cvplot — Plot cross-validation function after lasso

Description

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

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

Quick start

Graph the CV function after lasso, sqrtlasso, or elasticnet

    cvplot

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)

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)

As above, but graph the CV function as a function of the $\ell_1$-norm of the standardized coefficient vector

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

Menu

Statistics  >  Postestimation
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]

varspec is a varname, except after poivregress and xpoivregress, when it is either a varname or pred(varname).
### cvplot — Plot cross-validation function after lasso

<table>
<thead>
<tr>
<th>Options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main</strong></td>
<td></td>
</tr>
<tr>
<td><code>xunits(x_unit_spec)</code></td>
<td><em>x</em>-axis units (scale); default is <code>xunits(rlnlambda)</code>, where <code>rlnlambda</code> denotes λ on a reverse logarithmic scale</td>
</tr>
<tr>
<td><code>minmax</code></td>
<td>add labels for the minimum and maximum <em>x</em>-axis units</td>
</tr>
<tr>
<td><code>* for(varspec)</code></td>
<td>lasso for <code>varspec</code>; <code>ds</code>, <code>po</code>, and <code>xpo</code> commands only</td>
</tr>
<tr>
<td><code>* xfold(#)</code></td>
<td>lasso for the #th cross-fit fold; <code>xpo</code> commands only</td>
</tr>
<tr>
<td><code>* resample(#)</code></td>
<td>lasso for the #th resample; <code>xpo</code> commands with <code>resample</code> only</td>
</tr>
<tr>
<td><code>alpha(#)</code></td>
<td>graph CV function for α = #; default is the selected value α*; allowed after <code>elasticnet</code> only</td>
</tr>
<tr>
<td><code>lineopts(line_options)</code></td>
<td>affect rendition of the plotted lines</td>
</tr>
<tr>
<td><strong>S.E. plot</strong></td>
<td></td>
</tr>
<tr>
<td><code>se</code></td>
<td>show standard-error bands for the CV function</td>
</tr>
<tr>
<td><code>seopts(rcap_options)</code></td>
<td>affect rendition of the standard-error bands</td>
</tr>
<tr>
<td><strong>Reference lines</strong></td>
<td></td>
</tr>
<tr>
<td><code>cvlineopts(line_options)</code></td>
<td>affect rendition of reference line identifying the minimum of the CV function or other stopping rule</td>
</tr>
<tr>
<td><code>nocvline</code></td>
<td>suppress reference line identifying the minimum of the CV function or other stopping rule</td>
</tr>
<tr>
<td><code>lslineopts(line_options)</code></td>
<td>affect rendition of reference line identifying the value selected using <code>lassoselect</code></td>
</tr>
<tr>
<td><code>nolsline</code></td>
<td>suppress reference line identifying the value selected using <code>lassoselect</code></td>
</tr>
<tr>
<td><code>selineopts(line_options)</code></td>
<td>affect rendition of reference line identifying the value selected by the one-standard-error rule</td>
</tr>
<tr>
<td><code>[no]seline</code></td>
<td>draw or suppress reference line identifying the value selected by the one-standard-error rule; shown by default for <code>selection(cv, serule)</code></td>
</tr>
<tr>
<td><code>hrefline</code></td>
<td>add horizontal reference lines that intersect the vertical reference lines</td>
</tr>
<tr>
<td><code>rlabelopts(r_label_opts)</code></td>
<td>change look of labels for reference line</td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td></td>
</tr>
<tr>
<td><code>data(filename[, replace])</code></td>
<td>save plot data to <code>filename</code></td>
</tr>
</tbody>
</table>

### Options

#### Main

- **xunits(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`; `ds`, `po`, and `xpo` commands only

- **xfold(#)**: 
  - lasso for the #th cross-fit fold; `xpo` commands only

- **resample(#)**: 
  - lasso for the #th resample; `xpo` commands with `resample` only

- **alpha(#)**: 
  - graph CV function for α = #; default is the selected value α*; allowed after `elasticnet` only

- **lineopts(line_options)**: 
  - affect rendition of the plotted lines

#### S.E. plot

- **se**: 
  - show standard-error bands for the CV function

- **seopts(rcap_options)**: 
  - affect rendition of the standard-error bands

#### Reference lines

- **cvlineopts(line_options)**: 
  - affect rendition of reference line identifying the minimum of the CV function or other stopping rule

- **nocvline**: 
  - suppress reference line identifying the minimum of the CV function or other stopping rule

