Title

> **regress postestimation diagnostic plots —** Postestimation plots for regress

[Description](#)   [rvfplot](#)   [avplot](#)   [avplots](#)   [cprplot](#)   [acprplot](#)
[rvpplot](#)   [lvr2plot](#)   [Methods and formulas](#)   [References](#)   [Also see](#)

Description

The following postestimation commands are of special interest after regress:

Command	Description
rvfplot	residual-versus-fitted plot
avplot	added-variable plot
avplots	all added-variables plots in one image
cprplot	component-plus-residual plot
acprplot	augmented component-plus-residual plot
rvpplot	residual-versus-predictor plot
lvr2plot	leverage-versus-squared-residual plot

These commands are not appropriate with svy estimation results.

For a discussion of the terminology used in this entry, see the *Terminology* section of
Remarks and examples for predict in [R] **regress postestimation**.

rvfplot

Description for rvfplot

rvfplot graphs a residual-versus-fitted plot, a graph of the residuals against the fitted values.

Menu for rvfplot

Statistics > Linear models and related > Regression diagnostics > Residual-versus-fitted plot

Syntax for rvfplot

rvfplot [, *rvfplot_options*]

rvfplot_options	Description
Plot	
marker_options	change look of markers (color, size, etc.)
marker_label_options	add marker labels; change look or position
Add plots	
addplot(*plot*)	add plots to the generated graph
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than by() documented in [G-3] *twoway_options*

1

Options for rvfplot

> [Plot]

marker_options affect the rendition of markers drawn at the plotted points, including their shape, size, color, and outline; see [G-3] *marker_options*.

marker_label_options specify if and how the markers are to be labeled; see [G-3] *marker_label_options*.

> [Add plots]

addplot(*plot*) provides a way to add plots to the generated graph. See [G-3] *addplot_option*.

> [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 for saving the graph to disk (see [G-3] *saving_option*).

Remarks and examples for rvfplot

rvfplot graphs the residuals against the fitted values.

▷ Example 1

Using auto.dta described in [U] **1.2.2 Example datasets**, we will use regress to fit a model of price on weight, mpg, foreign, and the interaction of foreign with mpg. We specify foreign##c.mpg to obtain the interaction of foreign with mpg; see [U] **11.4.3 Factor variables**.

```
. use https://www.stata-press.com/data/r18/auto
(1978 automobile data)
```

```
. regress price weight foreign##c.mpg
```

Source	SS	df	MS		Number of obs	=	74
					F(4, 69)	=	21.22
Model	350319665	4	87579916.3		Prob > F	=	0.0000
Residual	284745731	69	4126749.72		R-squared	=	0.5516
					Adj R-squared	=	0.5256
Total	635065396	73	8699525.97		Root MSE	=	2031.4

price	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
weight	4.613589	.7254961	6.36	0.000	3.166263	6.060914
foreign						
Foreign	11240.33	2751.681	4.08	0.000	5750.878	16729.78
mpg	263.1875	110.7961	2.38	0.020	42.15527	484.2197
foreign#c.mpg						
Foreign	-307.2166	108.5307	-2.83	0.006	-523.7294	-90.70368
_cons	-14449.58	4425.72	-3.26	0.002	-23278.65	-5620.51

Once we have fit a model, we may use any of the regression diagnostics commands. `rvfplot` (read residual-versus-fitted plot) graphs the residuals against the fitted values:

