
- lassoselect command . . . . . 1485
- LATE . . . . . 353, 1356–1362
  - application . . . . . 1360–1362
  - assumptions . . . . . 1359
  - complier . . . . . 1357
  - estimator . . . . . 330
  - marginal treatment effects . . . 1362
  - monotonicity . . . . . 1359
  - Wald estimator . . . . . 1357
- latent class models . . . *see* finite mixture models
- least-absolute-deviations regression . . . . . 128–132, 685
  - censored . . . . . 960
- least-angle regression . . . . . 1480
- least-squares dummy-variable estimator . . . . . 273, 399–401
- leave-one-out cross-validation . . . . 1437, 1475
- LEF . . . . . *see* linear exponential family
- leverage . . . . . 264, 362
- likelihood-ratio test . . . . *see* hypothesis tests
- LIML estimator . . . . . 364
- lincom command . . . . . 494
- linear discriminant analysis . . . . . 1500
- linear exponential family . . . . . 583
  - definition . . . . . 589
  - examples . . . . . 583
  - Poisson example . . . . . 590
- linear mixed models . . . . . 277–286, 377
  - cluster-robust standard errors . . . . . 281–284
- linear probability model . . 471, 859, 862
- linear regression model . . . . . 96–205
  - basic results . . . . . 96–102
  - Bayesian example . . . . . 1528–1532, 1549–1551
  - bootstrap variance matrix . . . . 101, 117–120
  - Box–Cox transformation . . . . . 135
  - cluster FE example . . . . . 269–277
  - cluster FGLS example . . . . . 265–269
  - cluster OLS example . . . . . 256–265
  - cluster PA example . . . . . 266
  - cluster RE example . . . . . 267–269
  - cluster-robust variance . . . . . 99, 116
  - constraints on parameters . . . . 106, 293
  - decomposition analysis . . . . . 186–195
  - default standard errors . . . . . 98
  - difference in differences . . . . . 195–204
  - elasticities . . . . . 182
  - endogenous regressors . . . . . 241, 305–371
  - for one-sample  $t$  test . . . . . 121
  - for two-sample  $t$  test . . . . . 122
  - generalized least squares . . 245–303
  - heteroskedastic errors FGLS . . . 254
  - heteroskedastic-robust variance . . . . . 99
  - hypothesis tests . . . . . 106
  - influential observations . . . . . 126
  - instrumental variables . . . . . 305–371
  - lasso example . . . . . 1483–1491
  - linear probability model . . . . . 862
  - log versus linear regression . . . . 102
  - marginal effects . . . . . 175–186
  - Mata code for OLS . . . . . 145–146
  - measurement-error example . . . . 240
  - omnibus test . . . . . 137
  - panel-data basics . . . . . 373–419
  - panel-data extensions . . . . . 421–457
  - prediction . . . . . 149–175
  - prediction in logs . . . . . 152
  - prediction of actual value . . . . . 158
  - prediction of conditional mean . . . . . 158

- linear regression model, *continued*  
 prediction with binary regressor ..  
     ..... 153  
 regression in logs..... 102  
 residual analysis..... 124–126  
 retransformation problem.....  
     ..... 152–153  
 sampling weights..... 140–145  
 simulations to verify properties ...  
     ..... 232–242  
 specification analysis ..... 123–132  
 specification tests..... 132–140  
 standard error estimation.. 98–102  
 survey data..... 295–301  
 systems of equations..... 286–295  
 test against log-linear ..... 135  
 test interpretation..... 138–140  
 test of heteroskedasticity ..... 136,  
     253  
 test of normality..... 137  
 test of omitted variables..... 133  
 transform dependent variable.....  
     ..... 94–96  
 two-stage least squares..... 314  
 variance matrix estimation .....  
     ..... 98–102  
 weighted marginal effects..... 144  
 weighted prediction..... 144  
 weighted regression..... 142, 300  
 link function..... 590  
**linktest** command..... 134  
**list** command..... 39, 48  
 local average treatment effects.....see  
     LATE  
**local** command..... 24  
 local constant regression.. 78, 677, 1434  
 local linear regression .... 78, 677, 1435  
 local macros... 24–25, 29, 717, 718, 720  
 local polynomial regression..... 79,  
     676–679, 1379  
**locreg** command..... 885  
**log** command..... 16  
 log file..... 16  
 log–log survivor curve ..... 1114  
 log-integrated hazard curve..... 1114  
 log-linear regression ..... 102  
     Poisson regression alternative.....  
     ..... 1026  
     retransformation problem..... 152  
     test against linear..... 135  
 logical operators..... 14  
**logistic** command..... 861  
**logit** command..... 473, 861  
 logit model..... see binary outcome  
     models  
     brief summary..... 461  
**logitfe** command..... 1156  
 loglogistic model ..... 1124  
 lognormal data  
     linear regression model..... 102  
     tobit model..... 961–970  
 lognormal durations model ..... 1124  
**loneway** command..... 259  
 long command lines..... 18  
 long-form data ... 63, 414–418, 914–915  
 longitudinal data ..... see panel data  
**loocv** command..... 1475  
 looping commands ..... 26–29  
     **foreach** ..... 27  
     **forvalues** ..... 28  
     **while** ..... 28  
**lowess** command..... 80, 679  
 lowess regression..... 80, 679  
**lpoly** command..... 79, 677  
     compared with **npregress** ... 1448  
**lroc** command..... 875  
**lrtest** command..... 500
- ## M
- machine learning..... 1465–1525  
 augmented IPW ..... 1522  
 bagging..... 1499  
 big data ..... 1497  
 boosting ..... 1500, 1506  
 causal analysis ..... 1507–1524  
 classification ..... 1496, 1500–1501  
 cross-fit partialing out.. 1514–1515  
 double or debiased ..... 1514–1515  
 double selection..... 1515–1516



- machine learning, *continued*
- inference . . . . . 1465–1525
  - instrumental variables . . 1518–1521
  - lasso . . . . . 1479–1493, 1507–1524
  - logit . . . . . 1492–1493, 1517–1518
  - neural networks . . 1497–1498, 1505
  - oracle property . . . . . 1481
  - orthogonalization . . . . . 1513–1514
  - overview . . . . . 1496–1501
  - partial linear model . . . . 1507–1516
  - partialing out . . . . . 1509–1514
  - Poisson . . . . . 1492–1493, 1516–1517
  - prediction application . . 1501–1507
  - random forests . . . . . 1499, 1505
  - regression trees . . . . . 1498–1499
  - shrinkage estimators . . . 1477–1493
  - supervised learning . . . . 1496–1501
  - support vector machines . . . . 1501
  - unsupervised learning . . 1496, 1501
- macros . . . . . 23–26
- compared with scalars . . . . . 26
  - global . . . . . 24
  - local . . . . . 24–25
- Mallow's  $C_p$  measure . . . . . 1469
- marginal effects . . . . . 175–182, 466–472, 612–629
- at representative value . . 177, 181, 467, 613, 618–619
  - at the mean . . . . . 177, 467, 613, 617–618
  - average marginal effect . . . . . . . 113–116, 177, 178, 467–471, 613, 615–617
  - calculus method . . . . . 176, 466, 612
  - calculus versus finite difference . . . . . 616
  - elasticities . . . . . 182, 625–628
  - factor-variable notation . . 114, 179, 622–625
  - finite-difference method . . 176, 467, 613
  - in Bayesian model . . . . . 1574–1576
  - in binary outcome models . . . . . 877–880
  - in count-data models . . . . . 1033
- marginal effects, *continued*
- in linear mixed models . . . . . 1230
  - in linear regression model . . . . . 175–186
  - in log-linear model . . . . . 102
  - in multinomial outcome models . . . . . 902, 912, 920, 937, 941
  - in nonlinear models . . . . . 466–472, 612–629
  - in panel logit model . . . . 1160–1161
  - in panel ordered logit model . . 1166
  - in spatial regression . . . . 1421–1423
  - in tobit model . . . . . 956–959
  - in treatment-effects models . . 467, 1349, 1355
  - interacted regressors . . . . 113–116, 179, 623
  - introduction for AME . . . . 113–116
  - manual computation . . . . 620–622, 1355
  - polynomial regressors . . . . . 622
  - population average marginal effects . . . . . 620
  - predictive margins . . . . . 161–175, 609–612
  - semielasticities . . . . . 184
  - summary of types . . . . . 177, 467, 613
  - weighted . . . . . 144
  - which to use . . . . . 619
- marginal likelihood . . . . . 1533
- computation . . . . . 1567
- marginal treatment effects . . 628, 1362
- user-provided quantity . . . . . 619
- margins command . . 162, 468, 609–616, 1341, 1446, 1449
- margins,
- contrast command . . . . . 174
  - dydx() command . . . . . 115, 177, 614, 615
  - expression() command . . . . . 610
  - predict(cte) command . . . . 1349
  - pwcompare command . . . . . 171
  - subpop() command . . . . . 1349
- marginsplot command . . . . . 169, 1447

