corrgram — Tabulate and graph autocorrelations

Description

corrgram produces a table of the autocorrelations, partial autocorrelations, and portmanteau ($Q$) statistics. It also displays a character-based plot of the autocorrelations and partial autocorrelations. See [TS] wntestq for more information on the $Q$ statistic.

ac produces a correlogram (a graph of autocorrelations) with pointwise confidence intervals that is based on Bartlett’s formula for MA(q) processes.

pac produces a partial correlogram (a graph of partial autocorrelations) with confidence intervals calculated using a standard error of $1/\sqrt{n}$. The residual variances for each lag may optionally be included on the graph.

Quick start

Produce correlogram for $y$ using tsset data

corrgram y

As above, but limit the number of computed autocorrelations to 10

corrgram y, lags(10)

Plot the autocorrelation function for $y$

ac y

As above, and generate newv to hold the autocorrelations

ac y, generate(newv)

Plot partial autocorrelation function for $y$, and include standardized residual variances in the graph

pac y, srv

Menu

corrgram
Statistics > Time series > Graphs > Autocorrelations & partial autocorrelations

ac
Statistics > Time series > Graphs > Correlogram (ac)

pac
Statistics > Time series > Graphs > Partial correlogram (pac)
Syntax

Autocorrelations, partial autocorrelations, and portmanteau (Q) statistics

```
corrgram varname [if] [in] [ , corrgram_options ]
```

Graph autocorrelations with confidence intervals

```
ac varname [if] [in] [ , ac_options ]
```

Graph partial autocorrelations with confidence intervals

```
pac varname [if] [in] [ , pac_options ]
```

corrgram_options Description

<table>
<thead>
<tr>
<th>Main</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>lags(#)</code></td>
<td>calculate # autocorrelations</td>
</tr>
<tr>
<td><code>noplot</code></td>
<td>suppress character-based plots</td>
</tr>
<tr>
<td><code>yw</code></td>
<td>calculate partial autocorrelations by using Yule–Walker equations</td>
</tr>
</tbody>
</table>

ac_options Description

<table>
<thead>
<tr>
<th>Main</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>lags(#)</code></td>
<td>calculate # autocorrelations</td>
</tr>
<tr>
<td><code>generate(newvar)</code></td>
<td>generate a variable to hold the autocorrelations</td>
</tr>
<tr>
<td><code>level(#)</code></td>
<td>set confidence level; default is <code>level(95)</code></td>
</tr>
<tr>
<td><code>fft</code></td>
<td>calculate autocorrelation by using Fourier transforms</td>
</tr>
</tbody>
</table>

Plot

<table>
<thead>
<tr>
<th>Main</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>line_options</code></td>
<td>change look of dropped lines</td>
</tr>
<tr>
<td><code>marker_options</code></td>
<td>change look of markers (color, size, etc.)</td>
</tr>
<tr>
<td><code>marker_label_options</code></td>
<td>add marker labels; change look or position</td>
</tr>
</tbody>
</table>

CI plot

<table>
<thead>
<tr>
<th>Main</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>ciopts(area_options)</code></td>
<td>affect rendition of the confidence bands</td>
</tr>
</tbody>
</table>

Add plots

<table>
<thead>
<tr>
<th>Main</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>addplot(plot)</code></td>
<td>add other plots to the generated graph</td>
</tr>
</tbody>
</table>

Y axis, X axis, Titles, Legend, Overall

<table>
<thead>
<tr>
<th>Main</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>twoway_options</td>
<td>any options other than <code>by()</code> documented in [G-3] <code>twoway_options</code></td>
</tr>
</tbody>
</table>
Options for corrgram

Main

- `lags(#)`: specifies the number of autocorrelations to calculate. The default is to use \[\min(\lfloor n/2 \rfloor - 2, 40)\], where \(\lfloor n/2 \rfloor\) is the greatest integer less than or equal to \(n/2\).

- `noplot`: prevents the character-based plots from being in the listed table of autocorrelations and partial autocorrelations.

- `yw`: specifies that the partial autocorrelations be calculated using the Yule–Walker equations instead of using the default regression-based technique. `yw` cannot be used if `srv` is used.

