xcorr — Cross-correlogram for bivariate time series

Description

xcorr plots the sample cross-correlation function.

Quick start

Plot sample cross-correlation function for y1 and y2 using `tsset` data

```
xcorr y1 y2
```

Same as above, but display a table instead of a graph

```
xcorr y1 y2, table
```

Same as above, but only include 10 lags and leads of the cross-correlation function

```
xcorr y1 y2, table lags(10)
```

Same as above, and generate xcorrvar containing the cross-correlation values

```
xcorr y1 y2, table lags(10) generate(xcorrvar)
```

Menu

Statistics > Time series > Graphs > Cross-correlogram for bivariate time series

Syntax

    xcorr *varname*$_1$ *varname*$_2$ [*if*] [*in*] [, *options*]

options	Description
[Main]	
<u>gen</u>erate(*newvar*)	create *newvar* containing cross-correlation values
table	display a table instead of graphical output
noplot	do not include the character-based plot in tabular output
<u>lag</u>s(#)	include # lags and leads in graph
[Plot]	
base(#)	value to drop to; default is 0
marker_options	change look of markers (color, size, etc.)
marker_label_options	add marker labels; change look or position
line_options	change look of dropped lines
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*

You must tsset your data before using xcorr; see [TS] **tsset**.
varname$_1$ and *varname*$_2$ may contain time-series operators; see [U] **11.4.4 Time-series varlists**.

Options

        [Main]

generate(*newvar*) specifies a new variable to contain the cross-correlation values.

table requests that the results be presented as a table rather than the default graph.

noplot requests that the table not include the character-based plot of the cross-correlations.

lags(#) indicates the number of lags and leads to include in the graph. The default is to use
    min($\lfloor n/2 \rfloor - 2, 20$).

        [Plot]

base(#) specifies the value from which the lines should extend. The default is base(0).

marker_options, *marker_label_options*, and *line_options* affect the rendition of the plotted cross-
    correlations.

    marker_options specify the look of markers. This look includes the marker symbol, the marker size,
        and its 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*.

    line_options specify the look of the dropped lines, including pattern, width, and color; see
        [G-3] *line_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

▷ Example 1

We have a bivariate time series (Box et al. 2016 , Series J) on the input and output of a gas furnace, where 296 paired observations on the input (gas rate) and output (% CO_2) were recorded every 9 seconds. The cross-correlation function is given by

```
. use https://www.stata-press.com/data/r19/furnace
(TIMESLAB: Gas furnace)
. xcorr input output, xline(5) lags(40)
```
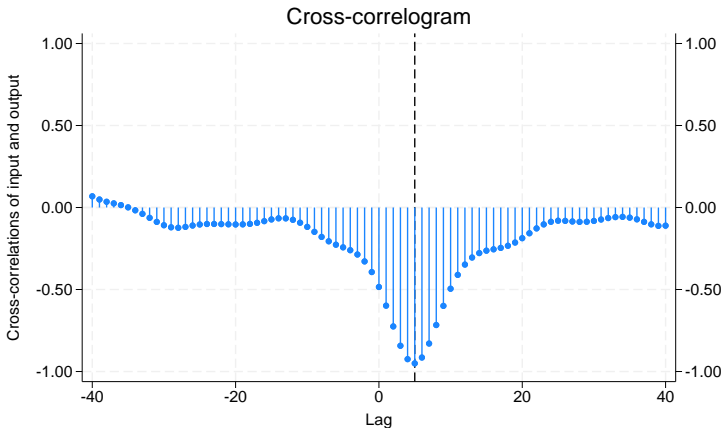
We included a vertical line at lag 5, because there is a well-defined peak at this value. This peak indicates that the output lags the input by five periods. Further, the fact that the correlations are negative indicates that as input (coded gas rate) is increased, output (% CO_2) decreases.

We may obtain the table of autocorrelations and the character-based plot of the cross-correlations (analogous to the univariate time-series command corrgram) by specifying the table option.

```
. xcorr input output, table
                        -1       0       1
   LAG      CORR      [Cross-correlation]

   -20     -0.1033                 |
   -19     -0.1027                 |
   -18     -0.0998                 |
   -17     -0.0932                 |
   -16     -0.0832                 |
   -15     -0.0727                 |
   -14     -0.0660                 |
   -13     -0.0662                 |
   -12     -0.0751                 |
   -11     -0.0927                 |
   -10     -0.1180                 |
    -9     -0.1484              —  |
    -8     -0.1793              —  |
    -7     -0.2059              —  |
    -6     -0.2266              —  |
    -5     -0.2429              —  |
    -4     -0.2604            —    |
    -3     -0.2865            —    |
    -2     -0.3287            —    |
    -1     -0.3936           —     |
     0     -0.4845          —      |
     1     -0.5985          —      |
     2     -0.7251        —        |
     3     -0.8429       —         |
     4     -0.9246      —          |
     5     -0.9503     —           |
     6     -0.9146      —          |
     7     -0.8294       —         |
     8     -0.7166        —        |
     9     -0.5998          —      |
    10     -0.4952          —      |
    11     -0.4107           —     |
    12     -0.3479            —    |
    13     -0.3049            —    |
    14     -0.2779            —    |
    15     -0.2632            —    |
    16     -0.2548            —    |
    17     -0.2463             —   |
    18     -0.2332             —   |
    19     -0.2135              —  |
    20     -0.1869              —  |
```

Once again, the well-defined peak is apparent in the plot.

Methods and formulas

The cross-covariance function of lag k for time series x_1 and x_2 is given by

$$\text{Cov}\left\{x_1(t), x_2(t+k)\right\} = R_{12}(k)$$

This function is not symmetric about lag zero; that is,

$$R_{12}(k) \neq R_{12}(-k)$$

We define the cross-correlation function as

$$\rho_{ij}(k) = \text{Corr}\left\{x_i(t), x_j(t+k)\right\} = \frac{R_{ij}(k)}{\sqrt{R_{ii}(0)R_{jj}(0)}}$$

where ρ_{11} and ρ_{22} are the autocorrelation functions for x_1 and x_2, respectively. The sequence $\rho_{12}(k)$ is the cross-correlation function and is drawn for lags $k \in (-Q, -Q+1, \ldots, -1, 0, 1, \ldots, Q-1, Q)$.

If $\rho_{12}(k) = 0$ for all lags, x_1 and x_2 are not cross-correlated.

References

Box, G. E. P., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. 2016. *Time Series Analysis: Forecasting and Control*. 5th ed. Hoboken, NJ: Wiley.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press. https://doi.org/10.2307/j. ctv14jx6sm.

Also see