Tutorial QQ - *Field: Finance –Business* -

TITLE: Wall Street – Stock Market Predictions using Times Series Analyses

Guest Author: Thomas Hill, Ph.D. Book co-author Gary Miner, Ph.D.

SUB-TITLE: How to use the STATISTICA software HELP files to find many additional EXAMPLES – CASE STUDIES.

On the *STATISTICA* software – HELP you will also find 7 additional EXAMPLES – TUTORIALS using "Time Series Analysis"; these 7 tutorials are titled:

Time Series Tutorials - case studies:

Example 2: Single Series ARIMA

Example 3: Interrupted ARIMA

Example 4: Seasonal and Non-seasonal Exponential Smoothing

Example 5: Seasonal Decomposition (Census Method I)

Example 6: X-11 Seasonal Decomposition (Census Method II)

Example 7: Distributed Lags Analysis

Example 8: Spectrum (Fourier) Analysis

These Tutorials / Examples / Case Studies are presented using the TIME SERIES MODULE in STATISTICA, using the Interactive Time Series Module. Any of these can be made into "nodes" for the Data Miner Workspace of STATISTICA, and run in a "Competitive Algorithm Comparison" like has been demonstrated in many of the previous tutorials of the Handbook, which would involve the user making a "customized node" in many cases. ***[How to make "Customized Nodes" for the Data Miner Workspace is explained in the HELP files, also – go to the DATA MINER section of the Table of Contents of the HELP to find specific discussions.].***

The following is taken from the STATISTICA software HELP, that is automatically loaded on the computer when the software is installed; if you want to follow this in the STATISTICA – HELP, please install the DVD – STATISTICA Data Miner that is bound with the book: HANDBOOK OF STATISTICAL ANALYSIS & DATA MINING APPLICATIONS.

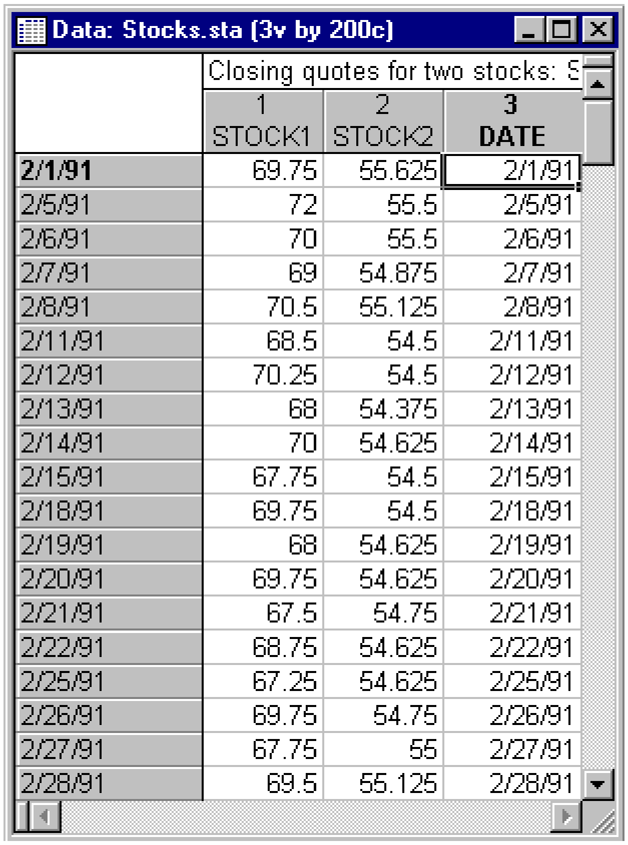
Below, on this Companion Web page for the "HANDBOOK" Tutorial / Example 1 is presented.

Example 1: Transformation of Variables: Wall Street – Stock Market Predictions using Times Series Analyses

Guest Author: Thomas Hill, Ph.D., and edited by 'Handbook' co-author Gary Miner, Ph.D.

