

Chapter 472

Autocorrelations

Introduction

The correlation between X_t and X_{t+k} is called the k^{th} order *autocorrelation* of X . The sample estimate of this autocorrelation, called r_k , is calculated using the formula:

$$r_k = \frac{\sum_{i=1}^{n-k} (X_i - \bar{X})(X_{i+k} - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

where

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Autocorrelations are used extensively in time series analysis. When plotted, they become the *correlogram* which is used during the identification phase of the Box-Jenkins method. The large sample standard error of the sample autocorrelations is simply $1/\sqrt{n}$ so that large sample confidence limits are $\pm 2/\sqrt{n}$.

The k^{th} order *partial autocorrelation* of X is the partial correlation between X_t and X_{t+k} , where the influence of $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$ have been removed. We use the following recursive formulae to calculate the partial autocorrelations.

$$\hat{\phi}_{k+1,j} = \hat{\phi}_{k,j} - \hat{\phi}_{k+1,k+1} \hat{\phi}_{k,k-j+1}$$

$$\hat{\phi}_{k+1,k+1} = \frac{r_{k+1} - \sum_{j=1}^k \hat{\phi}_{k,j} r_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{k,j} r_j}$$

The partial autocorrelations have the same large sample standard errors and confidence limits as do the autocorrelations. They are also used during the model identification phase of the Box-Jenkins method. For this same reason, the filter is not used by this procedure.

Data Structure

The data are entered in a single variable.

Missing Values

When missing values are found in the series, they are either replaced or omitted. The replacement value is the average of the nearest observation in the future and in the past or the nearest non-missing value in the past.

If you do not feel that this is a valid estimate of the missing value, you should manually enter a more reasonable estimate before using the algorithm. These missing value replacement methods are particularly poor for seasonal data. We recommend that you replace missing values manually before using the algorithm.

Procedure Options

This section describes the options available in this procedure.

Variables Tab

This panel specifies the variables used in the analysis.

Time Series Variable

Time Series Variable

Specify the variable on which to run the analysis.

Use Logarithms

Specifies that the log (base 10) transformation should be applied to the values of the variable.

Missing Values

Choose how missing (blank) values are processed.

The algorithm used in this procedure cannot tolerate missing values since each row is assumed to represent the next point in a time sequence. Hence, when missing values are found, they must be removed either by imputation (filling in with a reasonable value) or by skipping the row and pretending it does not exist.

Whenever possible, we recommend that you replace missing values manually.

Here are the available options.

Average the Adjacent Values

Replace the missing value with the average of the nearest values in the future (below) and in the past (above).

Carry the Previous Value Forward

Replace the missing value with the first non-missing value immediately above (previous) this value.

Omit Row from Calculations

Ignore the row in all calculations. Analyze the data as if the row was not on the database.

Data Adjustment Options

Remove Mean

Checking this option causes the series average to be subtracted from the data. This is almost always done.

Remove Trend

Checking this option causes the least squares trend line to be subtracted from the data. This is sometimes done, although differencing is usually used to remove trends.

Autocorrelations

Regular Differencing

This option lets you designate whether the original series, the first differences, or the second differences are analyzed. The first difference series, W , is calculated using the formula:

$$W_t = X_t - X_{t-1}$$

which may be written using the backshift operator, B , as:

$$W_t = (I - B)X_t$$

The second difference series, Z , is the first difference of the W series. The formula is:

$$Z_t = W_t - W_{t-1}$$

which may be written using the backshift operator, B , as:

$$Z_t = (I - B)^2 X_t$$

Seasonal Differencing

This option lets you designate whether the original series, the first seasonal differences, or the second seasonal differences are analyzed. Assuming the number of seasons is s , the first seasonal difference series, W , is calculated using the formula:

$$W_t = X_t - X_{t-s}$$

which may be written using the backshift operator, B , as:

$$W_t = (I - B^s)X_t$$

The second seasonal difference series, Z , is the first seasonal difference of the W series. The formula is:

$$Z_t = W_t - W_{t-s}$$

which may be written using the backshift operator, B , as:

$$Z_t = (I - B^s)^2 X_t$$

Seasonality Options

Seasons

Specify the number of seasons, s , in the series. Use '4' for quarterly data or '12' for monthly data. Note that this option is only used if seasonal differencing is used.

Reports Tab

The following options control which reports are displayed.

Select Reports

Autocorrelation Report - Partial Autocorrelation Report

Each of these options specifies whether the indicated report is displayed.

Autocorrelations

Report / Plot Options

Number of Autocorrelations

Specify the number of autocorrelations that are calculated and displayed. Note that the number of autocorrelations must be less than the number of rows of data.

Number of Partial Autocorrelations

Specify the number of partial autocorrelations (PAC's) that are calculated and displayed. Note that the number of partial autocorrelations must be less than the number of rows of data.