. rvfplot, yline(0)
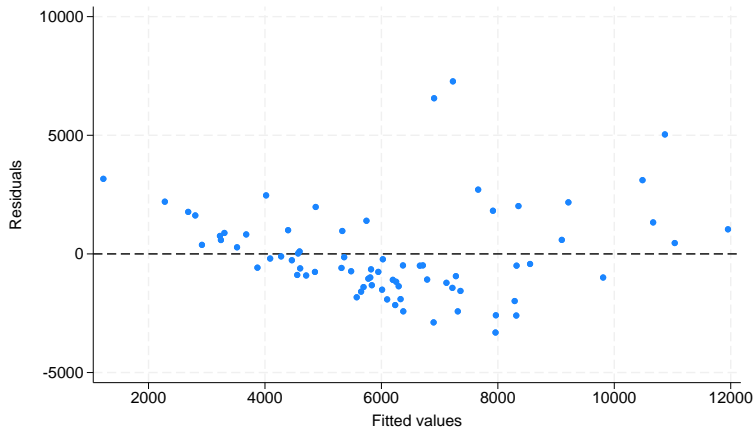
All the diagnostic plot commands allow the `graph twoway` and `graph twoway scatter` options; we specified a `yline(0)` to draw a line across the graph at $y = 0$; see [G-2] **graph twoway scatter**.

In a well-fitted model, there should be no pattern to the residuals plotted against the fitted values—something not true of our model. Ignoring the two outliers at the top center of the graph, we see curvature in the pattern of the residuals, suggesting a violation of the assumption that `price` is linear in our independent variables. We might also have seen increasing or decreasing variation in the residuals—heteroskedasticity. Any pattern whatsoever indicates a violation of the least-squares assumptions.

◁

avplot

Description for avplot

avplot graphs an added-variable plot (a.k.a. partial-regression leverage plot, partial regression plot, or adjusted partial residual plot) after regress. *indepvar* may be an independent variable (a.k.a. predictor, carrier, or covariate) that is currently in the model or not.

Menu for avplot

Statistics > Linear models and related > Regression diagnostics > Added-variable plot

Syntax for avplot

> avplot *indepvar* [, *avplot_options*]

avplot_options	Description
Plot	
marker_options	change look of markers (color, size, etc.)
marker_label_options	add marker labels; change look or position
Reference line	
rlopts(*cline_options*)	affect rendition of the reference line
Add plots	
addplot(*plot*)	add other plots to the generated graph
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than by() documented in [G-3] *twoway_options*

Options for avplot

Plot

marker_options affects the rendition of markers drawn at the plotted points, including their shape, size, color, and outline; see [G-3] *marker_options*.

marker_label_options specify if and how the markers are to be labeled; see [G-3] *marker_label_options*.

Reference line

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

Add plots

addplot(*plot*) provides a way to add plots to the generated graph. See [G-3] *addplot_option*.

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 for saving the graph to disk (see [G-3] *saving_option*).

Remarks and examples for avplot

`avplot` graphs an added-variable plot, also known as the partial-regression leverage plot.

One of the wonderful features of one-regressor regressions (regressions of y on one x) is that we can graph the data and the regression line. There is no easier way to understand the regression than to examine such a graph. Unfortunately, we cannot do this when we have more than one regressor. With two regressors, it is still theoretically possible—the graph must be drawn in three dimensions, but with three or more regressors no graph is possible.

The added-variable plot is an attempt to project multidimensional data back to the two-dimensional world for each of the original regressors. This is, of course, impossible without making some concessions. Call the coordinates on an added-variable plot y and x. The added-variable plot has the following properties:

- There is a one-to-one correspondence between (x_i, y_i) and the ith observation used in the original regression.

- A regression of y on x has the same coefficient and standard error (up to a degree-of-freedom adjustment) as the estimated coefficient and standard error for the regressor in the original regression.

- The "outlierness" of each observation in determining the slope is in some sense preserved.

It is equally important to note the properties that are not listed. The y and x coordinates of the added-variable plot cannot be used to identify functional form, or, at least, not well (see Mallows [1986]). In the construction of the added-variable plot, the relationship between y and x is forced to be linear.
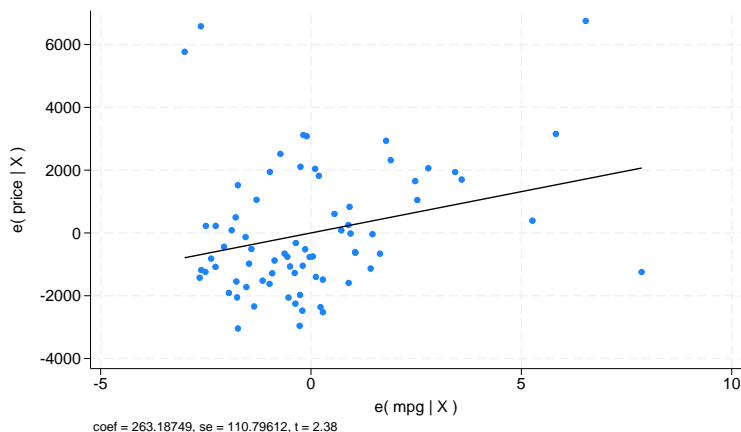
▷ Example 2