General Conventions and Options. In these examples, the general conventions used in the *Time Series* module to maintain the active work area (which functions as a queue of successive transformations of the input series; refer to Active Work Area) will be reviewed. Thus, transformations or the results of other analyses can be undone, saved, etc.

Transformations and the Active Work Area. The data file *Stocks.sta* contains the closing prices for two stocks over a 200-day period. Each trading week consists of exactly five trading days, and closing quotes for holidays (when the stock market was closed) were estimated. In this example, the two time series will be read into memory, some smoothing operations will be performed, several useful time series graphs will be produced, and an autocorrelation analysis of the stock prices will be performed. Open the *Stocks.sta* data file via the *File* - *Open Examples* menu; it is in the *Datasets* folder. The first few cases in the data file are shown below.



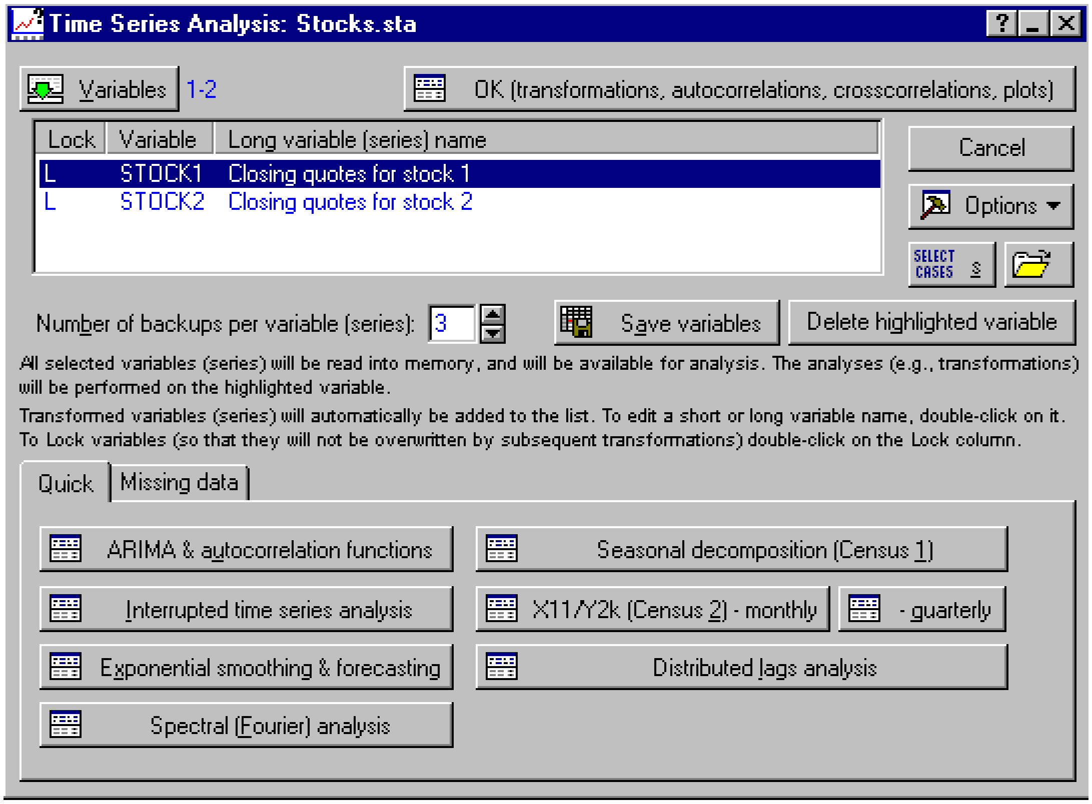
Data: Stocks.sta (3v by 200c)