- Markov chain Monte Carlo methods...  
 ..... 1527–1576  
 acceptance rate ..... 1540  
 acceptance rule ..... 1535, 1584  
 blocking parameters.... 1560–1561  
 burn-in draws.....1534  
 convergence.....1541, 1558  
 data augmentation ..... 1589–1595  
 diagnostics.....1541, 1588, 1593  
 different starting values..... 1543  
 draws example ..... 225–227  
 efficiency.....1540, 1588, 1593  
 efficiency of draws ..... 1554  
 efficiency statistic ..... 1542  
 for multiple imputation.....  
 ..... 1597–1607  
 Gelman–Rubin statistic..... 1545  
 Gibbs sampler example in Mata ..  
 ..... 1589–1595  
 Gibbs sampling algorithm.... 1536  
 Gibbs within Metropolis–Hastings  
 ..... 1561–1562  
 Metropolis algorithm.....1535  
 Metropolis–Hastings algorithm ...  
 ..... 1535  
 Metropolis–Hastings algorithm in  
 Mata ..... 1584–1589  
 Monte Carlo simulation error.....  
 ..... 1542  
 multiple chains..... 1543–1545  
 pitfalls..... 1556–1560  
 posterior draws .. 1534, 1541, 1547  
 prediction..... 1569–1572  
 proposal density ..... 1535  
 random-walk Metropolis–Hastings  
 algorithm ..... 1535  
 sensitivity analysis..... 1543  
 Mata..... 29, 727–754  
 ado-files using Mata..... 739  
 Bayesian model AME... 1575–1576  
 colon operators ..... 732  
 combining matrices ..... 736  
 commands in Mata..... 727  
 commands in Stata ..... 728
- Mata, *continued*  
 data augmentation example .....  
 ..... 1589–1595  
 declaration of program arguments  
 ..... 740  
 element-by-element operators..732  
 Gibbs sampler example..... 225,  
 1589–1595  
 GMM example ..... 852–854  
**help** command.....729  
 identity matrix.....730  
 matrix  
 cross products ..... 735  
 functions ..... 733  
 input ..... 729  
 inversion.....734  
 of constants.....730  
 operators ..... 732  
 subscripts.....735  
 Metropolis–Hastings example .....  
 ..... 1584–1589  
 Newton–Raphson example.....  
 ..... 821–824  
 OLS regression example... 145–146  
 optimization functions ... 741–754,  
 851–854  
 optimization in Mata.....741  
 overview ..... 727  
 Poisson regression example.....  
 ..... 821–824  
 program example ..... 738  
 programming in Mata ... 738–740  
 Stata commands in Mata ..... 728  
 Stata interface functions ..... 732  
 Stata matrix from Mata matrix...  
 ..... 737  
 Stata variables from Mata matrix  
 ..... 737  
 matching methods.....see treatment  
 effects  
 matrices in Mata.....see Mata

- matrices in Stata.....709–715  
   combining matrices ..... 712  
   matrix  
     cross products..... 713  
     functions ..... 713  
     input ..... 709  
     operators ..... 712  
     subscripts..... 711  
   OLS example..... 714  
   overview ..... 709
- matrix**  
   **accum** command..... 714  
   **command**..... 20, 709–715  
   **define** command ..... 710  
   **rownames** command..... 710  
   **vecaccum** command..... 714
- matrix algebra definitions.....97
- maximization ... see iterative methods
- maximum likelihood ..... 583–586  
   computational methods...819–851  
   definition of MLE..... 583  
   misspecified density ..... 583  
   pseudo-MLE..... 583  
   quasi-MLE..... 583
- maximum simulated likelihood estimator.....930–931
- MCMC...see Markov chain Monte Carlo methods
- mean absolute error ..... 1475
- mean group estimator ..... 454
- mean squared error ..... 1468
- measurement-error example ..... 240, 1240–1245
- median regression ..... 128–132, 685
- meglm** command..... 1225, 1226
- meintreg** command ..... 961
- MEM..... see marginal effects, at the mean
- menbreg** command ..... 1229
- meologit** command ..... 942
- meoprobit** command..... 942
- mepoisson** command..... 1226
- MER..... see marginal effects, at representative value
- merge** command ..... 63
- mestreg** command ..... 1137
- metobit** command.....961
- Metropolis–Hastings algorithm .....  
   ..... see Markov chain Monte Carlo methods
- mhtexp** command .....512
- mi**  
   **convert** command.....1606  
   **estimate** command.....1598  
   **impute** command..... 1598  
   **misstable** command ..... 1601  
   **register** command.....1602  
   **xeq** prefix.....1603
- minimization ... see iterative methods
- missing at random ..... 991, 1595
- missing completely at random.... 991, 1595
- missing data ..... 49–51, 991–995  
   complete-case analysis ..... 992  
   endogenous sample selection .. 994  
   imputation methods..... 995, 1595–1607  
   inverse-probability weighting .....  
     .....993–994  
   MAR ..... 991, 1595  
   MCAR ..... 991, 1595  
   mechanisms ..... 991–992  
   MNAR..... 991, 1595  
   multiple imputation... 1595–1607  
   panel attrition.....995–1018  
   sample selection.....1011–1018
- missing not at random..... 991, 1595
- missing-value codes ..... 49
- missing-values imputation..... 51, 1595–1607
- mixed** command ..... 279, 284
- mixed models  
   generalized linear model.....  
     ..... 1225–1231  
   linear.....277–286  
   nonlinear... 637, 1162, 1224–1231  
   Poisson example.....1226–1231
- mixlogit** command ..... 935
- mkspline** command.....671, 672