Options for ac and pac

Main

- `lags(#)`: specifies the number of autocorrelations to calculate. The default is to use \[\min(\lfloor n/2 \rfloor - 2, 40)\], where \(\lfloor n/2 \rfloor\) is the greatest integer less than or equal to \(n/2\).

- `generate(newvar)`: specifies a new variable to contain the autocorrelation (ac command) or partial autocorrelation (pac command) values. This option is required if the `nograph` option is used.
nograph (implied when using `generate()` in the dialog box) prevents `ac` and `pac` from constructing a graph. This option requires the `generate()` option.

`yw` (pac only) specifies that the partial autocorrelations be calculated using the Yule–Walker equations instead of using the default regression-based technique. `yw` cannot be used if `srv` is used.

`level(#)` specifies the confidence level, as a percentage, for the confidence bands in the `ac` or `pac` graph. The default is `level(95)` or as set by `set level`; see `[R] level`.

`fft` (ac only) specifies that the autocorrelations be calculated using two Fourier transforms. This technique can be faster than simply iterating over the requested number of lags.

---

`line_options, marker_options, and marker_label_options` affect the rendition of the plotted autocorrelations (with `ac`) or partial autocorrelations (with `pac`).

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

`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`.

---

`ciplot(area_options)` affects the rendition of the confidence bands; see [G-3] `area_options`.

---

`srv` (pac only) specifies that the standardized residual variances be plotted with the partial autocorrelations. `srv` cannot be used if `yw` is used.

`srv_options(marker_options)` (pac only) affects the rendition of the plotted standardized residual variances; see [G-3] `marker_options`. This option implies the `srv` option.

---

`addplot(plot)` adds specified plots to the generated graph; see [G-3] `addplot_option`.

---

`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

Remarks are presented under the following headings:

- Basic examples
- Video example
Basic examples

corrgram tabulates autocorrelations, partial autocorrelations, and portmanteau ($Q$) statistics and plots the autocorrelations and partial autocorrelations. The $Q$ statistics are the same as those produced by [TS] wntestq. ac produces graphs of the autocorrelations, and pac produces graphs of the partial autocorrelations. See Becketti (2020) for additional examples of how these commands are used in practice.

Example 1

Here we use the international airline passengers dataset (Box, Jenkins, and Reinsel 2008, Series G). This dataset has 144 observations on the monthly number of international airline passengers from 1949 through 1960. We can list the autocorrelations and partial autocorrelations by using corrgram.

```
. use https://www.stata-press.com/data/r16/air2
    (TIMESLAB: Airline passengers)
. corrgram air, lags(20)

LAG          AC         PAC           Q      Prob>Q  [Autocorrelation]  [Partial Autocor]
          -1           0          1          -1           0           1
1  0.9480       0.9589      132.14  0.0000
2  0.8756       -0.3298     245.65  0.0000
3  0.8067        0.2018     342.67  0.0000
4  0.7526       0.1450      427.74  0.0000
5  0.7138       0.2585      504.8   0.0000
6  0.6817       -0.0269     575.6   0.0000
7  0.6629       0.2043      643.04  0.0000
8  0.6556        0.1561     709.48  0.0000
9  0.6709       0.5686      779.59  0.0000
10 0.7027       0.2926      857.07  0.0000
11 0.7432       0.8402      944.39  0.0000
12 0.7604       0.6127     1036.5  0.0000
13 0.7127       -0.6660     1118   0.0000
14 0.6463       -0.3846     1185.6  0.0000
15 0.5859       0.0787      1241.5  0.0000
16 0.5380       -0.0266     1289   0.0000
17 0.4997       -0.0581     1330.4  0.0000
18 0.4687       -0.0435     1367   0.0000
19 0.4499        0.2773     1401.1  0.0000
20 0.4416       -0.0405     1434.1  0.0000
```
We can use `ac` to produce a graph of the autocorrelations.

```
. ac air, lags(20)
```

The data probably have a trend component as well as a seasonal component. First-differencing will mitigate the effects of the trend, and seasonal differencing will help control for seasonality. To accomplish this goal, we can use Stata’s time-series operators. Here we graph the partial autocorrelations after controlling for trends and seasonality. We also use `srv` to include the standardized residual variances.

```
. pac DS12.air, lags(20) srv
```