Closing quotes for two stocks: S

	1 STOCK1	2 STOCK2	3 DATE
2/1/91	69.75	55.625	2/1/91
2/5/91	72	55.5	2/5/91
2/6/91	70	55.5	2/6/91
2/7/91	69	54.875	2/7/91
2/8/91	70.5	55.125	2/8/91
2/11/91	68.5	54.5	2/11/91
2/12/91	70.25	54.5	2/12/91
2/13/91	68	54.375	2/13/91
2/14/91	70	54.625	2/14/91
2/15/91	67.75	54.5	2/15/91
2/18/91	69.75	54.5	2/18/91
2/19/91	68	54.625	2/19/91
2/20/91	69.75	54.625	2/20/91
2/21/91	67.5	54.75	2/21/91
2/22/91	68.75	54.625	2/22/91
2/25/91	67.25	54.625	2/25/91
2/26/91	69.75	54.75	2/26/91
2/27/91	67.75	55	2/27/91
2/28/91	69.5	55.125	2/28/91

Note that this file contains dates in two places, in the variable *Date* (variable 3) and as case names (in the first column of the spreadsheet). The dates were included in those two places to show how they can be used in plots and spreadsheets.

Specifying the Analysis. To start the analysis, select *Time Series/Forecasting* from the *Statistics - Advanced Linear/Nonlinear Models* menu to display the *Time Series Analysis* Startup Panel. Then, click the *Variables* button to display the standard variable selection dialog. Here, select variables *Stock1* and *Stock2* and then click the *OK* button. The opening dialog of the *File* module will now look like this.



All variables (series) and their transformations that are currently available for analysis are stored in the active work area and are listed in the scrollable edit fields at the top of the dialog. When you select new variables, the active work area will first be cleared, and then the selected variables will be read into the active work area (after all "holes" with missing data have been "patched;" see below and Active Work Area).

Highlighted variable. All subsequent analyses will be performed on the highlighted variable. For example, when you perform a transformation, then the currently highlighted variable will be transformed, and a new (transformed) variable will be appended to the active work area. To highlight a variable, simply click on it in any of the scrollable edit fields. For this example, highlight the variable *Stock1*.

Naming conventions. When a new (e.g., transformed) variable is appended to the active work area, it is assigned (1) the same short variable name as the original variable (that was transformed), and (2) a new long variable name that consists of the old long variable name (as much of it as will fit) and a brief description of the respective transformation that was performed. In this manner, as you perform successive transformations or analyses (e.g., successively difference a series), an automatic log of transformations will be maintained in the long variable names.

Editing variable names. Double-click in the column labeled *Variable* or *Long variable (series) name* to edit the short or long variable name for the series in the active work area. Note that the short and long names will only be changed in the active work area, not in the file (use the respective data spreadsheet operations to permanently change those names; see Variable Specs).

Number of backups per variable (series). All dialogs that contain the scrollable edit fields for highlighting a variable (series) for an analysis also contain a field for specifying the desired *Number of backups per variable (series)*. As described above, after a transformation (or other analysis) is performed on a series, the resulting transformed series (or residuals, forecasts, etc. in ARIMA) will be appended to the active work area, and the values of the series prior to the transformation will be maintained as a backup. The number of such backups that will be maintained in the active work area is controlled by this parameter. Thus, for example, if this parameter is set to *3*, and you have just performed the fourth transformation of an original variable, then the series with the data after the first transformation will be dropped from the active work area and replaced by the new (fourth) transformed series. Thus, series created by successive transformations will be appended to the active work area until there are as many backups as specified in this parameter; at that point the respective "oldest" transformed series will be replaced by the new one. Up to 99 backups can be kept of a single original variable. For this example, accept the default value of *3* backups.

Locking variables (series). The first column of the scrollable edit fields carries the header *Lock*. When you double-click in that column for a transformed variable, that variable will be locked in the active work area (or unlocked, if it was previously locked). An *L* will appear in that column to indicate that the variable is now locked. Locked variables will not be replaced as successive transformations exceed the current maximum number of backups (as described in the paragraph above). Note that original (untransformed) variables are always locked, and they cannot be unlocked.

Deleting variables (series) from the current work area. To delete a transformed variable from the active work area, use the *Delete highlighted variable* button. Original (untransformed) variables cannot be deleted.

