Glossary

adaptive lasso. Adaptive lasso is one of three methods that Stata provides for fitting lasso models. The other two methods are cross-validation and plugins. Adaptive lasso tends to include fewer covariates than cross-validation and more covariates than plugins. Adaptive lasso is not used for fitting square-root lasso and elastic-net models.

Bayesian information criterion, BIC. The Bayesian information criterion (BIC), also known as Schwarz criterion, is an information-based criterion used for model selection. It is given by the formula $-2 \times$ log likelihood $+ k \ln N$, where k is the number of parameters in the model and N is the sample size.

beta-min condition. Beta-min condition is a mathematical statement that the smallest nonzero coefficient, in absolute value, be sufficiently large in the true or best approximating model. The condition is seldom met in lasso models because lasso tends to omit covariates with small coefficients. That is not an issue for prediction, but it is for inference. Stata's double selection, partialing out, and cross-fit partialing out work around the issue.

coefficients of interest. See *covariates of interest and control covariates*.

control variables. See *covariates of interest and control covariates*.

covariates. Covariates, also known as explanatory variables and RHS variables, refer to the variables that appear or potentially appear on the right-hand side of a model and that predict the values of the outcome variable. This manual often refers to "potential covariates" and "selected covariates" to distinguish the variables that lasso considers from those it selects for inclusion in the model.

covariates of interest and control covariates. Covariates of interest and control covariates compose the covariates that are specified when fitting lasso models for inference. In these cases, the coefficients and standard errors for the covariates of interest are estimated and reported. The coefficients for the control covariates are not reported nor are they recoverable, but they nonetheless appear in the model to improve the measurement of the coefficients of interest.

Covariates of interest and control covariates are often called variables of interest and control variables.

The coefficients on the covariates of interest are called the coefficients of interest.

covariate selection. Covariate selection refers to processes that automatically select the covariates to be included in a model. Lasso, square-root lasso, and elastic net are three such processes. What makes them special is that they can handle so many potential covariates.

Covariate selection is handled at the same time as estimation. Covariates are included and excluded based on coefficient estimates. When estimates are 0, covariates are excluded.

cross-fitting. Cross-fitting is another term for double machine learning.

cross-validation (CV). Cross-validation (CV) is a method for fitting lasso models. The other methods that Stata provides are adaptive lasso and plugins.

The term in general refers to techniques that validate how well predictive models perform. Classic CV uses one dataset to fit the model and another to evaluate its predictions. When the term is used in connection with lasso, however, CV refers to K-fold CV, a technique that uses the same dataset to fit the model and to produce an estimate of how well the model would do if used to make out-of-sample predictions. See *folds*.

cross-validation function. The cross-validation (CV) function is calculated by first dividing the data into K folds. The model for each λ (and α for elastic net) is fit on data in all but one fold, and then the prediction on that excluded fold is computed and a measure of fit calculated. These K measures of fit are averaged to give the value of the CV function. For linear models, the CV function is the CV mean prediction error. For nonlinear models, the CV function is the CV mean deviance. CV finds the minimum of the CV function, and the value of λ (and α) that gives the minimum is the selected λ^* (and α^*).

cross-validation mean deviance. Cross-validation mean deviance is a cross-validation function that uses the observation-level deviance as a measure of fit.

cross-validation mean deviance ratio. Cross-validation mean deviance ratio is the cross-validation function using the mean of the deviance ratio as the measure of fit.

cross-validation mean prediction error. Cross-validation mean prediction error is the cross-validation function using the mean of the square of the prediction error as the measure of fit. For the linear model, the prediction error is the difference between the individual-level outcome and the linear prediction $\mathbf{x}_i'\boldsymbol{\beta}$.

data-generating process (DGP) and data-generating mechanism (DGM). Data-generating process (DGP) and data-generating mechanism (DGM) are synonyms for the underlying process that generated the data being analyzed. The scientific and statistical models that researchers fit are sometimes approximations of the DGP.

deviance. The deviance is a measure-of-fit statistic for linear and nonlinear likelihood-based models. The deviance for an observation i, D_i, is given by

$$D_i = -2(l_i - l_{\text{saturated}})$$

where l_i is the observation-level likelihood and $l_{\text{saturated}}$ is the value of the saturated likelihood.

deviance null. The deviance null is the mean of the deviance evaluated for the log likelihood of a model that only includes a constant.

deviance ratio. The deviance ratio is a measure-of-fit statistic for linear and nonlinear likelihood-based models. It is given by $D2$

$$D2 = 1 - \overline{D}/(D_{\text{null}})$$

where $\overline{D}$ is the mean of the deviance and D_{null} is the deviance null.