Let's use the same model as we used in example 1.

```
. use https://www.stata-press.com/data/r18/auto
(1978 automobile data)
. regress price weight foreign##c.mpg
  (output omitted)
```

We can now examine the added-variable plot for `mpg`.

```
. avplot mpg
```


coef = 263.18749, se = 110.79612, t = 2.38

This graph suggests a problem in determining the coefficient on mpg. Were this a one-regressor regression, the two points at the top-left corner and the one at the top right would cause us concern, and so it does in our more complicated multiple-regressor case. To identify the problem points, we retyped our command, modifying it to read avplot mpg, mlabel(make), and discovered that the two cars at the top left are the Cadillac Eldorado and the Lincoln Versailles; the point at the top right is the Cadillac Seville. These three cars account for 100% of the luxury cars in our data, suggesting that our model is misspecified. By the way, the point at the lower right of the graph, also cause for concern, is the Plymouth Arrow, our data entry error.

◁

❑ Technical note

Stata's avplot command can be used with regressors already in the model, as we just did, or with potential regressors not yet in the model. In either case, avplot will produce the correct graph. The name "added-variable plot" is unfortunate in the case when the variable is already among the list of regressors but is, we think, still preferable to the name "partial-regression leverage plot" assigned by Belsley, Kuh, and Welsch (1980, 30) and more in the spirit of the original use of such plots by Mosteller and Tukey (1977, 271–279). Welsch (1986, 403), however, disagrees: "I am sorry to see that Chatterjee and Hadi [1986] endorse the term 'added-variable plot' when X_j is part of the original model" and goes on to suggest the name "adjusted partial residual plot".

❑

avplots

Description for avplots

avplots graphs all the added-variable plots in one image.

Menu for avplots

Statistics > Linear models and related > Regression diagnostics > Added-variable plot

Syntax for avplots

avplots [, *avplots_options*]

avplots_options	Description
Plot	
marker_options	change look of markers (color, size, etc.)
marker_label_options	add marker labels; change look or position
combine_options	any of the options documented in [G-2] **graph combine**
Reference line	
rlopts(*cline_options*)	affect rendition of the reference line
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than by() documented in [G-3] *twoway_options*

Options for avplots

```
                Plot
```

marker_options affects the rendition of markers drawn at the plotted points, including their shape, size, color, and outline; see [G-3] *marker_options*.

marker_label_options specify if and how the markers are to be labeled; see [G-3] *marker_label_options*.

combine_options are any of the options documented in [G-2] **graph combine**. These include options for titling the graph (see [G-3] *title_options*) and for saving the graph to disk (see [G-3] *saving_option*).

```
                Reference line
```

`rlopts(`*cline_options*`)` affects the rendition of the reference line. See [G-3] *cline_options*.

```
                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 for saving the graph to disk (see [G-3] *saving_option*).

Remarks and examples for avplots

▷ Example 3

In example 2, we used `avplot` to examine the added-variable plot for `mpg` in our regression of `price` on `weight` and `foreign##c.mpg`. Now, let's use `avplots` to graph an added-variable plot for every regressor in the data.