- ml
  - check command.....834, 837
  - commands.....831–851
  - maximize command..... 832
  - method d2..... 849
  - method gf0..... 850
  - method lf..... 832
  - method lf0..... 844, 846
  - method lf1..... 847
  - method lf2..... 848
  - model command.....832, 844
  - search command..... 834
  - trace command..... 838
- mlevel command..... 845
- mlexp command..... 585
- mlmatbysum command..... 850
- mlmatsum command.....846, 850
- mlogit command..... 908
- mlsum command..... 845, 849
- mlvecsum command.....845, 850
- model approach versus weighted regression..... 143
- model diagnostics..... 629–632
- model selection..... 492, 630–631, 1466–1476
  - backward selection..... 493
  - based on predictive ability..... 1466–1476
  - based on statistical significance... 492–494
  - Bayesian..... 1567–1569
  - best subsets.....1476
  - conservative model selection..1481
  - consistent model selection....1481
  - count-data models..... 1067
  - cross-validation..... 1471–1475
  - finite mixture models..... 659
  - forward selection.....492
  - information criteria..... 630–631, 659, 1468
  - pretest bias..... 493
  - stepwise selection..... 492, 1476
- moment-based tests... see specification tests
- Monte Carlo methods... see simulation
- moptimize() Mata function...741–751
  - method d2..... 746
  - method gf\*.....749
  - method q0..... 749
  - method q1..... 749
  - Poisson example..... 742–751
- mprobit command.....929
- mtebinary command.....1362
- mtefe command.....1362
- mtest option..... 507, 509
- multicollinearity..... 840
- multinomial outcome models.. 901–947
  - additive random-utility model.... 904, 922
  - alternative-specific regressors.... 903, 906, 907, 914–922, 930–937
  - basic theory..... 901–903
  - bivariate probit model... 943–945
  - BLP market demand.....937
  - case-specific regressors... 903, 906, 907, 915, 929
  - clustered data.....942
  - cm commands... 915–922, 931–938
  - coefficient interpretation.. 910, 919
  - conditional logit model... 914–922
  - conditional versus multinomial logit.....919
  - data example..... 905–908
  - FMM application..... 1211–1214
  - Geweke–Hajivassiliou–Keane simulator..... 930
  - independence of irrelevant alternatives.....922
  - long-form data..... 914–915
  - marginal effects.... 902, 912, 920, 927, 934, 941
  - maximum likelihood estimator.... 902
  - maximum simulated likelihood estimator..... 930–931
  - model comparison..... 928
  - multinomial logit model.. 908–914
  - multinomial probit model..... 929–934

- multinomial outcome models, *continued*
  - multivariate outcomes . . . . . 943–946
  - nested logit model . . . . . 922–928
  - ordered logit model . . . . . 939
  - ordered outcome models . . . . . 938–942
  - ordered probit model . . . . . 939
  - overview . . . . . 901–905
  - predicted probabilities . . . . . 911, 920, 927, 934, 940, 945
  - prediction for new alternative . . . . . 920
  - probabilities . . . . . 902
  - random-parameters logit . . . . . 934–938
  - rank-ordered logit . . . . . 922
  - rank-ordered probit . . . . . 934
  - robust standard errors . . . . . 903, 951
  - Stata commands summary . . . . . 904
  - wide-form data . . . . . 914–915
- multiple imputation . . . . . 1595–1607
  - inference . . . . . 1595
- multivalued treatment-effects
  - example . . . . . 1324
- multproc** command . . . . . 507, 509
- mvdecode** command . . . . . 50
- N**
- natural spline . . . . . 672
- nbreg** command . . . . . 1035
- nearest-neighbors regression . . . . . 676
- negative binomial model . . . . . *see*
  - count-data models
- Nelson–Aalen estimate . . . . . 1108
- neural networks . . . . . 1497–1498, 1505
- Newton–Raphson algorithm . . . . . *see*
  - iterative methods
- nl** command . . . . . 474, 587
- nlcom** command . . . . . 495
- nlogit** command . . . . . 925
- nlogitgen** command . . . . . 924
- nlogittree** command . . . . . 925
- nlstur** command . . . . . 945
- nonlinear IV . . . . . 593, 851–854
  - additive versus multiplicative error . . . . . 1084
  - count-data example . . . . . 1084–1089
- nonlinear least squares . . . . . 587–589
  - estimator definition . . . . . 474, 587
  - example . . . . . 588, 836–837
  - for Poisson model . . . . . 1042–1043
  - introduction . . . . . 474–475
- nonlinear regression . . . . . 459–478, 579–641
  - cluster FE example . . . . . 637–640
  - cluster PA example . . . . . 633–635
  - cluster RE example . . . . . 635–637
  - clustered data example . . . . . 632–640
  - computational methods . . . . . 819–856
  - generalized linear models . . . . . 589–591
  - generalized method of moments . . . . . 591–596
  - introduction . . . . . 459–478
  - logit as example . . . . . 472–474
  - marginal effects . . . . . 466–472, 612–629
  - maximum likelihood . . . . . 583–586
  - mixed models . . . . . 637, 1162, 1224–1231
  - nonlinear IV . . . . . 851–854
  - nonlinear least squares . . . . . 474–475, 587–589
  - panel data . . . . . 1139–1190
  - Poisson as example . . . . . 580–640
  - prediction . . . . . 604–609
  - predictive margins . . . . . 609–612
  - probit as example . . . . . 462–472
  - standard errors . . . . . 597–604
  - Stata commands summary . . . . . 579
  - summary of methods . . . . . 582–597
  - systems of equations . . . . . 945
  - variance estimate . . . . . 597–604
- nonparametric methods . . . . . 73–82, 675–681, 1433–1463
  - bandwidth choice . . . . . 74, 1437, 1448
  - bootstrap standard errors . . . . . 1443
  - cluster-robust standard errors . . . . . 1443
  - cross-validation . . . . . 1437
  - for regression discontinuity design . . . . . 1373, 1375
  - graphs . . . . . 77–82, 678, 1447, 1448
  - kernel density estimation . . . . . 73–76

- nonparametric methods, *continued*  
 kernel function . . . . 73–74, 78, 676  
 kernel regression . . . . . 78, 676,  
 1434–1453  
 leave-one-out cross-validation . . . .  
 . . . . . 1437  
 Li–Racine kernel . . . . . 1436  
 local  
 constant . . . . . 78, 677, 1434  
 linear . . . . . 78, 677, 1435  
 linear *m*-estimation . . . . . 1438  
 logit . . . . . 885  
 polynomial . . . . . 79, 676–679, 1379  
 lowess regression . . . . . 80, 679  
 multiple regressors . . . . . 1433,  
 1450–1453  
 nearest neighbors . . . . . 676  
 observations not identified . . . . 1443  
 partial effects . . . . 1435, 1442, 1446  
 plugin bandwidth . . . . . 79, 81, 678,  
 1433  
 semiparametric regression . . . . .  
 . . . . . 680–681, 1453–1463  
 series regression . . . . . 1438–1439,  
 1449–1450  
 trimming . . . . . 1445  
**npregress kernel** command . . 82, 678,  
 1437  
 compared with **lpoly** . . . . . 1448  
**npregress series** command . . . . . 680,  
 1439  
 numerical derivatives . . . . . 828
- O**  
**oaxaca** command . . . . . 192  
 odds ratio for logit model . . . . . 474  
**ologit** command . . . . . 940  
 one-sample *t* test . . . . . 120  
 computed by OLS regression . . . 121  
 operators . . . . . 13–14  
**oprobit** command . . . . . 941  
 optimization . . . . . see iterative methods
- optimize()** Mata function . . . 751–754,  
 852  
 method **d2** . . . . . 852  
 method **gf2** . . . . . 752  
 Poisson example . . . . . 752–754  
**order** command . . . . . 60  
 ordered outcome models . . . . . see  
 multinomial outcome models  
 ordinary least squares . . . . . see linear  
 regression model  
**orthpoly** command . . . . . 668  
**outfile** command . . . . . 58  
**outreg** command . . . . . 109  
**outreg2** command . . . . . 109  
**outsheet** command . . . . . 58  
**outsum** command . . . . . 89  
 overidentifying restrictions test . . . . see  
 specification tests
- P**  
*p*-values . . . . . 480–485  
 bootstrap . . . . . 557  
 chi-squared versus *F* . . . . . 481  
 computation . . . . . 480–485  
 standard normal versus *t* . . . . 481,  
 484  
 used in Stata commands . . . . . 484  
 panel attrition . . . . . 995–1018  
 inverse-probability weighting . . . .  
 . . . . . 1004–1011  
 sample refreshment . . . . . 1018  
 sample selection . . . . . 1011–1018  
 panel data . . . . . 373–419, 421–457,  
 1139–1190  
 2SLS estimator . . . . . 430  
 Arellano–Bond estimator . . . . .  
 . . . . . 428–445  
 attrition . . . . . 995–1018  
 augmented mean group estimator  
 . . . . . 454  
 balanced panel . . . . 374, 1000–1004  
 balanced panel creation . . . . . 1001  
 basics . . . . . 373–419  
 between estimator . . . . . 401–402  
 between variation . . 383, 407, 1147

