cvplot — Plot cross-validation function after lasso

Description

cvplot graphs the cross-validation (CV) function after a lasso fit using selection(cv), selection(adapt
selection(bic), or selection(none).

cvplot can be used after lasso, elasticnet, sqrtlasso, telasso, or any of the lasso inference
commands.

Quick start

Graph the CV function after lasso, sqrtlasso, or elasticnet

    cvplot

Same as above, and draw a reference line identifying the value selected by the one-standard-error rule

    cvplot, seline

Graph the CV function after elasticnet for the $\alpha = 0.5$ lasso

    cvplot, alpha(.5)

After any of the ds or po commands, graph the CV function for the dependent variable y

    cvplot, for(y)

Same as above, and show standard error bands for the CV function

    cvplot, for(y) se

After an xpo command without resample, graph the CV function for x in cross-fit fold 2

    cvplot, for(x) xfold(2)

After an xpo command with resample, graph the CV function for x in cross-fit fold 2 for the first resample

    cvplot, for(x) xfold(2) resample(1)

Same as above, but graph the CV function as a function of the ℓ_1-norm of the standardized coefficient
vector

    cvplot, for(x) xfold(2) resample(1) xunits(l1norm)

After telasso, graph the CV function for the outcome variable x at treatment level 1

    cvplot, for(y) tlevel(1)

Menu

Statistics > Postestimation

1

Syntax

After lasso, sqrtlasso, *and* elasticnet

    cvplot [, *options*]

After ds *and* po *commands*

    cvplot, for(*varspec*) [*options*]

After xpo *commands without* resample

    cvplot, for(*varspec*) xfold(#) [*options*]

After xpo *commands with* resample

    cvplot, for(*varspec*) xfold(#) resample(#) [*options*]

After telasso *for the outcome variable*

    cvplot, for(*varspec*) tlevel(#) [*options*]

After telasso *for the treatment variable*

    cvplot, for(*varspec*) [*options*]

After telasso *for the outcome variable with cross-fitting but without* resample

    cvplot, for(*varspec*) tlevel(#) xfold(#) [*options*]

After telasso *for the treatment variable with cross-fitting but without* resample

    cvplot, for(*varspec*) xfold(#) [*options*]

After telasso *for the outcome variable with cross-fitting and* resample

    cvplot, for(*varspec*) tlevel(#) xfold(#) resample(#) [*options*]

After telasso *for the treatment variable with cross-fitting and* resample

    cvplot, for(*varspec*) xfold(#) resample(#) [*options*]

varspec is *varname*, except after poivregress and xpoivregress, when it is either *varname* or
    pred(*varname*).

options	Description
Main	
<u>xunits</u>(*x_unit_spec*)	*x*-axis units (scale); default is xunits(rlnlambda), where rlnlambda denotes λ on a reverse logarithmic scale
minmax	add labels for the minimum and maximum *x*-axis units
* for(*varspec*)	lasso for *varspec*; telasso, ds, po, and xpo commands only
* xfold(*#*)	lasso for the #th cross-fit fold; xpo commands and telasso with xfolds only
* resample(*#*)	lasso for the #th resample; xpo commands and telasso with resample only
* tlevel(*#*)	lasso for the outcome model with the treatment level #; telasso only
alpha(*#*)	graph CV function for $\alpha = \#$; default is the selected value α^*; allowed after elasticnet only
<u>lineopts</u>(*cline_options*)	affect rendition of the plotted lines
S.E. plot	
se	show standard error bands for the CV function
<u>seopts</u>(*rcap_options*)	affect rendition of the standard error bands
Reference lines	
<u>cvlineopts</u>(*cline_options*)	affect rendition of reference line identifying the minimum of the CV function or other stopping rule
<u>nocvline</u>	suppress reference line identifying the minimum of the CV function or other stopping rule
<u>lslineopts</u>(*cline_options*)	affect rendition of reference line identifying the value selected using lassoselect
<u>nolsline</u>	suppress reference line identifying the value selected using lassoselect
<u>selineopts</u>(*cline_options*)	affect rendition of reference line identifying the value selected by the one-standard-error rule
[no]<u>seline</u>	draw or suppress reference line identifying the value selected by the one-standard-error rule; shown by default for selection(cv, serule)
<u>hrefline</u>	add horizontal reference lines that intersect the vertical reference lines
<u>rlabelopts</u>(*r_label_opts*)	change look of labels for reference line
Data	
data(*filename* [, replace])	save plot data to *filename*
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than by() documented in [G-3] *twoway_options*

*for(*varspec*) is required for all ds, po, and xpo commands and for telasso.

xfold(*#*) is required for all xpo commands and for telasso when the option xfolds(*#*) was specified.

resample(*#*) is required for xpo and for telasso when the option resample(*#*) was specified.

tlevel(*#*) is required for the outcome model in telasso.

x_unit_spec	Description
rlnlambda	λ on a reverse logarithmic scale; the default
lnlambda	λ on a logarithmic scale
l1norm	ℓ_1-norm of standardized coefficient vector
l1normraw	ℓ_1-norm of unstandardized coefficient vector

r_label_opts	Description
labgap(*size*)	margin between tick and label
labstyle(*textstyle*)	overall style of label
labsize(*textsizestyle*)	size of label
labcolor(*colorstyle*)	color and opacity of label

Options

> ### Main

xunits(*x_unit_spec*) specifies the x-axis units used for graphing the CV function. The following *x_unit_spec*s are available:

> rlnlambda specifies x-axis units λ on a reverse logarithmic scale. This is the default.