```
. avplots
```


◁
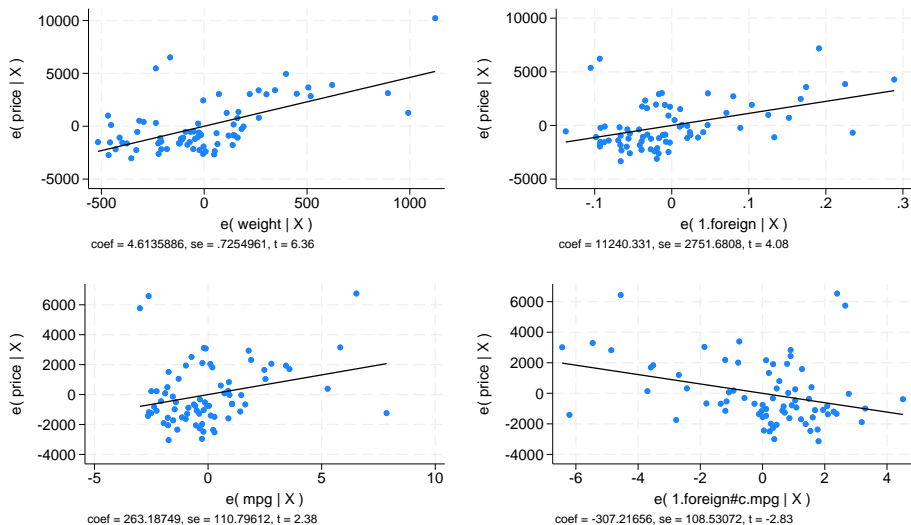
cprplot

Description for cprplot

cprplot graphs a component-plus-residual plot (a.k.a. partial residual plot) after regress. *indepvar* must be an independent variable that is currently in the model.

Menu for cprplot

Statistics > Linear models and related > Regression diagnostics > Component-plus-residual plot

Syntax for cprplot

cprplot *indepvar* [, *cprplot_options*]

cprplot_options	Description
Plot	
marker_options	change look of markers (color, size, etc.)
marker_label_options	add marker labels; change look or position
Reference line	
rlopts(*cline_options*)	affect rendition of the reference line
Options	
lowess	add a lowess smooth of the plotted points
lsopts(*lowess_options*)	affect rendition of the lowess smooth
mspline	add median spline of the plotted points
msopts(*mspline_options*)	affect rendition of the spline
Add plots	
addplot(*plot*)	add other plots to the generated graph
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than by() documented in [G-3] *twoway_options*

Options for cprplot

> Plot

marker_options affects the rendition of markers drawn at the plotted points, including their shape, size, color, and outline; see [G-3] *marker_options*.

marker_label_options specify if and how the markers are to be labeled; see [G-3] *marker_label_options*.

> Reference line

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

Options

lowess adds a lowess smooth of the plotted points to assist in detecting nonlinearities.

lsopts(*lowess_options*) affects the rendition of the lowess smooth. For an explanation of these options, especially the bwidth() option, see [R] **lowess**. Specifying lsopts() implies the lowess option.

mspline adds a median spline of the plotted points to assist in detecting nonlinearities.

msopts(*mspline_options*) affects the rendition of the spline. For an explanation of these options, especially the bands() option, see [G-2] **graph twoway mspline**. Specifying msopts() implies the mspline option.

Add plots

addplot(*plot*) provides a way to add other plots to the generated graph. See [G-3] *addplot_option*.

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 for saving the graph to disk (see [G-3] *saving_option*).

Remarks and examples for cprplot

Added-variable plots are successful at identifying outliers, but they cannot be used to identify functional form. The component-plus-residual plot (Ezekiel 1924; Larsen and McCleary 1972) is another attempt at projecting multidimensional data into a two-dimensional form, but with different properties. Although the added-variable plot can identify outliers, the component-plus-residual plot cannot. It can, however, be used to examine the functional form assumptions of the model. Both plots have the property that a regression line through the coordinates has a slope equal to the estimated coefficient in the regression model.

▷ Example 4

We illustrate component-plus-residual plots using a variation of auto.dta.

