Chapter 473

Cross-Correlations

Introduction

The cross correlation between X_t and Y_{t+k} is called the kth 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.

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.

Setup

To run this example, complete the following steps:

1   Open the Intel example dataset
- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Intel** and click **OK**.

2   Specify the Cross-Correlations procedure options
- Find and open the **Cross-Correlations** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.
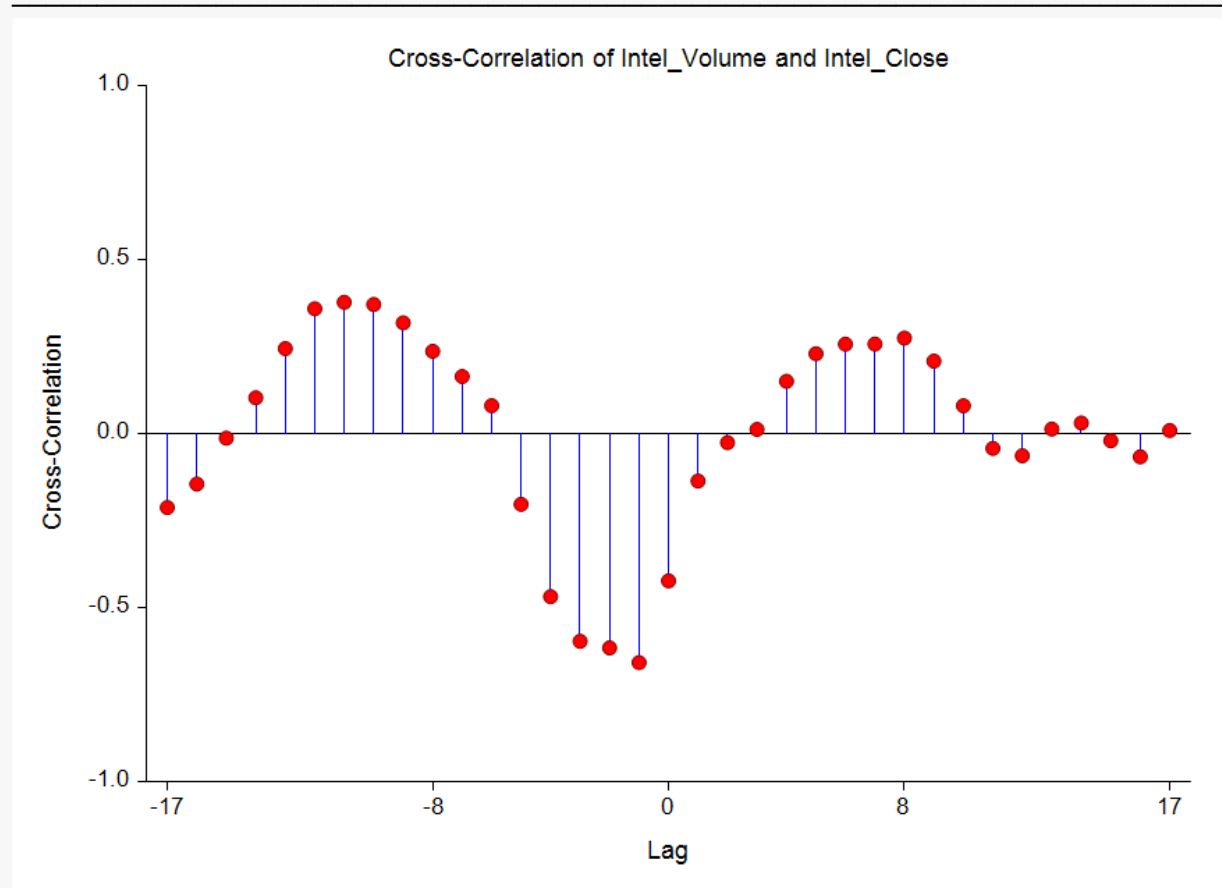
Variables Tab

Y Variable ....**Intel_Volume**
X Variable ....**Intel_Close**

3   Run the procedure
- Click the **Run** button to perform the calculations and generate the output.

Cross Correlation Plot

Cross-Correlation Plot

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

Cross-Correlations of Intel_Volume and Intel_Close

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 Plots

Data Plots
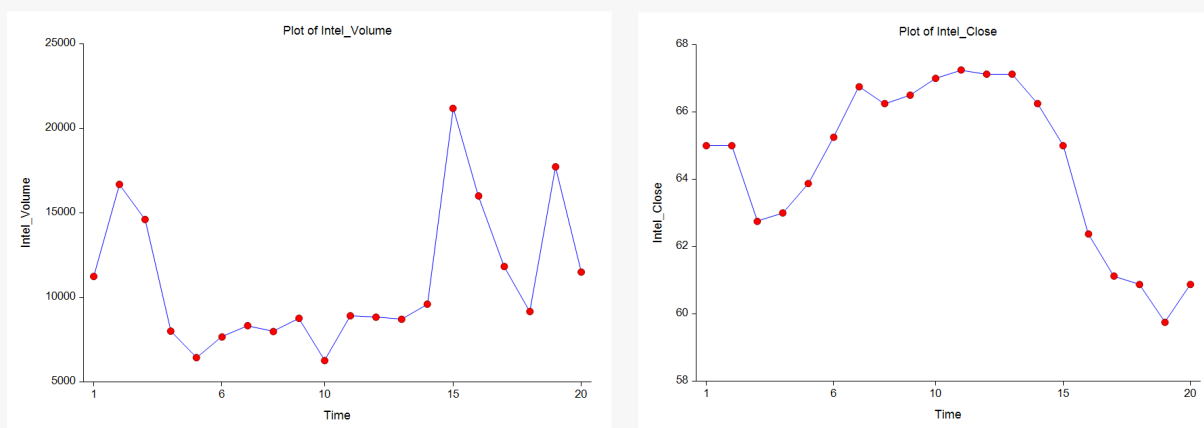
This section displays plots of the data values.