- panel data, *continued*
- bias correction . . . . . 1155
  - binary outcome models . . . . .
    - . . . . . 1148–1163
  - cluster–robust inference . . 377–378, 1144
  - cointegration . . . . . 455–456
  - common correlated estimator . . 454
  - comparison of estimators . . 406–412
  - correlated random-effects model . .
    - . . . . 376, 405, 1142, 1156
  - count-data models . . . . . 1172–1184
  - data management . . . . . 414–418
  - data summary . . . . . 379–394
  - difference in differences . . . . . 1363
  - dynamic
    - linear model . . . . . 428–445
    - logit model . . . . . 1162
    - systems model . . . . . 445
  - endogeneity . . . . . 375, 1147
  - endogenous regressors . . . 421–445, 1187–1188
  - exogeneity . . . . . 414
  - few clusters . . . . . 378, 561–567
  - first-difference
    - estimator . . . . . 412–414
    - IV estimator . . . . . 429
    - model . . . . . 429
  - fixed effects . . . . . 401, 1140
  - fixed versus random effects . . . 409
  - fixed-effects estimator . . . 397–401, 451, 1154, 1179, 1185
  - fixed-effects model . . 375, 397, 1140
  - generalized estimating equations . .
    - . . . . . 1141
  - GMM estimator . . . . . 430
  - Hausman test . . . . . 409
  - Hausman–Taylor estimator . . . . .
    - . . . . . 425–427
  - heterogeneous panels . . . . . 454
  - in spatial regression models . . . . .
    - . . . . . 1428–1429
  - incidental parameters problem . . . .
    - . . . . . 1140
- panel data, *continued*
- individual-effects model . . 375, 397, 402, 451, 1139
  - individual-invariant regressor . . . . .
    - . . . . . 374, 385, 401
  - instrumental-variables estimator . .
    - . . . . . 421–445
  - interactive-effects estimator . . . 453
  - least-squares dummy-variable estimator . . . . . 399–401
  - linear
    - extensions . . . . . 421–457
    - mixed models . . . . . 377
    - model overview . . . . . 373–379
  - logit model . . . . . 1149–1167
  - long panel . . . . . 374, 445–456
  - long-form data . . . . . 414–418
  - marginal effects . . . . . 1160–1161
  - mean group estimator . . . . . 454
  - missing data . . . . . 995–1018
  - mixed nonlinear models . . . . . 1162
  - MNAR assumption . . . . . 1011
  - multiplicative-effects model . . 1140
  - Mundlak correction . . . . 405, 1142, 1156
  - negative binomial model . . . . .
    - . . . . . 1183–1184
  - nonlinear
    - example . . . . . 1145–1147
    - models . . . . . 1139–1190
    - models overview . . . . . 1139–1144
  - ordered logit model . . . . 1164–1167
  - ordered probit model . . 1164–1167
  - Poisson model . . . . . 1173–1182
  - pooled estimator . . . . . 390–391, 394–396, 447–451, 1149, 1174, 1184
  - pooled model . . . . . 376, 394, 1141
  - population-averaged estimator . . . .
    - . . . . 394–396, 1151, 1175
  - population-averaged model . . . 376, 1141
  - prediction . . . 411, 1160–1161, 1181
  - probit models . . . . . 1163
  - quantile regression . . . . . 1184–1187

- panel data, *continued*
- $R^2$  ..... 406
  - random-coefficients model .... 377, 1141
  - random-effects estimator ..... .. 402–406, 451, 1153, 1177
  - random-effects model .... 376, 402, 1140, 1168
  - sample refreshment ..... 1018
  - sample-selection model ..... .. 1169–1172
  - separate regressions ..... 453–455
  - short panel ..... 373, 428
  - spatial correlation in panel .... 449
  - Stata linear commands summary .. .. 379
  - Stata nonlinear commands ... 1144
  - systems IV estimator ..... 445
  - time-invariant regressor .. 374, 385, 390, 391, 397, 399, 404, 409, 413, 423, 425, 427, 433
  - time-series autocorrelations ... 391, 393
  - time-series operators ..... 391
  - time-series plots ..... 386
  - tobit models ..... 1167–1172
  - two-way fixed effects ..... 401, 406
  - two-way-effects model ... 377, 447
  - unbalanced panel ... 374, 375, 383, 406, 1000–1004, 1146
  - unit roots ..... 455–456
  - variance components ..... 406
  - wide-form data ..... 414–418
  - within
    - estimator ..... 397–401
    - scatterplot ..... 389
    - variation ..... 383, 386, 407
    - within variation ..... 1147
- partial identification ..... 1019
- partial linear model ... 681, 1453–1456, 1461
- lasso estimation ..... 1507–1516
- partialing-out estimator ..... 1509–1514
- instrumental variables .. 1518–1521
  - nonlinear models ..... 1509
- pca** command ..... 1494
- percentiles ..... 88
- permutation tests .. see hypothesis tests
- piecewise linear regression ..... 669
- plots ..... see graphs
- plugin** command ..... 29
- poisson** command ..... 584, 1029
- Poisson model .. see count-data models
- poivre** command ..... 1519, 1520
- polynomials
- fractional ..... 667
  - global ..... 665–668, 680
  - local ..... 676–679
  - orthogonal ..... 668
- population-averaged model ... see panel data
- poregress** command ..... 1510
- post** command ..... 218
- postclose** command ..... 218
- posterior distribution ..... 1533
- for normal model ..... 1548–1549
  - for normal regression ..... 1551
  - MCMC draws ..... 1534
- posterior odds ratio ..... 1568–1569
- posterior predictive distribution .. 1569
- postestimation commands summary ... .. 105, 586
- postfile** command ..... 218, 517
- potential outcomes ..... 1271–1272
- potential-outcome means ..... see treatment effects
- power calculations ..... 519–528, 1276–1281
- limitations ..... 1281
- power** commands ..... 519–528
- power onemean** command ..... 520
- power twomeans** command ..... 1276
- prchange** command ..... 879
- prcounts** command ..... 1039
- precision parameter ..... 1548
- predict** command ..... 150, 411, 604
- prediction ..... 149–161, 604–609, 1465–1525
- at a specified value ..... 607
  - Bayesian ..... 1569–1572