```
. use https://www.stata-press.com/data/r18/auto1
(Automobile models)

. regress price mpg weight
```

Source	SS	df	MS	Number of obs	=	74
				F(2, 71)	=	14.90
Model	187716578	2	93858289	Prob > F	=	0.0000
Residual	447348818	71	6300687.58	R-squared	=	0.2956
				Adj R-squared	=	0.2757
Total	635065396	73	8699525.97	Root MSE	=	2510.1

price	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
mpg	-55.9393	75.24136	-0.74	0.460	-205.9663	94.08771
weight	1.710992	.5861682	2.92	0.005	.5422063	2.879779
_cons	2197.9	3190.768	0.69	0.493	-4164.311	8560.11

In fact, we know that the effects of mpg in this model are nonlinear—if we added mpg squared to the model, its coefficient would have a t statistic of 2.38, the t statistic on mpg would become -2.48, and weight's effect would become about one-third of its current value and become statistically insignificant. Pretend that we do not know this.

The component-plus-residual plot for mpg is

```
. cprplot mpg, mspline msopts(bands(13))
```
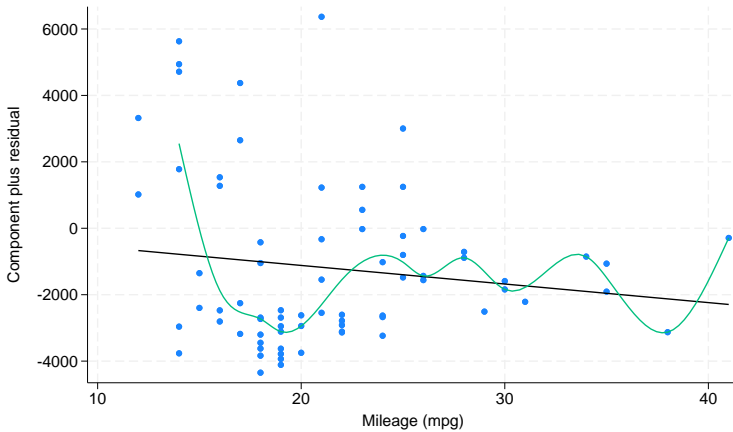
We are supposed to examine the above graph for nonlinearities or, equivalently, ask if the regression line, which has slope equal to the estimated effect of mpg in the original model, fits the data adequately. To assist our eyes, we added a median spline. Perhaps some people may detect nonlinearity from this graph, but we assert that if we had not previously revealed the nonlinearity of mpg and if we had not added the median spline, the graph would not overly bother us.

◁

acprplot

Description for acprplot

acprplot graphs an augmented component-plus-residual plot (a.k.a. augmented partial residual plot) as described by Mallows (1986). This seems to work better than the component-plus-residual plot for identifying nonlinearities in the data.

Menu for acprplot

Statistics > Linear models and related > Regression diagnostics > Augmented component-plus-residual plot

Syntax for acprplot

> acprplot *indepvar* [, *acprplot_options*]

acprplot_options	Description
Plot	
marker_options	change look of markers (color, size, etc.)
marker_label_options	add marker labels; change look or position
Reference line	
rlopts(*cline_options*)	affect rendition of the reference line
Options	
lowess	add a lowess smooth of the plotted points
lsopts(*lowess_options*)	affect rendition of the lowess smooth
mspline	add median spline of the plotted points
msopts(*mspline_options*)	affect rendition of the spline
Add plots	
addplot(*plot*)	add other plots to the generated graph
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than by() documented in [G-3] *twoway_options*

Options for acprplot

Plot

marker_options affects the rendition of markers drawn at the plotted points, including their shape, size, color, and outline; see [G-3] *marker_options*.

marker_label_options specify if and how the markers are to be labeled; see [G-3] *marker_label_options*.

Reference line

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

Options

lowess adds a lowess smooth of the plotted points to assist in detecting nonlinearities.

lsopts(*lowess_options*) affects the rendition of the lowess smooth. For an explanation of these options, especially the bwidth() option, see [R] **lowess**. Specifying lsopts() implies the lowess option.

mspline adds a median spline of the plotted points to assist in detecting nonlinearities.

msopts(*mspline_options*) affects the rendition of the spline. For an explanation of these options, especially the bands() option, see [G-2] **graph twoway mspline**. Specifying msopts() implies the mspline option.

___ Add plots _____

addplot(*plot*) provides a way to add other plots to the generated graph. See [G-3] *addplot_option*.

___ 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 for saving the graph to disk (see [G-3] *saving_option*).

Remarks and examples for acprplot

In the *cprplot* section above, we discussed the component-plus-residual plot. Mallows (1986) proposed an augmented component-plus-residual plot that is often more sensitive to detecting nonlinearity.

▷ Example 5

Let's compare the augmented component-plus-residual plot with the component-plus-residual plot of example 4.