double machine learning (DML). Double machine learning (DML) is a method for estimating the coefficients of interest and their standard errors. When lasso is used for inference, you specify the covariates of interest and the potential control covariates. DML is a family of techniques that combine sample splitting and robust moment conditions. See *double selection, partialing out, and cross-fit partialing out*.

double selection, partialing out, and cross-fit partialing out. Double selection, partialing out, and cross-fit partialing out are three different estimation techniques for performing inference on a subset of the coefficients in a lasso model. Stata provides these techniques for linear, logit, probit, Poisson, and instrumental-variables models. Cross-fit partialing out is also known as double machine learning (DML). Also see [LASSO] **Lasso inference intro**.

ds. A shorthand that we use in this manual to refer to all the double-selection inference commands—`dsregress`, `dslogit`, and `dspoisson`.

elastic net. Elastic net is a penalized estimator designed to be less likely than lasso to exclude highly collinear covariates. Stata's `elasticnet` command fits elastic-net models using cross-validation.

excluded covariates. See *covariate selection*.

folds and K-fold cross-validation. Folds and K-fold cross-validation refer to a technique for estimating how well a model would perform in out-of-sample prediction without actually having a second dataset. The same data that were used to fit the model are then divided into K approximately equal-sized, mutually exclusive subsamples called folds. For each fold k, the model is refit on the data in the other $K-1$ folds, and that result is then used to make predictions for the values in fold k. When the process is complete for all K folds, the predictions in the combined folds are compared with actual values. The number of folds, K, is usually set to 10.

included covariates. See *covariate selection*.

inference. Inference means statistical inference or scientific inference. It involves using samples of data to infer the values of parameters in the underlying population along with measures of their likely accuracy. The likely accuracy is stated in terms of probabilities, confidence intervals, credence intervals, standard errors, and other statistical measures.

Inference can also refer to scientific inference. Scientific inference is statistical inference on a causal parameter. These parameters characterize cause-and-effect relationships. Does more education cause higher incomes, or is it simply a proxy that is associated with higher incomes because those who have it are judged to be smarter or have more drive to succeed or simply spent more time with the right people? If the interest were in simply making statistical predictions, it would not matter.

in-sample R^2. The in-sample R^2 is the R^2 evaluated at the sample where the model is fit.

knots. Knots are the values of the penalty parameters at which variables in the model change.

lambda and alpha. Lambda and alpha (λ and α) are lasso's and elastic-net's penalty parameters.

Lambda is lasso's and square-root lasso's penalty parameter. Lambda is greater than or equal to 0. When it is 0, all possible covariates are included in the model. At its largest value (which is model dependent), no covariates are included. Thus, lambda orders the models.

Alpha is elastic-net's penalty parameter. Alpha is bounded by 0 and 1, inclusive. When alpha is 0, the elastic net becomes ridge regression. When alpha is 1, the elastic net becomes lasso.

lasso. Lasso has different meanings in this glossary, depending on usage.

First, we use lasso to mean lasso, the word that started as LASSO because it was an acronym for "least absolute shrinkage and selection operator".

Second, we use lasso to mean lasso and square-root lasso, which are two different types of lasso. See *square-root lasso*.

Third, we use lasso to mean lasso, square-root lasso, and elastic net. Elastic net is yet another type of lasso that uses a different penalty function. See *elastic net*.

Lasso in the broadest sense is widely used for prediction and covariate selection.

Lasso in the narrowest sense is implemented by Stata's `lasso` command. It fits linear, logit, probit, Poisson, and Cox models. It fits them using any of four methods: cross-validation, adaptive lasso, plugins (not available for Cox models), and the Bayesian information criterion function.

Square-root lasso is implemented by Stata's `sqrtlasso` command. It fits linear models using cross-validation or plugins.

Elastic net is implemented by Stata's `elasticnet` command. It fits linear, logit, probit, Poisson, and Cox models. It uses cross-validation.

Regardless of the particular lasso used, these methods estimate coefficients on potential covariates. Covariates are included and excluded based on the estimate. When estimates are 0, covariates are excluded.

lasso selection. See *covariate selection*.

nonzero coefficients. Nonzero coefficients are the coefficients estimated for the selected covariates.

not-selected covariates. Not-selected covariates is a synonym for excluded covariates; see *covariate selection*.

outcome variable. Outcome variable, also known as dependent variable and LHS variable, refers to the variable whose values are predicted by the independent variables, which are also known as covariates and RHS variables.

out-of-sample R^2. The out-of-sample R^2 is the R^2 evaluated for a sample distinct from the one for which the model was fit.

penalized coefficients. Penalized coefficients are the coefficient estimates produced by lasso when the covariates are not standardized to have a mean of 0 and standard deviation of 1.

penalized estimators. Penalized estimators are statistical estimators that minimize a measure of fit that includes a penalty term. That term penalizes models based on their complexity. Lasso, square-root lasso, and elastic net are penalized estimators.

