

lasso postestimation — Postestimation tools for lasso for prediction

[Postestimation commands](#)
[predict](#)
[Remarks and examples](#)
[References](#)
[Also see](#)

Postestimation commands

The following postestimation commands are of special interest after `lasso`, `sqrtlasso`, and `elasticnet`:

Command	Description
<code>bicplot</code>	plot Bayesian information criterion function
<code>coefpath</code>	plot path of coefficients
<code>cvplot</code>	plot cross-validation function
<code>lassocoef</code>	display selected coefficients
<code>lassogof</code>	goodness of fit after lasso for prediction
<code>lassoinfo</code>	information about lasso estimation results
<code>lassoknots</code>	knot table of coefficient selection and measures of fit
<code>lassoselect</code>	select alternative λ^* (and α^* for <code>elasticnet</code>)

The following standard postestimation commands are also available:

Command	Description
<code>estat summarize</code>	summary statistics for the estimation sample
<code>estimates</code>	cataloging estimation results
<code>predict</code>	linear predictions

predict

Description for predict

`predict` creates a new variable containing predictions such as linear predictions, probabilities when the model is `logit` or `probit`, or number of events when the model is `Poisson`.

Menu for predict

Statistics > Postestimation

Syntax for predict

```
predict [type] newvar [if] [in] [, statistic options]
```

<i>statistic</i>	Description
Main	
<code>xb</code>	linear predictions; the default for the <code>linear</code> model
<code>pr</code>	probability of a positive outcome; the default for the <code>logit</code> and <code>probit</code> models
<code>n</code>	number of events; the default for the <code>poisson</code> model
<code>ir</code>	incidence rate; optional for the <code>poisson</code> model

Option `pr` is not allowed when the model is `linear` or `poisson`.

Options `n` and `ir` are allowed only when the model is `poisson`.

<i>options</i>	Description
Main	
<code>penalized</code>	use penalized coefficients; the default
<code>postselection</code>	use postselection (unpenalized) coefficients
<code>nooffset</code>	ignore the offset or exposure variable (if any)

Options for predict

Main

`xb`, the default for the `linear` model, calculates linear predictions.

`pr`, the default for the `logit` and `probit` models, calculates the probability of a positive event.

`n`, the default for the `poisson` model, calculates the number of events.

`ir` applies to `poisson` models only. It calculates the incidence rate $\exp(\mathbf{x}\beta')$, which is the predicted number of events when exposure is 1. Specifying `ir` is equivalent to specifying `n` when neither `offset()` nor `exposure()` was specified when the model was fit.

`penalized` specifies that penalized coefficients be used to calculate predictions. This is the default. Penalized coefficients are those estimated by lasso in the calculation of the lasso penalty. See [Methods and formulas](#) in `[LASSO] lasso`.

`postselection` specifies that postselection coefficients be used to calculate predictions. Postselection coefficients are calculated by taking the variables selected by lasso and refitting the model with the appropriate ordinary estimator: linear regression for `linear` models, logistic regression for `logit` models, probit regression for `probit` models, and Poisson regression for `poisson` models.

`nooffset` is relevant only if you specified `offset()` or `exposure()` when you fit the model. It modifies the calculations made by `predict` so that they ignore the offset or exposure variable; the linear prediction is treated as $\mathbf{x}\beta'$ rather than $\mathbf{x}\beta' + \text{offset}$ or $\mathbf{x}\beta' + \ln(\text{exposure})$. For the poisson model, specifying `predict ...`, `nooffset` is equivalent to specifying `predict ..., ir`.

Remarks and examples

[stata.com](https://www.stata.com)

By default, `predict` after `lasso` uses the penalized coefficient estimates to predict the outcome. Specifying the option `postselection` causes `predict` to use the postselection coefficients to calculate predictions. Postselection coefficients are calculated by taking the variables selected by lasso and refitting the model with the unpenalized estimator.

Belloni and Chernozhukov (2013) and Belloni et al. (2012) provide results under which predictions using postselection coefficients perform at least well as predictions using penalized coefficients. Their results are only for linear models. Their conditions essentially limit the cases to ones in which the covariates selected by the lasso are close to the set of covariates that best approximates the outcome. Said plainly, this means that under the conditions for which lasso provides valid predictions, the postselection coefficients should do slightly better than the penalized coefficients in most cases, and in other cases they should be about the same.

Rather than relying on theorems, standard practice in prediction applications uses split-sample techniques to find which of several models produces the best predictions. One standard practice in prediction applications is to randomly split the sample into training and testing samples. Using the training data, the coefficients for several competing predictors are computed. Using the testing data, an out-of-sample prediction error is computed for each of the predictors whose coefficients were estimated on the training data. The predictor with smallest out-of-sample prediction error is preferred. This practice is illustrated in [LASSO] [lassogof](#).

References

- Belloni, A., D. Chen, V. Chernozhukov, and C. B. Hansen. 2012. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80: 2369–2429. <https://doi.org/10.3982/ECTA9626>.
- Belloni, A., and V. Chernozhukov. 2013. Least squares after model selection in high-dimensional sparse models. *Bernoulli* 19: 521–547. <https://doi.org/10.3150/11-BEJ410>.

Also see

- [LASSO] [lasso examples](#) — Examples of lasso for prediction
- [LASSO] [elasticnet](#) — Elastic net for prediction and model selection
- [LASSO] [lasso](#) — Lasso for prediction and model selection
- [LASSO] [sqrtlasso](#) — Square-root lasso for prediction and model selection
- [U] [20 Estimation and postestimation commands](#)