Precision

Specify the precision of numbers in the report. Single precision will display seven-place accuracy, while the double precision will display thirteen-place accuracy. Note that all reports are formatted for single precision only.

Variable Names

Specify whether to use variable names or (the longer) variable labels in report headings.

Plots Tab

This section controls the inclusion and the settings of the plots.

Select Plots

Autocorrelation Plot - Data Plot

Each of these options specifies whether the indicated plot is displayed. Click the plot format button to change the plot settings.

Horizontal Axis Variable if there are Missing or Filtered Values

Horizontal Variable

This option controls the spacing on the horizontal axis when missing or filtered values occur.

Your choices are

Actual Row Number

Use the actual row number of each row from the dataset along the horizontal axis.

Sequence Number

Use the sequence (relative row) number formed by ignoring any missing or filtered values.

Storage Tab

Both the autocorrelations and the partial autocorrelations may be stored on the current dataset for further analysis. This group of options lets you designate which statistics (if any) should be stored and which columns should receive these statistics.

Note that the variables you specify must already have been named on the current dataset.

Note that existing data is replaced. Be careful that you do not specify columns that contain important data.

Data Storage Columns

Autocorrelations

The autocorrelations are stored in this column.

Partial Autocorrelations

The partial autocorrelations are stored in this column.

Example 1 – Generating Autocorrelations of a Series

This section presents an example of how to generate autocorrelations of a series. The Spots variable in the Sunspot dataset will be used.

You may follow along here by making the appropriate entries or load the completed template **Example 1** by clicking on Open Example Template from the File menu of the Autocorrelations window.

1 Open the Sunspot dataset.

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **Sunspot.NCSS**.
- Click **Open**.

2 Open the Autocorrelations window.

- Using the Analysis menu or the Procedure Navigator, find and select the **Autocorrelations** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

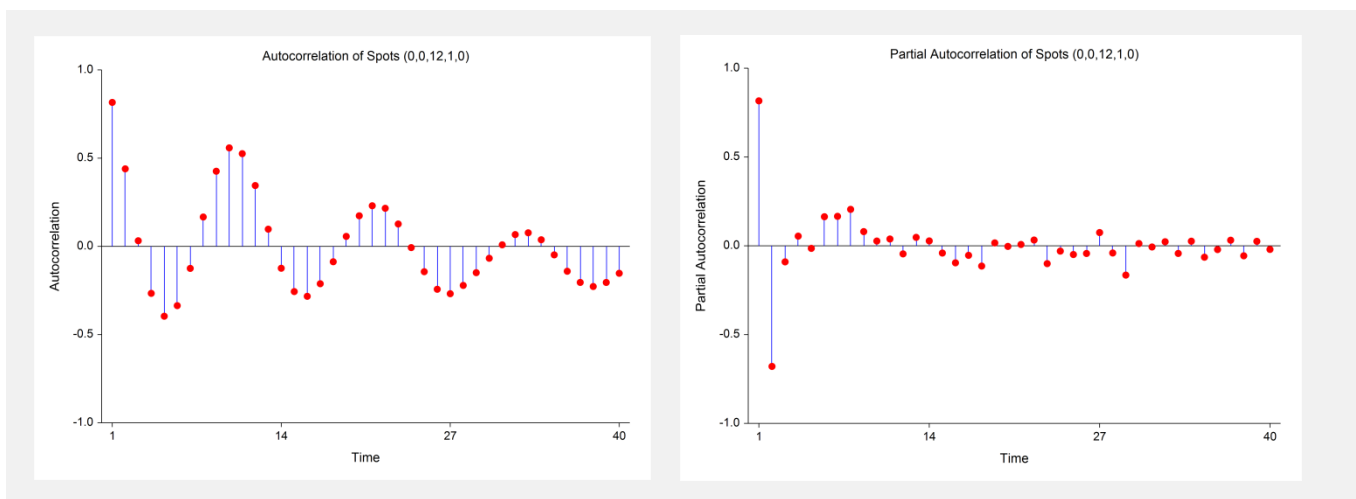
3 Specify the variables.

- On the Autocorrelations window, select the **Variables tab**.
- Double-click in the **Time Series Variable** box. This will bring up the variable selection window.
- Select **Spots** from the list of variables and then click **Ok**.

4 Run the procedure.

- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

Autocorrelation Plots Section



This section displays the autocorrelations and partial autocorrelations in a plot format.