- prediction, *continued*
- in binary outcome models .....
    - .....869–876
  - in count-data models... 1036–1041
  - in duration models ..... 1112, 1121
  - in levels from log model ..... 152
  - in levels from tobit in logs .....
    - .....986–989
  - in linear mixed models ..... 283
  - in linear panel-data model .... 411
  - in linear regression model .....
    - .....149–175
  - in nonlinear mixed models... 1230, 1258
  - in nonlinear models ..... 604–609
  - in panel logit model .... 1160–1161
  - in panel Poisson model ..... 1181
  - in sample ..... 149–157, 1466–1472
  - in spatial regression models .....
    - ..... 1421–1423
  - in two-part model ..... 988
  - machine learning ..... 1465–1525
  - machine learning application .....
    - ..... 1501–1507
  - of actual value ..... 158, 606
  - of conditional mean ..... 158, 606
  - of multinomial model probabilities .....
    - ..... 911
  - of nonlinear quantity ..... 605
  - out of sample ..... 157–161, 606, 1466–1475
  - weighted ..... 144
  - with binary regressor ..... 153
- predictive margins ... 161–175, 609–612
- at a specified value ..... 611
  - at means ..... 611
  - average ..... 610
  - contrasts of predictive margins ...
    - ..... 173
  - for a categorical variable.. 164, 612
  - for a continuous variable ..... 168
  - in nonlinear models ..... 609–612
  - pairwise comparisons ..... 171
  - plots of predictive margins .... 169
- predictive means ..... 161–175
- predictnl** command ..... 605
- prefix** command ..... 8
- preserve** command .. 61, 155, 607, 620
- principal components ..... 1494–1496, 1501, 1505
- standardize regressors ..... 1494
- prior distribution ..... 1533, 1550–1551
- can identify an unidentified model ..... 1557
  - conjugate prior ..... 1550
  - example of specification ..... 1552
  - flat prior ..... 1550, 1581
  - hierarchical prior ..... 1562–1566
  - inverse-gamma prior ..... 1550
  - Jeffreys prior ..... 1550
  - noninformative prior ..... 1550
  - normal prior ..... 1550
  - uninformative prior can be informative ..... 1556
  - Wishart prior ..... 1551, 1598
  - Zellner’s *g*-prior ..... 1551
- probit** command ..... 462, 861
- probit model ..... *see* binary outcome models
- brief summary ..... 461
- probitfe** command ..... 1156
- program define** command ..... 716
- program drop** command ..... 716
- programs ..... 716–725
- bootstrap example... 550, 557, 569
  - central limit theorem example ....
    - ..... 216–219
  - checking program ..... 837–843
  - debugging ..... 722
  - in Mata ..... 738–740
  - in Stata ..... 216, 716–725
  - Monte Carlo integration .. 229–230
  - named positional arguments... 718
  - OLS with chi-squared errors .....
    - ..... 233–239
  - overview ..... 716
  - parsing syntax ..... 719
  - positional arguments ..... 717

- programs, *continued*
- r-class example..... 719
  - temporary variables..... 718
- propensity score .... *see also* treatment effects
- definition ..... 1291
- propensity-score overlap ..... 1315
- proportional hazards model ..... *see* duration models
- proportions data regression ... 862, 897
- prvalue command..... 623, 876, 1040
- pscore command..... 1321
- pseudo- $R^2$  ..... 629, 870
- pvar command ..... 445
- pwcorr command..... 103
- pylearn package ..... 1523
- python command..... 29, 1523
- Q**
- qcount command..... 1091
- qnorm command..... 126
- qplot command..... 689
- qqplot command..... 126
- qreg command ..... 129, 686
- qreg2 command..... 693
- quadchk command ..... 1152
- quadratic discriminant analysis ... 1500
- quadrature methods ..... 227, 1152
- adaptive Gauss–Hermite ..... 227
  - for multinomial probit ..... 929
- quantile regression..... 683–708, 1388–1401
- bootstrap robust variance ..... 687
  - censored data ..... 699
  - coefficient comparison across quantiles ..... 692, 698
  - coefficient interpretation ..... 690
  - computation ..... 686
  - conditional quantiles ..... 684–706, 1332–1334, 1388–1389
  - count data ..... 1090–1096
  - counterfactual analysis..... 1398
  - data example ..... 688–699
  - definition of quantiles ..... 683
- quantile regression, *continued*
- endogenous data ..... 699
  - endogenous regressors .. 1388–1394
  - endogenous treatment effects ..... 1388–1401
  - generated data example .. 699, 703
  - inverse-propensity score weighting ..... 1395
  - loss functions ..... 684
  - marginal effects ..... 692
  - panel data ..... 1184–1187
  - quantile estimator..... 685
  - rank invariance..... 1393–1394
  - rank similarity ..... 1393–1394
  - recentered influence function .. 1396
  - retransformation ..... 692
  - robust standard errors... 686, 693
  - test of coefficient equality ..... 696
  - test of heteroskedasticity ..... 695
  - treatment effect ..... 703–706, 1332–1334
  - unconditional quantile treatment effects ..... 1394–1401
  - unconditional quantiles..... 1394–1401
  - variance matrix estimation .... 686
- query command..... 19
- R**
- $R^2$  ..... 629
- for linear panel models..... 406
  - partial  $R^2$ ..... 340
  - pseudo- $R^2$  ..... 629, 870
- random effects .. *see* clustered data, *see* panel data
- random forests ..... 1499, 1505
- random-number generation..... *see* simulation
- randomization tests..... *see* hypothesis tests
- randomized controlled trials..... 1272–1290
- adding covariates..... 1282–1290
  - assumptions ..... 1272
  - covariate balance ..... 1281

- randomized controlled trials, *continued*  
   design-based ..... 1289  
   limitations ..... 1290  
   OHIE example .... 1305–1323, 1360  
   optimal sample size ..... 1274  
   power calculations ..... 1276–1281  
   two-sample *t* test ..... 1274  
**rbeta()** function ..... 212  
**rbinomial()** function ..... 212  
**rchi2()** function ..... 211  
 r-class commands ..... 20, 216, 719  
**rdbwselect** command ..... 1383  
**rdplot** command ..... 1376  
**rdrobust** command ..... 1379  
 receiver operator characteristics curve ..  
   ..... 875  
 recentered influence function estimator  
   ..... 1396  
**recode** command ..... 53  
 recursive models ..... 308  
 recursive structure ..... 1342  
**reg3** command ..... 367  
**rego** command ..... 194  
**regress** command ..... 8, 22, 103, 273  
 regression decomposition analysis  
   Oaxaca–Blinder ..... 186–193  
   Shapley relative importance .....  
   ..... 193–195  
 regression discontinuity design .....  
   ..... 1369–1388  
   binning ..... 1374  
   fuzzy RD ..... 1384–1386  
   kink RD ..... 1388  
   local approach ..... 1371  
   nonparametric methods ..... 1373,  
   1375  
   parametric approach ..... 1371  
   sharp RD ..... 1370–1384  
 regression splines ..... 668–675, 680  
   basis expansions ..... 674  
   cubic splines ..... 672  
   knots ..... 669  
   natural spline ..... 672  
   piecewise linear regression ..... 669  
 regression splines, *continued*  
   restricted spline ..... 672  
   smoothing splines ..... 674, 1459  
 regression trees ..... 1498–1499  
 regressor balance ..... 1317  
 relational operators ..... 14  
**rename** command ..... 46  
**replace** command ..... 49, 52  
 report generation ..... 147  
 resampling methods ..... *see* bootstrap  
   methods  
**reshape** command ... 63, 200, 414, 418,  
   914  
 residual augmentation approach .... *see*  
   control function estimator  
 residuals  
   definitions of various residuals ....  
   ..... 631–632  
   residual plots ..... 124  
**restore** command ... 61, 155, 607, 620  
 return code ..... 14, 724  
**return** command ..... 233  
**return list** command ..... 20  
**rforest** command ..... 1505  
**rgamma()** function ..... 212, 1024  
 ridge regression ..... 1477–1478  
**rifreg** command ..... 1397  
**rivtest** command ..... 360  
**\_rmcoll** command ..... 841  
**rnbinomial()** function ..... 214, 1025  
**rnormal()** function ..... 210  
 robust regression ... 123–124, 128–132  
   influential observations ..... 126  
 robust standard errors ..... *see*  
   variance–covariance matrix  
 root mean squared error ..... 1474  
**rpoisson()** function ... 212, 1023, 1024  
**rreg** command ..... 128  
**rseed()** option ..... 1529  
**rt()** function ..... 211  
 Rubin causal model ..... *see* potential  
   outcomes  
**runiform()** function ..... 208  
**rvfplot** command ..... 124