> lnlambda specifies x-axis units λ on a logarithmic scale.

> l1norm specifies x-axis units ℓ_1-norm of the standardized coefficient vector.

> l1normraw specifies x-axis units ℓ_1-norm of the unstandardized coefficient vector.

minmax adds labels for the minimum and maximum x-axis units to the graph of the CV function.

for(*varspec*) specifies a particular lasso after telasso or after a ds, po, or xpo estimation command fit using the option selection(cv), selection(adaptive), or selection(bic). For all commands except poivregress and xpoivregress, *varspec* is always *varname*.

> For the ds, po, and xpo commands except poivregress and xpoivregress, *varspec* is either *depvar*, the dependent variable, or one of *varsofinterest* for which inference is done.

> For poivregress and xpoivregress, *varspec* is either *varname* or pred(*varname*). The lasso for *depvar* is specified with its *varname*. Each of the endogenous variables have two lassos, specified by *varname* and pred(*varname*). The exogenous variables of interest each have only one lasso, and it is specified by pred(*varname*).

> For telasso, *varspec* is either the outcome variable or the treatment variable.

> This option is required after telasso and after the ds, po, and xpo commands.

xfold(#) specifies a particular lasso after an xpo estimation command or after telasso when the option xfolds(#) was specified. For each variable to be fit with a lasso, K lassos are done, one for each cross-fit fold, where K is the number of folds. This option specifies which fold, where $\# = 1, 2, \ldots, K$. xfold(#) is required after an xpo command and after telasso when the option xfolds(#) was specified.

resample(#) specifies a particular lasso after an xpo estimation command or after telasso fit using the
option resample(#). For each variable to be fit with a lasso, $R \times K$ lassos are done, where R is the
number of resamples and K is the number of cross-fitting folds. This option specifies which resample,
where $\# = 1, 2, \ldots, R$. resample(#), along with xfold(#), is required after an xpo command and
after telasso with resampling.

tlevel(#) specifies the lasso for the outcome variable at the specified treatment level after telasso.
This option is required to refer to the outcome model after telasso.

alpha(#) graphs the CV function for $\alpha = \#$. The default is alpha(α^*), where α^* is the selected α.
alpha(#) may only be specified after elasticnet.

lineopts(*cline_options*) affects the rendition of the plotted line. See [G-3] *cline_options*.

_____ S.E. plot _____

se shows standard error bands for the CV function.

seopts(*rcap_options*) affects the rendition of the standard error bands. See [G-3] *rcap_options*.

_____ Reference lines _____

cvlineopts(*cline_options*) affects the rendition of the reference line identifying the minimum CV
value, the value selected when the stopping tolerance is reached, or the grid-minimum value. See
[G-3] *cline_options*.

nocvline suppresses the reference line identifying the minimum CV value, the value selected when the
stopping tolerance is reached, or the grid-minimum value.

lslineopts(*cline_options*) affects the rendition of the reference line identifying the value selected
using lassoselect. See [G-3] *cline_options*.

nolsline suppresses the reference line identifying the value selected using lassoselect.

selineopts(*cline_options*) affects the rendition of the reference line identifying the value selected by
the one-standard-error rule. See [G-3] *cline_options*.

[no]seline draws or suppresses a reference line identifying the value selected by the one-standard-error
rule. By default, the line is shown when selection(cv, serule) was the selection method for the
lasso. For other selection methods, the line is not shown by default.

hrefline adds horizontal reference lines that intersect the vertical reference lines.

rlabelopts(*r_label_opts*) changes the look of labels for the reference line. The label options
labgap(*relativesize*), labstyle(*textstyle*), labsize(*textsizestyle*), and labcolor(*colorstyle*)
specify details about how the labels are presented. See [G-4] *size*, [G-4] *textstyle*, [G-4] *textsizestyle*,
and [G-4] *colorstyle*.

_____ Data _____

data(*filename* [, replace]) saves the plot data to a Stata data file.

_____ Y axis, X axis, Titles, Legend, Overall _____

twoway_options are any of the options documented in [G-3] *twoway_options*, excluding by(). These
include options for titling the graph (see [G-3] *title_options*) and options for saving the graph to disk
(see [G-3] *saving_option*).

Remarks and examples

CV plots graph the CV function over the search grid for the lasso penalty parameter λ. For linear models, the CV function is the mean squared error of the predictions in the CV samples. For logit and Poisson models, the CV function is the mean deviance in the CV samples.

The search grid can be shown as the log of the lasso penalty parameter λ, xunits(lnlambda); the reverse of that scale, xunits(rlnlambda); the ℓ_1-norm of the standardized coefficients, xunits(l1norm); or the ℓ_1-norm of the unstandardized coefficients, xunits(l1normraw). The reverse log of lambda is the default because it represents the CV search path over λ, with the first λ tried on the left and the last λ tried on the right.

CV plots can be drawn after any command that directly searches over a grid of λ's—that is, after any command that used the option selection(cv), selection(adaptive), or selection(none). They can be drawn after commands lasso, elasticnet, sqrtlasso, telasso, or any of the 11 lasso inference commands.

Examples that demonstrate how to use cvplot after the lasso command can be found in *The CV function* in [LASSO] **lasso**.

Examples after elasticnet can be found starting in example 2 of [LASSO] **elasticnet**.

Also see

[LASSO] **lasso postestimation** — Postestimation tools for lasso for prediction

[LASSO] **lasso inference postestimation** — Postestimation tools for lasso inferential models

[CAUSAL] **telasso postestimation** — Postestimation tools for telasso