```
. use https://www.stata-press.com/data/r18/auto
(1978 automobile data)
. regress price weight foreign##c.mpg
(output omitted)
. acprplot mpg, mspline msopts(bands(13))
```
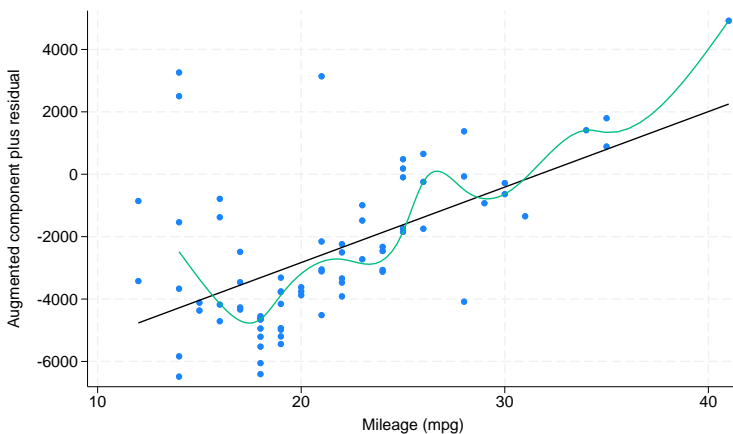

It does do somewhat better.

◁

rvpplot

Description for rvpplot

rvpplot graphs a residual-versus-predictor plot (a.k.a. independent variable plot or carrier plot), a graph of the residuals against the specified predictor.

Menu for rvpplot

Statistics > Linear models and related > Regression diagnostics > Residual-versus-predictor plot

Syntax for rvpplot

rvpplot *indepvar* [, *rvpplot_options*]

rvpplot_options	Description
Plot	
marker_options	change look of markers (color, size, etc.)
marker_label_options	add marker labels; change look or position
Add plots	
addplot(*plot*)	add plots to the generated graph
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than by() documented in [G-3] *twoway_options*

Options for rvpplot

⌐ Plot ⌐

marker_options affects the rendition of markers drawn at the plotted points, including their shape, size, color, and outline; see [G-3] *marker_options*.

marker_label_options specify if and how the markers are to be labeled; see [G-3] *marker_label_options*.

⌐ Add plots ⌐

addplot(*plot*) provides a way to add other plots to the generated graph; see [G-3] *addplot_option*.

⌐ 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 for saving the graph to disk (see [G-3] *saving_option*).

Remarks and examples for rvpplot

The residual-versus-predictor plot is a simple way to look for violations of the regression assumptions. If the assumptions are correct, there should be no pattern on the graph.

▷ Example 6

Let's use our model of `price` on `mpg` and `weight`.