Saving variables (series) in the current work area. Use the *Save variables* button to save the variables (series) in the active work area. You can save all variables or only selected variables.

Missing Data. Practically all time series analyses require that all data are observed, and that there are no "holes" with missing data in the time series. As long as the missing data are at the end of the series (trailing missing data) or the beginning of the series (leading missing data), the missing data will simply be ignored. Missing data embedded in the series have to be replaced in some way. The *Time Series* module offers a range of different methods for dealing with missing data in this case, which are described in the *Time Series Analysis Startup Panel - Missing Data* tab topic. For this example, select the missing data *Interpolation from adjacent points* option button on the *Missing Data* tab.

Note that the chosen missing data replacement method will be used not only when reading selected variables from the data file into the active work area, but it will also be used when time series transformations result in embedded missing data. For example, suppose an input series contains a few 0's (zeros), and you request a log transformation. Since the log of 0 is undefined, *STATISTICA* will replace those observations with missing data; then, in a second pass through the series, those missing data will be replaced according to the method chosen on the opening dialog.

Overall mean. In this method, all missing data will simply be replaced by the overall mean of the series. Very often, when the series is not stationary (see Introductory Overview), or when there are large systematic fluctuations in the values of the series, this method may not be appropriate. On the other hand, the overall mean is often the best *a priori* (unbiased) guess for the missing data.

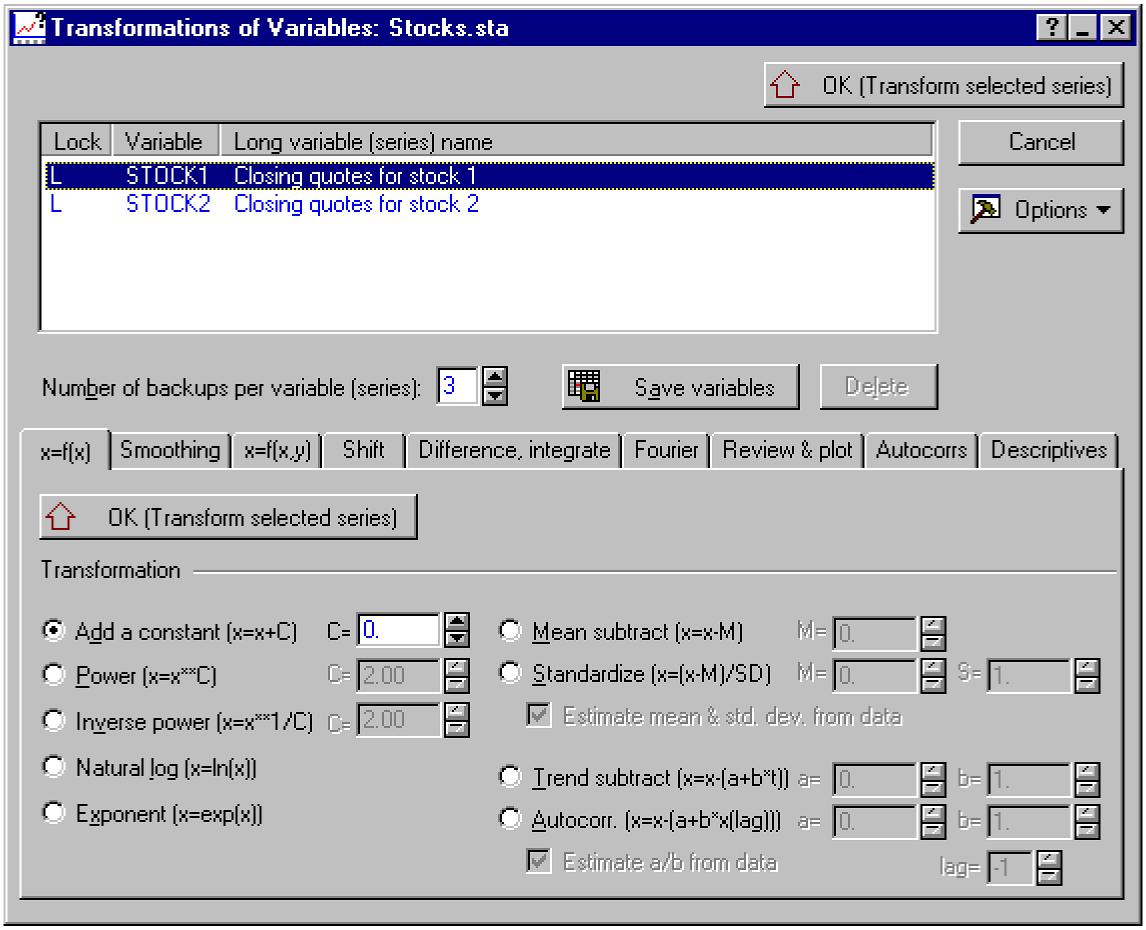
Interpolation from adjacent points. In this method, the missing data are computed by interpolation from the adjacent non-missing points. Graphically, this method amounts to replacing missing data by connecting with a straight line the point just prior to the missing data with the point just following the missing data. Thus, this method in a sense assumes that there is some serial correlation in the data, that is, that each observation is to some extent related to and therefore most similar to the previous observation.