See [U] 11.4.4 Time-series varlists for more information about time-series operators.
Video example

Correlograms and partial correlograms

Stored results

corrgram stores the following in r():

Scalars

- r(lags) number of lags
- r(ac#) AC for lag #
- r(pac#) PAC for lag #
- r(q#) Q for lag #

Matrices

- r(AC) vector of autocorrelations
- r(PAC) vector of partial autocorrelations
- r(Q) vector of Q statistics

Methods and formulas

Box, Jenkins, and Reinsel (2008, sec. 2.1.4); Newton (1988); Chatfield (2004); and Hamilton (1994) provide excellent descriptions of correlograms. Newton (1988) also discusses the calculation of the various quantities.

The autocovariance function for a time series \( x_1, x_2, \ldots, x_n \) is defined for \( |v| < n \) as

\[
\hat{R}(v) = \frac{1}{n} \sum_{i=1}^{n-|v|} (x_i - \bar{x})(x_{i+v} - \bar{x})
\]

where \( \bar{x} \) is the sample mean, and the autocorrelation function is then defined as

\[
\hat{\rho}_v = \frac{\hat{R}(v)}{\hat{R}(0)}
\]

The variance of \( \hat{\rho}_v \) is given by Bartlett’s formula for MA(q) processes. From Brockwell and Davis (2016, 92), we have

\[
\text{Var}(\hat{\rho}_v) = \begin{cases} 
\frac{1}{n} & v = 1 \\
\frac{1}{n} \left( 1 + 2 \sum_{i=1}^{v-1} \hat{\rho}^2(i) \right) & v > 1
\end{cases}
\]

The partial autocorrelation at lag \( v \) measures the correlation between \( x_t \) and \( x_{t+v} \) after the effects of \( x_{t+1}, \ldots, x_{t+v-1} \) have been removed. By default, corrgram and pac use a regression-based method to estimate it. We run an OLS regression of \( x_t \) on \( x_{t-1}, \ldots, x_{t-v} \) and a constant term. The estimated coefficient on \( x_{t-v} \) is our estimate of the \( v \)th partial autocorrelation. The residual variance is the estimated variance of that regression, which we then standardize by dividing by \( \hat{R}(0) \).

If the \( \text{yw} \) option is specified, corrgram and pac use the Yule–Walker equations to estimate the partial autocorrelations. Per Enders (2015, 64–65), let \( \phi_{vv} \) denote the \( v \)th partial autocorrelation coefficient. We then have

\[
\hat{\phi}_{11} = \hat{\rho}_1
\]
and for $v > 1$

$$\hat{\phi}_{vv} = \frac{\hat{\rho}_v - \sum_{j=1}^{v-1} \hat{\phi}_{v-1,j} \hat{\rho}_{v-j}}{1 - \sum_{j=1}^{v-1} \hat{\phi}_{v-1,j} \hat{\rho}_j}$$

and

$$\hat{\phi}_{vj} = \hat{\phi}_{v-1,j} - \hat{\phi}_{vv} \hat{\phi}_{v-1,v-j} \quad j = 1, 2, \ldots, v - 1$$

Unlike the regression-based method, the Yule–Walker equations-based method ensures that the first-sample partial autocorrelation equal the first-sample autocorrelation coefficient, as must be true in the population; see Greene (2008, 725).

McCullough (1998) discusses other methods of estimating $\hat{\phi}_{vv}$; he finds that relative to other methods, such as linear regression, the Yule–Walker equations-based method performs poorly, in part because it is susceptible to numerical error. Box, Jenkins, and Reinsel (2008, 69) also caution against using the Yule–Walker equations-based method, especially with data that are nearly nonstationary.

**Acknowledgment**

The *ac* and *pac* commands are based on the *ac* and *pac* commands written by Sean Becketti (1992), a past editor of the *Stata Technical Bulletin* and author of the Stata Press book *Introduction to Time Series Using Stata*.

**References**


**Also see**

[TS] pergram — Periodogram

[TS] tsset — Declare data to be time-series data

[TS] wntestq — Portmanteau (Q) test for white noise