```
. use https://www.stata-press.com/data/r18/auto
(1978 automobile data)
. regress price weight foreign##c.mpg
 (output omitted )
. rvpplot mpg, yline(0)
```
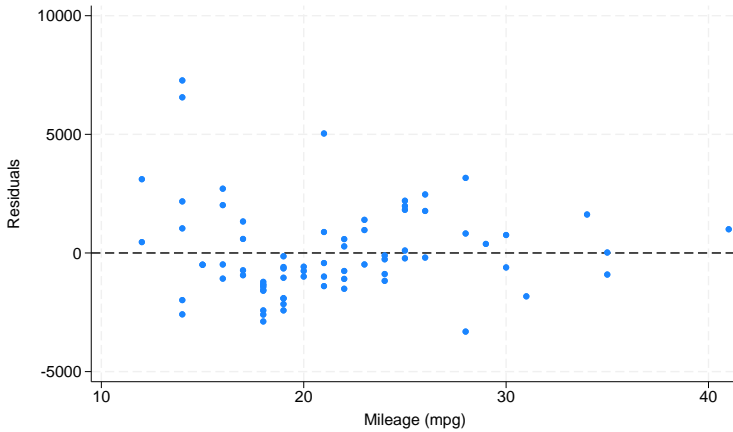

Remember, any pattern counts as a problem, and in this graph, we see that the variation in the residuals decreases as `mpg` increases.

◁

lvr2plot

Description for lvr2plot

lvr2plot graphs a leverage-versus-squared-residual plot (a.k.a. L-R plot).

Menu for lvr2plot

Statistics > Linear models and related > Regression diagnostics > Leverage-versus-squared-residual plot

Syntax for lvr2plot

> lvr2plot [, *lvr2plot_options*]

lvr2plot_options	Description
Plot	
marker_options	change look of markers (color, size, etc.)
marker_label_options	add marker labels; change look or position
Add plots	
addplot(*plot*)	add other plots to the generated graph
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than by() documented in [G-3] *twoway_options*

Options for lvr2plot

> Plot

marker_options affects the rendition of markers drawn at the plotted points, including their shape, size, color, and outline; see [G-3] *marker_options*.

marker_label_options specify if and how the markers are to be labeled; see [G-3] *marker_label_options*.

> Add plots

addplot(*plot*) provides a way to add other plots to the generated graph. See [G-3] *addplot_option*.

> 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 for saving the graph to disk (see [G-3] *saving_option*).

Remarks and examples for lvr2plot

One of the most useful diagnostic graphs is provided by `lvr2plot` (leverage-versus-residual-squared plot), a graph of leverage against the (normalized) residuals squared.

▷ Example 7

We illustrate `lvr2plot` using our model in example 1.

```
. use https://www.stata-press.com/data/r18/auto
(1978 automobile data)
. regress price weight foreign##c.mpg
  (output omitted )
. lvr2plot
```
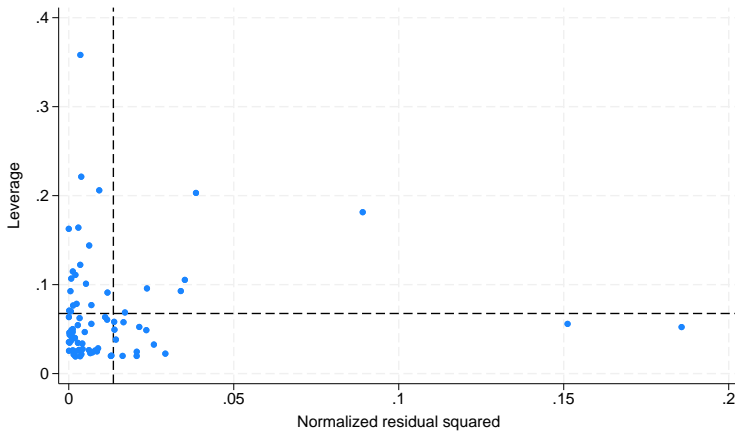

The lines on the chart show the average values of leverage and the (normalized) residuals squared. Points above the horizontal line have higher-than-average leverage; points to the right of the vertical line have larger-than-average residuals.

One point immediately catches our eye, and four more make us pause. The point at the top of the graph has high leverage and a smaller-than-average residual. The other points that bother us all have higher-than-average leverage, two with smaller-than-average residuals and two with larger-than-average residuals.

A less pretty but more useful version of the above graph specifies that `make` be used as the symbol (see [G-3] *marker_label_options*):