Mean of N adjacent points. In this method the missing data are computed from the mean of the N adjacent points on both sides of the "hole" of missing data. For example, when N is left at its default value of *1*, then missing data will be replaced by the average of the value just prior to the missing data and the value immediately following the missing data. In general, this method implies that the data in the region or window specified by the N parameter are more similar to each other than points that are further away.

Median of N adjacent points. This method is essentially the same as that described above, except that missing data are replaced by the median of the N non-missing adjacent points.

Predicted values from linear trend regression. In this method, *STATISTICA* will fit a least-squares regression line to the time series. The missing data will then be replaced by the values predicted by this regression line. This method implies that the most salient (or strongest) feature of the series is its linear trend across time.

Reviewing the Time Series. Now proceed with the analysis and review the closing quotes for the two series. Click the *OK (transformations, autocorrelations, crosscorrelations, plots)* button to display the *Transformations of Variables* dialog.



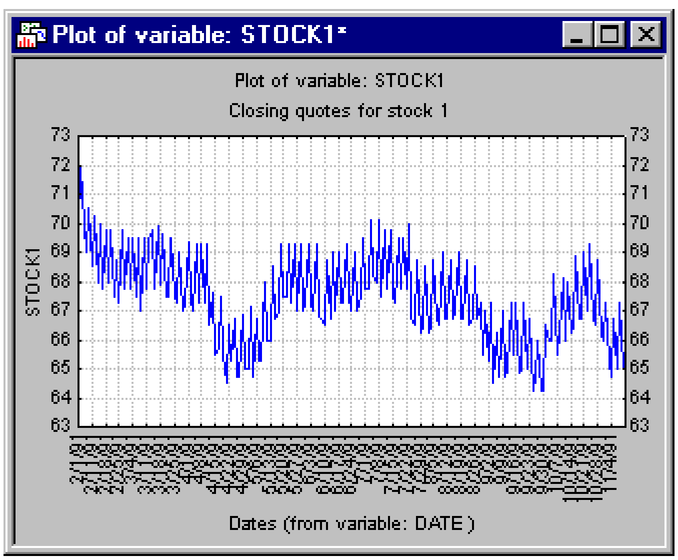
Several options for reviewing time series in the active work area are shown on Review & plot tab. The *Label data points with* box contains options that will determine how the horizontal (time) axis is scaled and labeled. Note that those options will affect all plots in the *Time Series* module (there are numerous options for plotting time series available with all procedures in the *Time Series* module). The example data file *Stocks.sta* contains dates in variable *Date* and in the case names and either of these can be used to label the horizontal axes in plots (select the appropriate option button).