- **lslineopts(line_options)**: 
  - affect rendition of reference line identifying the value selected using `lassoselect`

- **nolsline**: 
  - suppress reference line identifying the value selected using `lassoselect`

- **selineopts(line_options)**: 
  - affect rendition of reference line identifying the value selected by the one-standard-error rule

- **[no]seline**: 
  - draw or suppress reference line identifying the value selected by the one-standard-error rule; shown by default for `selection(cv, serule)`

- **hrefline**: 
  - add horizontal reference lines that intersect the vertical reference lines

- **rlabelopts(r_label_opts)**: 
  - change look of labels for reference line

#### Data

- **data(filename[, replace])**: 
  - save plot data to `filename`

### Options

#### 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. `xfold(#)` is required for all `xpo` commands. `resample(#)` is required for `xpo` when the option `resample(#)` was specified.

### x_unit_spec

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>rlnlambda</code></td>
<td>λ on a reverse logarithmic scale; the default</td>
</tr>
<tr>
<td><code>lnlambda</code></td>
<td>λ on a logarithmic scale</td>
</tr>
<tr>
<td><code>l1norm</code></td>
<td>ℓ₁-norm of standardized coefficient vector</td>
</tr>
<tr>
<td><code>l1normraw</code></td>
<td>ℓ₁-norm of unstandardized coefficient vector</td>
</tr>
<tr>
<td>Description</td>
<td>r_label_opts</td>
</tr>
<tr>
<td>------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>margin between tick and label</td>
<td>labgap(size)</td>
</tr>
<tr>
<td>overall style of label</td>
<td>labstyle(textstyle)</td>
</tr>
<tr>
<td>size of label</td>
<td>labsize(textsize)</td>
</tr>
<tr>
<td>color and opacity of label</td>
<td>labcolor(colorstyle)</td>
</tr>
</tbody>
</table>

**Options**

- **Main**

  - `xunits(x_unit_spec)` specifies the x-axis units used for graphing the CV function. The following `x_unit_specs` are available:
    - `rlnlambda` specifies x-axis units $\lambda$ on a reverse logarithmic scale. This is the default.
    - `lnlambda` specifies x-axis units $\lambda$ on a logarithmic scale.
    - `l1norm` specifies x-axis units $\ell_1$-norm of the standardized coefficient vector.
    - `l1normraw` specifies x-axis units $\ell_1$-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 a `ds`, a `po`, or an `xpo` estimation command fit using the option `selection(cv)` or `selection(adaptive)`. For all commands except `poivregress` and `xpoivregress`, `varspec` is always a `varname`; it 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)`.
    - This option is required after `ds`, `po`, and `xpo` commands.
  - `xfold(#)` specifies a particular lasso after an `xpo` estimation command. 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 $# = 1, 2, \ldots, K$. It is required after an `xpo` command.
  - `resample(#)` specifies a particular lasso after an `xpo` estimation command 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 $# = 1, 2, \ldots, R$. This option, along with `xfold(#)`, is required after an `xpo` command with resampling.
  - `alpha(#)` graphs the CV function for $\alpha = #$. The default is `alpha(\alpha^*)`, where $\alpha^*$ is the selected $\alpha$. `alpha(#)` may only be specified after `elasticnet`.
  - `lineopts(cline_options)` affects the rendition of the plotted line. See [G-3] `cline_options`. 


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.

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.

[b]noseline[b] 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.


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

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  stata.com

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 either the log of the lasso penalty parameter λ, xunits(lnlambda); the reverse of that scale, xunits(rlnlambda); the ℓ₁-norm of the standardized coefficients, xunits(l1norm); or the ℓ₁-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 \(\lambda\)'s—that is, after any command that used the option \texttt{selection(cv)}, \texttt{selection(adaptive)}, or \texttt{selection(none)}. They can be drawn after commands \texttt{lasso}, \texttt{elasticnet}, \texttt{sqrtlasso}, or any of the 11 lasso inference commands.

Examples that demonstrate how to use \texttt{cvplot} after the \texttt{lasso} command can be found in \textit{The CV function} in [LASSO] \texttt{lasso}.

Examples after \texttt{elasticnet} can be found starting in example 2 of [LASSO] \texttt{elasticnet}.

Also see

[LASSO] \texttt{lasso postestimation} — Postestimation tools for lasso for prediction

[LASSO] \texttt{lasso inference postestimation} — Postestimation tools for lasso inferential models