```
. lvr2plot, mlabel(make) mlabp(0) m(none) mlabsize(small)
```
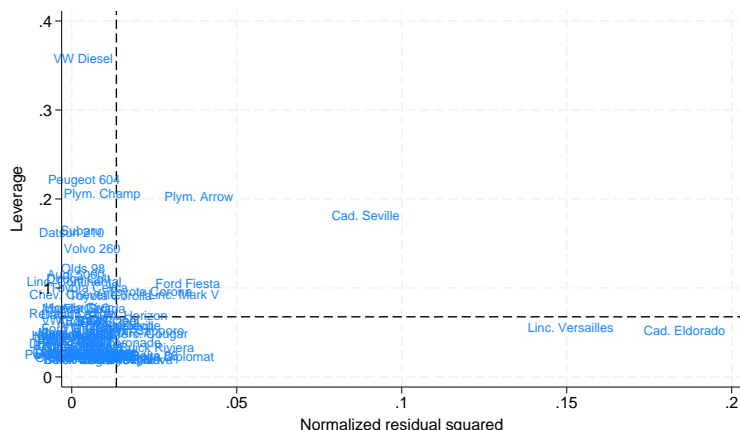

The VW Diesel, Plymouth Champ, Plymouth Arrow, and Peugeot 604 are the points that cause us the most concern. When we further examine our data, we discover that the VW Diesel is the only diesel in our data and that the data for the Plymouth Arrow were entered incorrectly into the computer. No such simple explanations were found for the Plymouth Champ and Peugeot 604.

◁

Methods and formulas

See Hamilton (2013, 209–214) and Kohler and Kreuter (2012, sec. 9.3) for a discussion of these diagnostic graphs.

The `lvr2plot` command plots leverage against the squares of the normalized residuals. The normalized residuals are defined as $\widehat{e}_{n_j} = \widehat{e}_j / (\sum_i \widehat{e}_i^2)^{1/2}$.

References

Belsley, D. A., E. Kuh, and R. E. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Chatterjee, S., and A. S. Hadi. 1986. Influential observations, high leverage points, and outliers in linear regression. *Statistical Science* 1: 379–393. https://doi.org/10.1214/ss/1177013622.

Cox, N. J. 2004. Speaking Stata: Graphing model diagnostics. *Stata Journal* 4: 449–475.

Ezekiel, M. 1924. A method of handling curvilinear correlation for any number of variables. *Journal of the American Statistical Association* 19: 431–453. https://doi.org/10.2307/2281561.

Gallup, J. L. 2019. Added-variable plots with confidence intervals. *Stata Journal* 19: 598–614.

Hamilton, L. C. 2013. *Statistics with Stata: Updated for Version 12*. 8th ed. Boston: Brooks/Cole.

Hoaglin, D. C., and R. E. Welsch. 1978. The hat matrix in regression and ANOVA. *American Statistician* 32: 17–22. https://doi.org/10.1080/00031305.1978.10479237.

Kohler, U., and F. Kreuter. 2012. *Data Analysis Using Stata*. 3rd ed. College Station, TX: Stata Press.

Larsen, W. A., and S. J. McCleary. 1972. The use of partial residual plots in regression analysis. *Technometrics* 14: 781–790. https://doi.org/10.1080/00401706.1972.10488966.

Lindsey, C., and S. J. Sheather. 2010a. Optimal power transformation via inverse response plots. *Stata Journal* 10: 200–214.

——. 2010b. Model fit assessment via marginal model plots. *Stata Journal* 10: 215–225.

Mallows, C. L. 1986. Augmented partial residuals. *Technometrics* 28: 313–319. https://doi.org/10.2307/1268980.

Mosteller, C. F., and J. W. Tukey. 1977. *Data Analysis and Regression: A Second Course in Statistics.* Reading, MA: Addison–Wesley.

Pedace, R. 2013. *Econometrics for Dummies.* Hoboken, NJ: Wiley.

Welsch, R. E. 1986. Comment [on Chatterjee and Hadi 1986]. *Statistical Science* 1: 403–405. https://doi.org/10.1214/ss/1177013625.

Also see

[R] **regress** — Linear regression

[R] **regress postestimation** — Postestimation tools for regress

[R] **regress postestimation time series** — Postestimation tools for regress with time series

[U] **20 Estimation and postestimation commands**