Note that case names may also contain information other than dates; for example, significant discrete events affecting the series (e.g., release of news affecting the stock prices) could also be noted in the case names and be used as labels in plots. For this example, label the horizontal axis in the plots with the dates in the variable *Date*. Select the *Dates from a variable* option button, and select that variable from the subsequent variable selection dialog and then click the *OK* button. Because in this example series, each trading week consists of five days (Monday through Friday), select the *Scale X axis in plots manually* check box and enter as the minimum (*Min*=) *1* (start with the first day), and the step size (*Step*=) *5*. Then, click the *Review highlighted variable* button to produce the following spreadsheet.

Data: Listing of variable: STO...

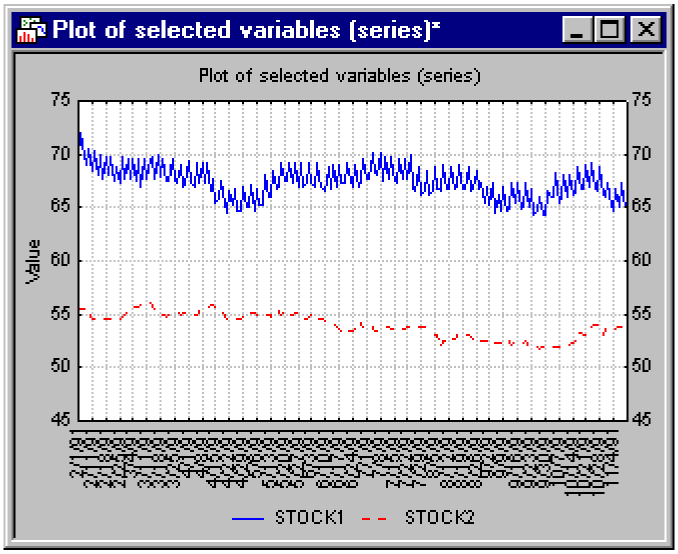
DATE (Dates)	Listing of variable: STOCK1 (St Closing quotes for stock 1 Value
2/1/91	69.75000
2/5/91	72.00000
2/6/91	70.00000
2/7/91	69.00000
2/8/91	70.50000
2/11/91	68.50000
2/12/91	70.25000
2/13/91	68.00000
2/14/91	70.00000
2/15/91	67.75000
2/18/91	69.75000
2/19/91	68.00000
2/20/91	69.75000
2/21/91	67.50000
2/22/91	68.75000
2/25/91	67.25000
2/26/91	69.75000
2/27/91	67.75000
2/28/91	69.50000

You can plot this series by clicking the *Plot* button next to the *Review highlighted variable* button on the Review & plot tab.

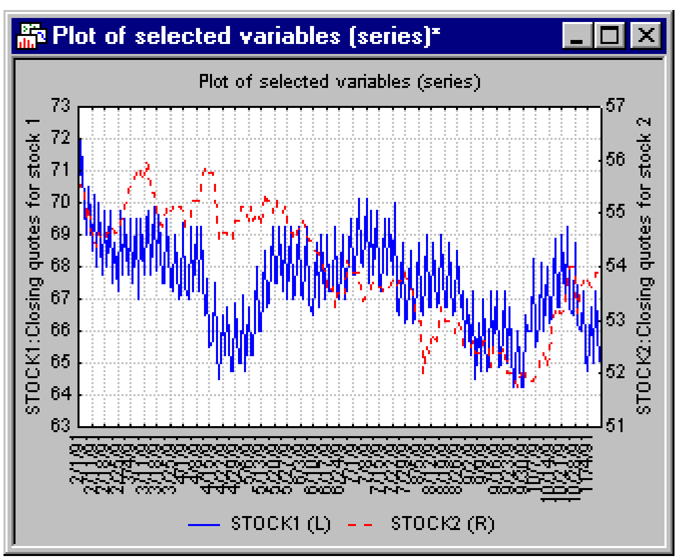


To plot both stocks simultaneously, click the *Plot* button next to the *Review multiple variables* button, and then select the variables (series) to be displayed or plotted in the

Select variables for the Spreadsheet/plot dialog. In this example, select both variables and then click the *OK* button.