**S**

- sample** command . . . . . 59
- sample-selection model . . . see selection models
- sampling weights . . . . . 140, 295
  - in linear regression model . . . . .
  - . . . . . 140–145, 295–301
  - weighted mean . . . . . 141, 299
- save** command . . . . . 57
- saveold** command . . . . . 57
- scalars . . . . . 19, 26
  - compared with macros . . . . . 26
- scatterplot . . . . . 76
- Schoenfeld residual . . . . . 1115
- score test . . . . . see hypothesis tests
- search** command . . . . . 5
- seed for bootstrap . . . . . 539
- seed for random-number generation . . . . . 208
- seemingly unrelated estimations . . . 294
- seemingly unrelated regression
  - equations . . . . . 286–295
  - cross-equation restrictions test . . . . . 291
  - feasible GLS estimation . . . 287–293
  - FGLS example . . . . . 288
  - imposing cross-equation restrictions . . . . . 293
  - nonlinear . . . . . 945
  - robust standard errors . . . . . 290
  - SUR estimator . . . . . 288
  - SUR model . . . . . 287
  - test of error independence . . . . 290
- selection models . . . . . 974–986
  - binary outcomes . . . . . 977
  - copula models . . . . . 982–986
  - exclusion restrictions . . . . . 980
  - heckit model . . . . . 974
  - identification . . . . . 977
  - inverse of the Mills ratio . . . . . 978
  - maximum likelihood estimation . . . . . 976–978
  - model definition . . . . . 975
  - model likelihood . . . . . 975
  - nonnormal selection . . . . . 982–986
- selection models, *continued*
  - ordered outcomes . . . . . 977
  - panel attrition . . . . . 1011–1018
  - panel sample selection . . 1169–1172
  - prediction . . . . . 988
  - sample selection . . . 994, 1169–1172
  - two-step estimation . . 552, 978–981
  - type-2 tobit . . . . . 974
- selection on observables only . . 991, see also unconfoundedness
- sem** command . . . . . 1237
- SEM path builder . . . . . 1240
- semielasticities . . . . . 625–628
  - in linear regression model . . . . 184
  - in Poisson model . . . . . 1033
- semipar** command
  - observations not identified . . 1454
- semiparametric methods . . . . . 680–681, 1453–1463
  - generalized additive model . . . . . 1458–1461
  - partial linear model . . . 1453–1456, 1461
  - Robinson differencing estimator . . . . . 1453
  - semiparametric least squares . . 1457
  - single-index models . . . . 1456–1458
  - trimming . . . . . 1454
- series regression . . . . . 680, 1438–1439, 1449–1450
- set**
  - command . . . . . 19
  - dots** command . . . . . 119
  - iterlog** command . . . . . 130
  - matsize** command . . . . . 273
  - seed** command . . . . 208, 539, 687, 1529
- shrinkage estimators . . . . . 1477–1493
  - bias-variance tradeoff . . . . . 1477
  - elastic net . . 1480–1481, 1489–1491
  - lasso . . . . . 1479–1493
  - least-angle regression . . . . . 1480
  - logit . . . . . 1492–1493
  - Poisson . . . . . 1492–1493
  - probit . . . . . 1492–1493

- shrinkage estimators, *continued*  
 ridge regression ..... 1477–1478  
 standardize regressors ..... 1478  
**simulate** prefix ..... 217  
 simulation ..... 207–244  
 asymptotic power ..... 518  
 Bayesian methods ..... 1527–1578  
 bias of estimator ..... 235  
 bias of standard error estimator ..  
 ..... 236  
 bootstrap example ..... 569  
 central limit theorem example ....  
 ..... 214–219  
 Cholesky decomposition ..... 224  
 computing integrals ..... 227–232  
 direct transformation ..... 221  
 distribution of sample mean .....  
 ..... 214–219  
 draws from multivariate normal ..  
 ..... 223  
 draws from negative binomial .....  
 ..... 1024  
 draws from normal ..... 210  
 draws from Poisson ..... 1023  
 draws from truncated normal .. 222  
 draws from uniform ..... 208  
 draws of continuous variates... 211  
 draws of discrete variates ..... 212  
 draws using Markov chain Monte  
 Carlo method ..... 225–227  
 endogeneity example ..... 308–310  
 endogenous regressors example ...  
 ..... 241  
 FGLS heteroskedastic errors exam-  
 ple ..... 251  
 Gibbs sampler example ..... 225  
 Halton sequences ..... 230, 931  
 Hammersley sequences ... 230, 931  
 inconsistency of estimator ..... 240  
 interpreting simulation output ....  
 ..... 235  
 inverse-probability  
 transformation ..... 220  
 linear regression example .. 232–242  
 measurement-error example ... 240
- simulation, *continued*  
 mixture of distributions ..... 222  
 Monte Carlo integration .. 229–232  
 OLS with chi-squared errors .....  
 ..... 233–239  
 pseudorandom numbers... 208–214  
 seed ..... 208  
 test power computation ..... 238,  
 516–518  
 test size computation ..... 237,  
 513–515  
 using actual data ..... 515  
 using **postfile** command ..... 218  
 using **simulate** prefix .... 217, 233  
 weak-instruments example .....  
 ..... 333–337  
 simultaneous equations model .....  
 ..... 306–310  
 3SLS estimation ..... 367–368  
 binary endogenous ..... 889  
 first-stage equation ..... 308  
 recursive system ..... 308  
 reduced form ..... 307  
 structural equation ..... 308  
 structural model ..... 306  
 single-index models ..... 681  
 coefficient interpretation .. 470, 614  
 definition ..... 614  
 numerical derivatives for ..... 828  
 semiparametric estimation .....  
 ..... 1456–1458  
 semiparametric least squares .. 1457  
 skewness measure ..... 88  
**sktest** command ..... 973  
**sls** command ..... 1457  
**sort** command ..... 60  
 spatial regression ..... 1405–1431  
 contiguity matrix ..... 1411  
 direct effect ..... 1423  
 FGLS with spatial errors ..... 1416  
 generalized spatial 2SLS estimation  
 ..... 1418, 1424  
 geospatial data ..... 1407–1411  
 geospatial distance ..... 1410  
 geospatial shapefile ..... 1408

- spatial regression, *continued*
- heat maps ..... 1409
  - indirect effect ..... 1423
  - instrumental variables ..... 1419, 1427–1428
  - inverse-distance matrix ..... 1411
  - latitude and longitude coordinates ..... 1410
  - marginal effects ..... 1421–1423
  - ML estimation ..... 1420
  - Moran  $I$  test ..... 1414
  - normalized weighting matrix .. 1412
  - OLS with spatial errors .. 1415–1416
  - overview of spatial models ..... 1406–1407
  - panel data ..... 1428–1429
  - planar coordinates ..... 1410
  - prediction ..... 1421–1423
  - SAR(1) in error model .. 1424–1425
  - SAR(1) in mean model .. 1418–1423
  - SARAR(1,1) model ..... 1425–1426
  - SARAR( $p, q$ ) model ..... 1426
  - spatial autoregressive models ..... 1406–1407, 1417–1429
  - spatial correlation test ..... 1413
  - spatial dependence in error ..... 1414–1416, 1424–1425
  - spatial HAC standard errors ..... 1415–1416
  - spatial lag variables ..... 1413
  - spatial spillovers ..... 1423
  - spatial weighting matrix ..... 1406, 1411–1413
  - total impact ..... 1422
- spdistance** command ..... 1411
- specification analysis ..... 123–132
- specification tests ... 132–140, 529–531, 632
- binary outcome models ..... 883
  - chi-squared goodness-of-fit test ..... 530
  - classification test ..... 872
  - first-stage  $F$  statistic ..... 341–351
  - for fixed effects ..... 409–411
- specification tests, *continued*
- for serially uncorrelated panel errors ..... 435
  - for tobit model ..... 967–970
  - goodness-of-fit test ..... 870
  - Hausman test ..... 409–411, 531, 553–554
  - information matrix test ..... 530
  - moment-based tests ..... 529
  - of endogeneity .. 322–325, 553–554, 1083
  - of heteroskedasticity ..... 136, 253, 695
  - of omitted variables ..... 133
  - of overdispersion ..... 1032
  - of overidentifying restrictions ..... 325–327, 436, 531, 1250
  - of underidentification ..... 341
  - of weak instruments ..... 341–351
  - panel cointegration ..... 455–456
  - panel unit roots ..... 455–456
  - RESET test ..... 134
  - robust tests ..... 361
  - test of heteroskedasticity ..... 136
- spgenerate** command ..... 1413
- spivregress** command ..... 1427
- spline regression .. *see* regression splines
- splitsample** command ..... 1470
- spmatrix** command ..... 1411
- spregress** command ..... 1417
- spset** command ..... 1408
- spshape2dta** command ..... 39, 1408
- spxtregress** command ..... 1428
- sqreg** command ..... 687
- sqrtlasso** command ..... 1482
- st\_addvar()** Mata function .. 737, 1587
- st\_data()** Mata function ..... 731
- st\_matrix()** Mata function .. 146, 731, 737
- st\_store()** Mata function .. 737, 1587
- st\_view()** Mata function ..... 145, 731
- stable unit-treatment value assumption ..... 1291
- standard error estimation ..... *see* variance–covariance matrix