What distinguishes lasso from elastic net, and is the only thing that distinguishes them, is the particular form of the penalty term. Lasso uses the sum of the absolute values of the coefficients for the included covariates. Elastic net uses the same penalty term plus the sum of the squared coefficients. The additional term is designed to prevent exclusion of highly collinear covariates.

Square-root lasso uses the same penalty term as lasso, but the form of the objective function to which the penalty is added differs.

penalty loadings. Penalty loadings refer to coefficient-specific penalty weights in adaptive lasso and plugins. Allowing coefficients to have different penalty weights improves the model chosen by lasso, square-root lasso, and elastic net.

penalty parameter. Penalty parameter is the formal term for lambda (λ), lasso's and square-root lasso's penalty parameter, and alpha (α), elastic-net's penalty parameter. See *lambda and alpha*.

plugins. Plugins are the method for fitting lasso and square-root lasso models, but not elastic-net models. It is an alternative to cross-validation. Cross-validation tends to include more covariates than are justified, at least in comparison with the best approximating model. Plugins were developed to address this problem. Plugins have the added advantage of being quicker to execute, but they will sometimes miss important covariates that cross-validation will find.

po. A shorthand that we use in this manual to refer to all the partialing-out inference commands—`poregress`, `pologit`, `popoisson`, and `poivregress`.

postlasso coefficients. Postlasso coefficients, also known as postselection coefficients, are the estimated coefficients you would obtain if you refit the model selected by lasso. To be clear about it, you fit a linear model by using lasso. It selected covariates. You then refit the model on those covariates by using `regress`, `logit`, etc. Those are the postselection coefficients, and they will differ from those produced by lasso. They will differ because lasso is a shrinkage estimator, and that leads to the question: which are better for prediction?

There is no definitive answer to that question. The best answer we can give you is to use split samples and `lassogof` to evaluate both sets of predictions and choose the better one.

For your information, Stata's lasso commands—`lasso`, `sqrtlasso`, and `elasticnet`—provide both the lasso and the postselection coefficients. The lasso coefficients are stored in `e(b)`. The postselection coefficients are stored in `e(b_postselection)`. You can do in-sample and out-of-sample prediction with `predict`. `predict` by default uses the lasso coefficients. Specify option `postselection`, and it uses the postselection coefficients.

potential covariates. See *covariates*.

prediction and predictive modeling. Prediction and predictive modeling refer to predicting values of the outcome variable based on covariates. Prediction is what lasso was originally designed to do. The variables on which the predictions are based do not necessarily have a cause-and-effect relationship with the outcome. They might be proxies for the cause and effects. Also see *inference*.

regularized estimator. Regularized estimators is another term used for penalized estimators. See *penalized estimators*.

R^2. R^2 is a measure of goodness of fit. It tells you what fraction of the variance of the outcome is explained by your model.

sample splitting. Sample splitting is a way of creating two or more smaller datasets from one dataset. Observations are randomly assigned to subsamples. Stata's `splitsample` command does this. Samples are sometimes split to use the resulting subsamples in different ways. One could use the first subsample to fit the model and the second subsample to evaluate its predictions.

saturated likelihood. The saturated likelihood is the likelihood for a model that has as many estimated parameters as data points.

selected covariates. Selected covariates is synonym for included covariates; see *covariate selection*.

sparsity assumption. Sparsity assumption refers to a requirement for lasso to produce reliable results. That requirement is that the true model that lasso seeks has few variables, where "few" is measured relative to the number of observations in the dataset used to fit the model.

square-root lasso. Square-root lasso is a variation on lasso. Development of square-root lassos was motivated by a desire to better fit linear models with homoskedastic but not normal errors, but it can also be used with heteroskedastic errors. Stata's `sqrtlasso` command fits square-root lassos.

standardized coefficients. Standardized coefficients are the coefficient estimates produced by lasso when the covariates are standardized to have a mean of 0 and standard deviation of 1.

variable selection. See *covariate selection*.

variables of interest. See *covariates of interest and control covariates*.

xpo. A shorthand that we use in this manual to refer to all the cross-fit partialing-out inference commands—`xporegress`, `xpologit`, `xpopoisson`, and `xpoivregress`.