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 kth order partial autocorrelation of X is the partial correlation between X_t and X_{t+k}, where the influence of $X_{t+1}, X_{t+2}, \ldots, X_{t+k-1}$ 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.

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.

Setup

To run this example, complete the following steps:

1 Open the Sunspot example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Sunspot** and click **OK**.

2 Specify the Autocorrelations procedure options

- Find and open the **Autocorrelations** 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

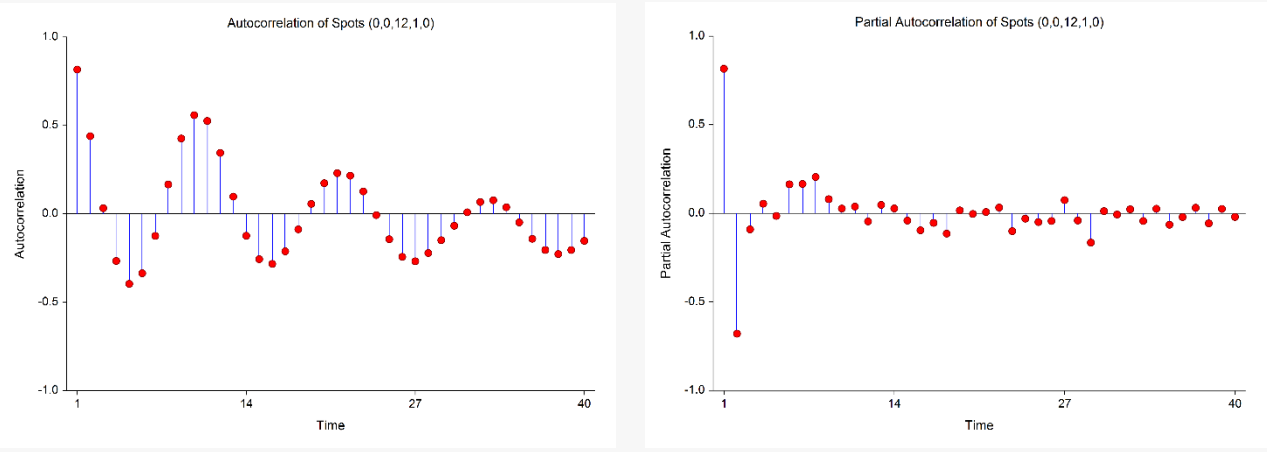
Time Series Variable .....**Spots**

3 Run the procedure

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

Autocorrelation Plots Section

Autocorrelation Plots Section



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

Autocorrelations

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 Autocorrelations

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

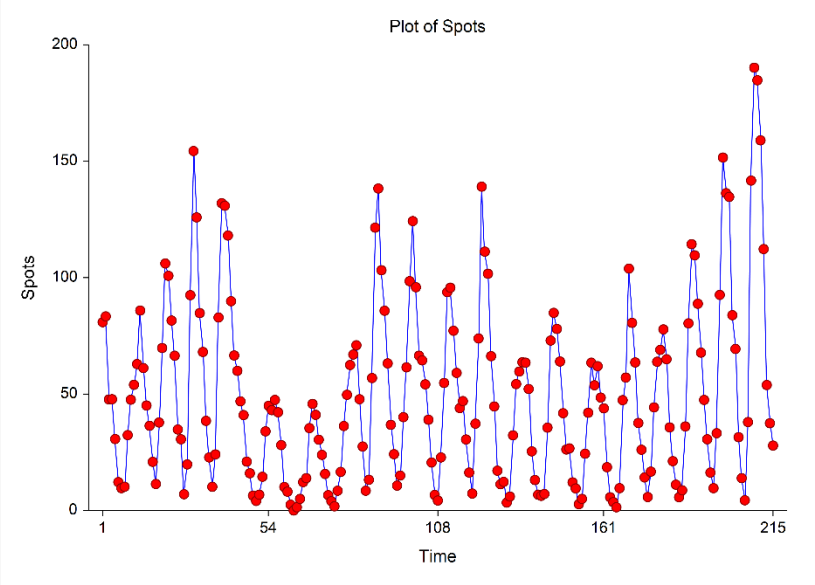
Lag	Correlation	Lag	Correlation	Lag	Correlation	Lag	Correlation
1	0.816234	11	0.038335	21	0.007443	31	-0.006587
2	-0.679072	12	-0.045487	22	0.032548	32	0.022903
3	-0.090380	13	0.047515	23	-0.100418	33	-0.042974
4	0.054429	14	0.027539	24	-0.029899	34	0.025960
5	-0.014413	15	-0.040822	25	-0.049139	35	-0.064242
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

Data Plot Section



This section displays a plot of the data values.