- standardize regressors ..... 1478
  - for principal components..... 1494
- standardized regression ..... 148
- Stata documentation ..... 2–5
- Stata Journal*..... 3
- Stata Technical Bulletin* ..... 3
- `stata()` Mata function ..... 728
- `statsby` prefix ..... 453
- `stcox` command..... 1110
- `stcoxkm` command ..... 1114
- `stcurve` command ..... 1113
- `stdescribe` command..... 1103
- `stepwise` command ..... 492
- `stjoin` command..... 1131
- stochastic production frontier models..
  - ..... 1172
- `stofreal()` Mata function ..... 1587
- `stphplot` command..... 1114
- `streg` command..... 1119
- string data ..... 36, 38
- structural equation models.. 1231–1261
  - cluster-robust standard errors...
    - ..... 1250
  - estimation .. 1233–1234, 1249–1253
  - examples..... 1231, 1239, 1253
  - generalized SEM..... 1251–1261
  - introduction..... 1232–1234
  - linear IV ..... 1247–1249
  - linear regression example.....
    - ..... 1234–1237
  - linear SEM ..... 1232–1251
  - measurement-error example.....
    - ..... 1240–1247
  - model specification..... 1237
  - multiple response FMM example ..
    - ..... 1259–1261
  - path diagram ..... 1243, 1245
  - Poisson example..... 1254–1259
  - robust standard errors ..... 1250
  - test of overidentifying restrictions
    - ..... 1250
  - with endogeneity ..... 1256–1259
- structural model
  - for endogenous Poisson.....
    - ..... 1079–1084
  - for endogenous probit .... 888–893
- `sts graph` command ..... 1107
- `sts list` command..... 1105
- `stset` command..... 1102
- `stsplit` command ..... 1129
- `stsum` command..... 1103
- `stvary` command..... 1103
- subsampling... see bootstrap methods
- `suest` command..... 190, 294
- `summarize` command ..... 6, 20, 87
- summary statistics.. see data summary
  - examples
- `sumcluster` package..... 264
- supervised learning ..... 1497–1501
- support vector machines ..... 1501
- SUR model .... see seemingly unrelated
  - regression equations
- `sureg` command..... 288
- survey data..... 140–145, 295–301
  - clustering..... 295
  - commands for survey data .... 297
  - complex survey data ..... 295
  - stratification ..... 295
  - weighted mean ..... 141, 299
  - weighted regression..... 142, 300
  - weighting ..... 295
- survival data..... see duration models
- `svmachines` command..... 1501
- `svy` prefix ..... 295
- `svy: regress` command..... 300
- `svydescribe` command..... 298
- `svymean` command..... 299
- `svyset` command..... 297
- `syntax`..... 5–6
  - basic command syntax ..... 5
  - parsing syntax ..... 719
- `syntax` command..... 720
- `synth` command..... 1366
- synthetic control..... 1365–1369, 1402
- `synthrunner` command..... 1367

- systems of equations
  - linear regressions . . . . . 286–295
  - linear simultaneous equations . . . . . 367–368
  - nonlinear regressions . . . . . 945
- sysuse** command . . . . . 7
- T**
- tab2** command . . . . . 90
- table** command . . . . . 89, 110
- tables of frequencies . . . . . 89–91
- tables of regression output . . . . . 107
  - in Word or L<sup>A</sup>T<sub>E</sub>X . . . . . 147
- tables of summary statistics . . . . . 91–92
- tabstat** command . . . . . 92
- tabulate** command . . . . . 48, 55, 87
- tebalance** command . . . . . 1295, 1317
- teffects**
  - aipw** command . . . 1296, 1318, 1329
  - commands summary . . . 1302–1304
  - ipw** command . . . . . 1294, 1315
  - ipwra** command . . . . . 1320
  - nnmatch** command . . . . . 1299, 1322
  - psmatch** command . . . . . 1298, 1320
  - ra** command . . . . . 1293, 1313, 1326
- telasso** command . . . . . 1522
- tempvar** command . . . . . 718
- teoverlap** command . . . . . 1294, 1315
- test** command . . . . . 106, 487
- test of overidentifying restrictions .. see specification tests
- test sample . . . . . 1471, 1501
- testnl** command . . . . . 492
- testparm** command . . . . . 487
- tests . . . . . see hypothesis tests
- text data . . . . . 34
- three-stage least squares . . . . . 367–368
- time-series data . . . . . 59–60
  - within panel . . . . . 391
- time-series operators . . . . . 391
- tnbreg** command . . . . . 1043
- tobcm** command . . . . . 969
- tobit** command . . . . . 952
  - tobit models . . . . . 949–1020
    - censored data . . . . . 950
    - censored least-absolute deviations . . . . . 960
    - clustered data . . . . . 961
    - density . . . . . 951
    - diagnostic tests . . . . . 966
    - endogenous regressor . . . . . 989–990
    - endogenous-switching regression model . . . . . 990
    - example . . . . . 953–960
    - FMM application . . . . . 1215–1217
    - for lognormal data . . . . . 961–970
    - generalized residuals . . . . . 966
    - interval regression . . . . . 960, 1172
    - marginal effects . . . . . 956–959
    - maximum likelihood estimation . . . . . 951
    - panel-data estimators . . 1167–1172
    - partial observability . . . . . 950
    - prediction . . . . . 955
    - prediction from log model . . . . . 986–989
    - random-effects model . . . . . 1168
    - selection model . . . . . 974–986
    - test of heteroskedasticity . . . . . 970
    - test of normality . . . . . 969
    - tobit model . . . . . 950–952
    - truncated data . . . . . 959
    - truncated mean . . . . . 950
    - two-limit tobit . . . . . 965
    - two-part model . . . . . 970–974
    - unknown censoring point . . . . . 952
- tokens()** Mata function . . . . . 145
- tostring** command . . . . . 36
- tpoisson** command . . . . . 1043
- trace** command . . . . . 724
- training sample . . . . . 1471, 1501
- treatment effects . . . . . 1269–1403
  - assumptions . . . . . 1290–1292
  - ATE . . . . . 1271, 1291
  - ATET . . . . . 1272, 1291
  - augmented IPW . . . . . 1296, 1318, 1329–1332, 1522
  - balancing . . . . . 1295