Plotting Two Series with Different Scales. As you can see, the closing quotes for *Stock2* are generally lower than those for *Stock1*. You can independently scale the vertical axes for those two series to obtain the best vertical resolution possible for each series. Click the *Plot two var lists with different scales* button and in the resulting *Select variables for the Spreadsheet/plot* dialog select to plot *Stock1* against the left y-axis, and *Stock2* against the right y-axis.

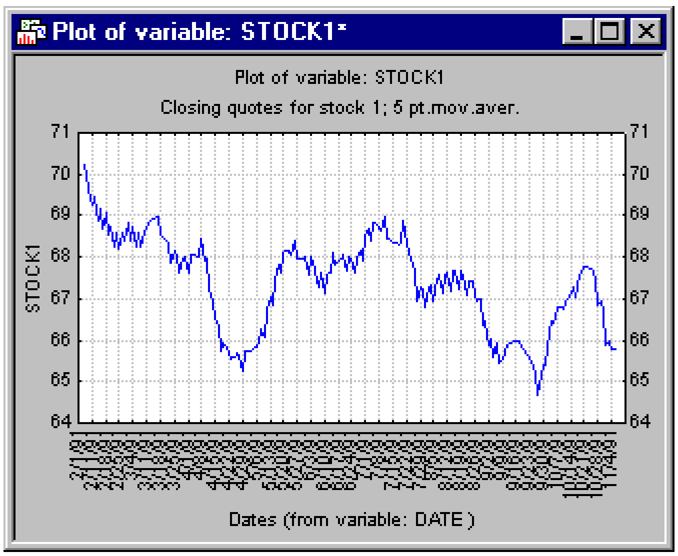


This plot allows you to compare the pattern or movement of the two series across time more clearly.

Transforming Time Series. Now perform a few transformations. On the *Transformations of Variables* dialog, highlight the first series (*Stock1*). The various tabs on this dialog show all common transformations for time series data. Some of those transformations require that you select a second variable, for example, for *Residualizing* a time series via the *x=f(x,y)* tab. For this example, we will specify a simple (unweighted) 5-point moving average transformation for series *Stock1* using the *Smoothing* tab. Select the *N-pts. mov. averg.* option button, and specify *5* as the window width in the *N=* box.