Autocorrelations

Autocorrelations Section

Autocorrelation Section

Autocorrelations of SPOTS (0,0,12,1,0)

Lag	Correlation	Lag	Correlation	Lag	Correlation	Lag	Correlation
1	0.816234	11	0.525615	21	0.230319	31	0.008908
2	0.439589	12	0.344871	22	0.216010	32	0.066994
3	0.031927	13	0.097503	23	0.126930	33	0.077056
4	-0.266327	14	-0.123974	24	-0.006907	34	0.037571
5	-0.395920	15	-0.256157	25	-0.143509	35	-0.048184
6	-0.335935	16	-0.283001	26	-0.243137	36	-0.141013
7	-0.124787	17	-0.211754	27	-0.268284	37	-0.204330
8	0.166522	18	-0.087193	28	-0.221329	38	-0.227475
9	0.426074	19	0.056621	29	-0.149028	39	-0.204460
10	0.558426	20	0.173492	30	-0.067392	40	-0.152599

Significant if |Correlation| > 0.136399

This section shows the values of the autocorrelations for the specified number of lags. The numbers in parentheses, (d,D,s,M,T), are defined as follows:

- d** is the regular differencing order.
- D** is the seasonal differencing order.
- s** is the number of seasons (ignored if D is 0).
- M** is 1 if the mean is subtracted, 0 otherwise.
- T** is 1 if the trend is subtracted, 0 otherwise.

Partial Autocorrelation Section

Partial Autocorrelation Section

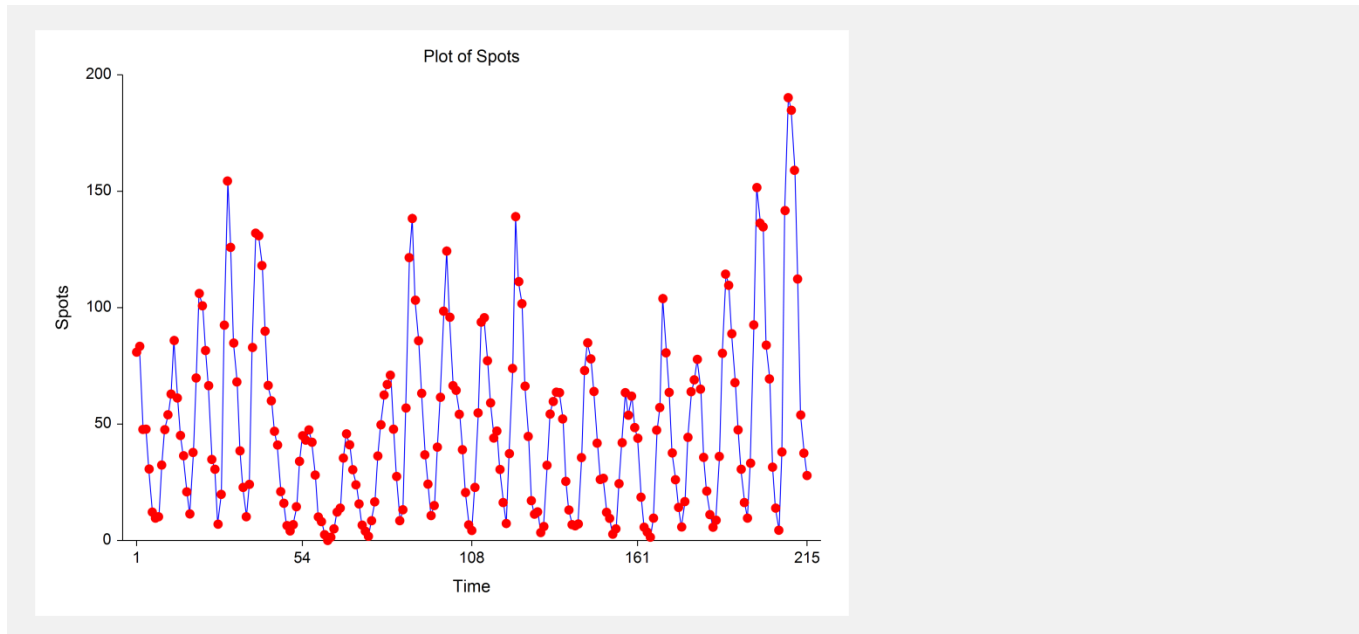
Partial Autocorrelations of SPOTS (0,0,12,1,0)

Lag	Correlation	Lag	Correlation	Lag	Correlation	Lag	Correlation
1	0.816234	11	0.038335	21	0.007443	31	-0.006587
2	-0.679072	12	-0.045487	22	0.032548	32	0.022903
3	-0.090380	13	0.047515	23	-0.100418	33	-0.042974
4	0.054429	14	0.027539	24	-0.029899	34	0.025960
5	-0.014413	15	-0.040822	25	-0.049139	35	-0.064242
6	0.163731	16	-0.095774	26	-0.043122	36	-0.021117
7	0.165977	17	-0.053699	27	0.074602	37	0.031502
8	0.205197	18	-0.114117	28	-0.040450	38	-0.056303
9	0.079963	19	0.016408	29	-0.164996	39	0.025227
10	0.026876	20	-0.003383	30	0.012685	40	-0.020340

Significant if |Correlation| > 0.136399

This section shows the values of the partial autocorrelations for the specified number of lags. The numbers in parentheses are defined above.

Data Plot Section



This section displays a plot of the data values.