

Chapter 473

Cross-Correlations

Introduction

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

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

where

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

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

The time index, k , is allowed to be either positive or negative. The large sample standard error of the sample cross correlations is simply $1/\sqrt{n}$ so that large sample confidence limits are $\pm 2/\sqrt{n}$.

Data Structure

The data are entered in two columns.

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

Specify the variables on which to run the analysis.

Time Series Variables

Y Variable

Specify the first variable to be cross correlated.

X Variable

Specify the second variable to be cross correlated.

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.

Reports Tab

The following options control which reports are displayed.

Select Reports

Cross-Correlation Report

This option specifies whether the indicated report is displayed.

Report Options

Number of Cross-Correlations

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

Variable Names

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

Cross-Correlations

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.

Plots Tab

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

Select Plots

Cross-Correlation Plot - Data Plots

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

Data Storage Columns

Cross Correlations

The cross correlations are stored in this column.

Example 1 – Generating Cross-Correlations of Two Series

This section presents an example of how to generate cross correlations of two series. The Intel_Volume and Intel_Close variables in the Intel 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 Cross-Correlations window.

1 Open the Intel dataset.

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

2 Open the Cross-Correlations window.

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

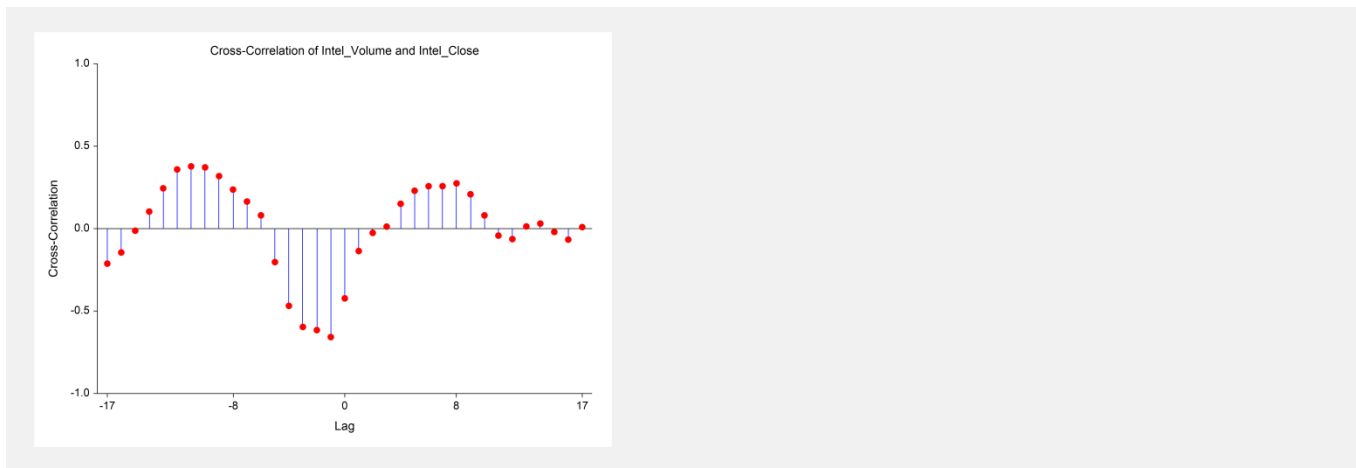
3 Specify the variables.

- On the Cross-Correlations window, select the **Variables tab**.
- Double-click in the **Y Variable** box. This will bring up the variable selection window.
- Select **Intel_Volume** from the list of variables and then click **Ok**.
- Double-click in the **X Variable** box. This will bring up the variable selection window.
- Select **Intel_Close** 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.

Cross Correlation Plot Section



This section displays the cross correlations from both positive and negative lags. The value at lag 0 is the simple correlation between these two variables.

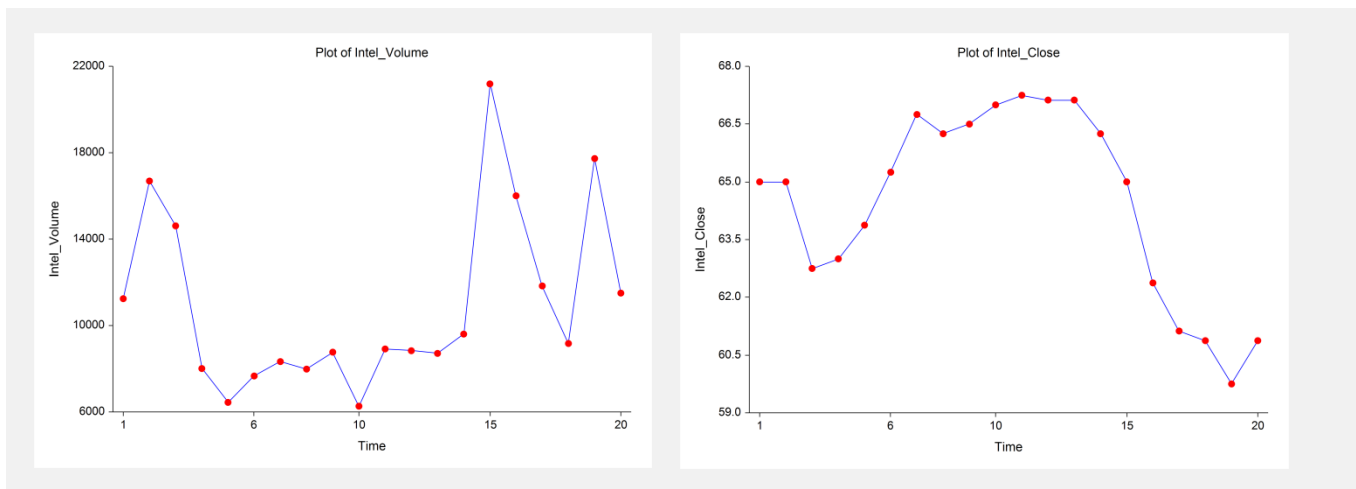
Cross Correlations Section

Cross-Correlations of Intel_Volume

Lag	Correlation	Lag	Correlation	Lag	Correlation	Lag	Correlation
-17	-0.212089	-8	0.236876	1	-0.135799	10	0.080439
-16	-0.144536	-7	0.164298	2	-0.025696	11	-0.042123
-15	-0.012648	-6	0.080595	3	0.012290	12	-0.063397
-14	0.103407	-5	-0.202975	4	0.150682	13	0.013433
-13	0.244696	-4	-0.468207	5	0.229830	14	0.030845
-12	0.359095	-3	-0.596296	6	0.257356	15	-0.020010
-11	0.377600	-2	-0.615427	7	0.257654	16	-0.066433
-10	0.371336	-1	-0.657680	8	0.274812	17	0.009300
-9	0.318597	0	-0.422771	9	0.208603		

This section shows the values of the cross correlations for the specified number of lags.

Data Plot Section



This section displays plots of the data values.