- treatment effects, *continued*
- blocking ..... 1321
  - complier ..... 1357
  - conditional independence assumption ..... 1290
  - conditional mean assumption ..... 1291
  - counterfactual ..... 1271
  - defier ..... 1357
  - difference in differences .. 195–204, 1363–1364
  - doubly robust methods ..... 1295–1296
  - endogenous treatment .. 1337–1403
  - endogenous treatment parametric methods ..... 1338–1356
  - exogenous treatment ... 1269–1336
  - heterogeneous effects .. 1292, 1338, 1357
  - in Poisson model ..... 1353–1356
  - inference ..... 1301–1302
  - intention-to-treat effect ..... 1307, 1360
  - inverse-probability weighting ..... 1293–1294, 1315–1318
  - IPW regression adjustment... 1296, 1320
  - lasso for AIPW ..... 1522
  - LATE ..... 1356–1362
  - marginal treatment effects ... 1362
  - margins** command ..... 1341, 1349
  - matching assumption ..... 1291
  - matching methods ..... 1296–1299
  - monotonicity ..... 1359
  - multilevel treatments .. 1323–1334, 1343–1348
  - nearest-neighbor matching... 1299, 1322–1323
  - OHIE example .... 1305–1323, 1360
  - overlap assumption ..... 1291
  - POMs ..... 1293
  - poparms** command ..... 1332, 1334
  - potential outcomes ..... 1271–1272
  - propensity score ..... 1291
- treatment effects, *continued*
- propensity-score matching ..... 1297–1299, 1320–1322
  - propensity-score overlap ..... 1294
  - quantile effects ..... 1332–1334, 1388–1401
  - quantile endogenous effects ..... 1388–1394
  - quantile regression ..... 703–706
  - regression adjustment.. 1292–1293, 1313–1314, 1326–1329
  - regression discontinuity design .... 1369–1388
  - stable unit-treatment value assumption ..... 1291
  - synthetic control ..... 1365–1369
  - unconfoundedness assumption .... 1291, 1300
- treatment–control comparison ..... 199
- triangular system ..... 1342, 1348
- truncated data ... see tobit models, see count-data models
- truncreg** command ..... 959
- tsset** command ..... 59, 209
- ttest** command ..... 93, 120
- compared with **regress** ..... 120
- two-part model
- for count data ..... 1044–1050
  - for log expenditure data .. 970–974
  - prediction ..... 988
- two-sample *t* test ..... 121, 1274
- computed by OLS regression ... 122
- two-sample 2SLS estimator ..... 366
- two-stage least-squares estimator ... see instrumental variables
- two-step estimator
- bootstrap standard error .. 551–552
  - sample-selection model ... 978–981
  - stacked GMM standard error ..... 594–596, 980
- two-way cluster-robust ..... 261–263
- twoway** command ..... 76
- type** command ..... 15

**U**

unconfoundedness . . . . 991, 1291, 1508  
 unobserved heterogeneity . . . 1191–1261  
   summary . . . . . 1192  
 unsupervised learning . . . . . 1501  
   principal components . . . . . 1501  
**update** command . . . . . 2  
**use** command . . . . . 38

**V**

variable names and labels . . . . . 46–47  
 variance–covariance matrix  
   **vce(bootstrap)** option . . . . . 465  
 variance–covariance matrix . . . 101–102,  
   597–604  
   bootstrap estimate . . 101, 117–120,  
     539–544, 604  
   cluster–robust . . . 99, 116, 258–265,  
     377–378, 602, 1144  
   default estimate . . . . . 98, 601  
   design-based . . . . . 1289  
   dyad–robust . . . . . 263  
   for FGLS . . . . . 247–249  
   for IV . . . . . 315  
   for *m*-estimator . . . . . 598  
   for nonlinear estimators . . 597–604  
   for OLS . . . . . 98–102  
   for SUR . . . . . 290  
   for two-step estimator . . . 551–552,  
     594–596, 980  
   HAC estimate . . . . . 603  
   heteroskedastic–robust . . . 99, 601  
   jackknife estimate . . . 575–576, 604  
   outer product of the gradient . 601  
   robust when to use . . . . . 583, 599  
   sandwich form . . . . . 598  
   spatial HAC . . . . . 1415–1416  
   two-way cluster–robust . . 261–263  
   unconditional . . . . . 178, 615, 620,  
     1263, 1343, 1349  
 VCE . . . see variance–covariance matrix  
**vce()** option . . see variance–covariance  
   matrix  
**vce(bootstrap)** option . . 101, 540, 604

**vce(cluster *clustvar*)** option . . . . . 99,  
   602, 620  
**vce(hac)** option . . . . . 603  
**vce(jackknife)** option . . . . . 576, 604  
**vce(oim)** option . . . . . 601  
**vce(opg)** option . . . . . 601  
**vce(robust)** option . . . . . 99, 601  
**vce(unconditional)** option . . 178, 615,  
   620, 640, 1263, 1343, 1349  
**vcemway** command . . . . . 261  
**vselect** command . . . . . 1476

**W**

Wald  
   confidence intervals . . . . . see  
     confidence intervals  
   estimator . . . . . 1357  
   test . . . . . see hypothesis tests  
 weak instruments . . . . . see instrumental  
   variables  
**weakiv** command . . . . . 361  
**weakivtest** command . . . . . 346  
 Weibull model . . . . . 1119  
   FMM application . . . . . 1220–1223  
 weight . . . . . 6  
 weighted mean . . . . . 141  
   survey data . . . . . 299  
 weighted regression . . . . . see  
   inverse-probability weighting  
   survey data . . . . . 300  
**while** command . . . . . 28  
 wide-form data . . . 63, 414–418, 914–915  
 wild gradient bootstrap . . . . . 686  
 wildcards . . . . . 13

**X**

**xi** prefix . . . . . 56  
**x\_ols** command . . . . . 1415  
**xpoivregress** command . . . . . 1521  
**xporegress** command . . . . . 1515  
**xtabond** command . . . . . 429  
**xtabond2** command . . . . . 440  
**xtbalance** command . . . . . 1001  
**xtcloglog** command . . . . . 1148  
**xtcointtest** command . . . . . 455

xtdata command..... 388  
 xtdescribe command.. 278, 383, 1224  
 xtdidregress command..... 202, 1364  
 xtdpd command..... 438  
 xtdpdsys command..... 436  
 xteintreg command..... 1261  
 xteoprobit command..... 1261  
 xteprobit command..... 1261  
 xteregress command..... 1261  
 xtfrontier command..... 1172  
 xtgee command ... 395, 634, 880, 1150  
 xtglsl command..... 448, 450  
 xtheckman command..... 1169  
 xthtaylor command..... 426  
 xtintreg command..... 1172  
 xtivreg command..... 422  
 xtline command..... 386  
 xtlogit command..... 880, 1150  
 xtlogitfe command..... 639, 881  
 xtmg command..... 455  
 xtnbreg command..... 1173  
 xtologit command..... 942, 1164  
 xtoprobit command..... 942  
 xtoverid command..... 411  
 xtpcse command..... 448  
 xtpmg command..... 455  
 xtpoisson command..... 1173  
 xtprobit command..... 1148  
 xtprobitfe command..... 639, 881  
 xtreg commands..... 265, 378  
 xtreg,  
     be command..... 402  
     fe command..... 271, 398  
     mle command..... 403  
     pa command..... 266, 395  
     re command..... 267, 403  
 xtregar command..... 451  
 xtsccl command..... 449, 1429  
 xtset command..... 266, 382  
 xtstreg command..... 1137  
 xtsum command..... 384  
 xttab command..... 385  
 xttobit command..... 1168  
 xttrans command..... 386  
 xtunitroot command..... 455

**Z**

zero-inflated models .... see count-data  
     models  
 zero-truncated models .. see count-data  
     models  
 zinb command..... 1071  
 zioprobit command..... 941  
 zip command..... 1071