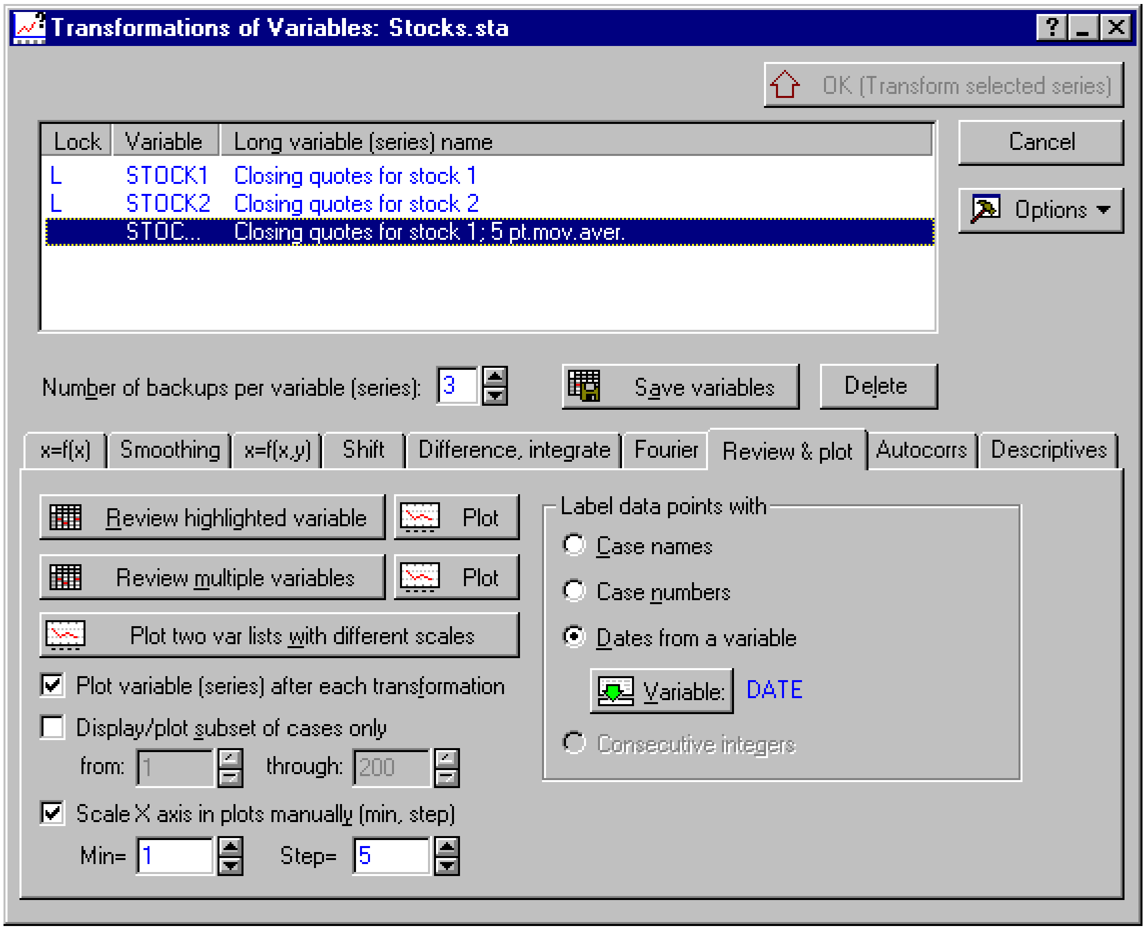
Then, click the *OK (Transform selected series)* button, and the moving average transformation will be performed. When all cases have been transformed, then, by default (i.e., if the *Plot variable (series) after each transformation* check box is selected on the *Review & plot* tab), the transformed series will be plotted.

As you can see below, compared to the plot of the raw (untransformed) series (see above), the transformed series is much less "jagged," and the general trend over the trading days reflected by the data is much clearer.



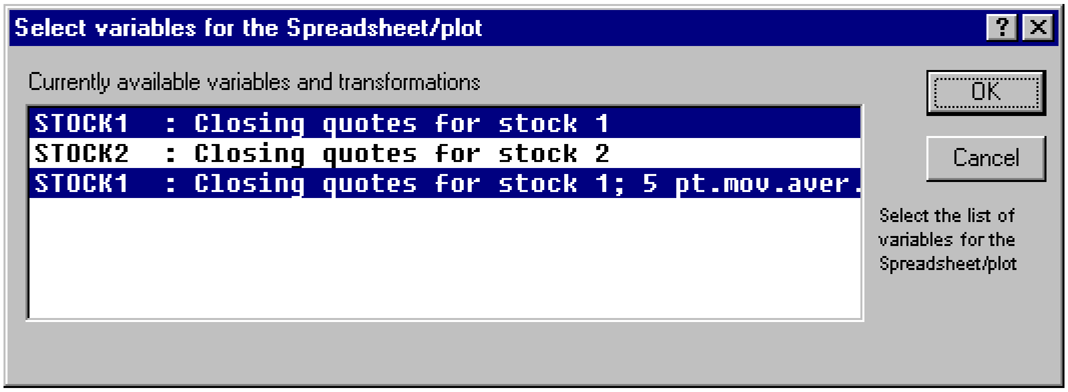
Now return to the *Transformations of Variables* dialog.

The Updated Active Work Area. As described above, the transformed (smoothed) series has been appended to the active work area.

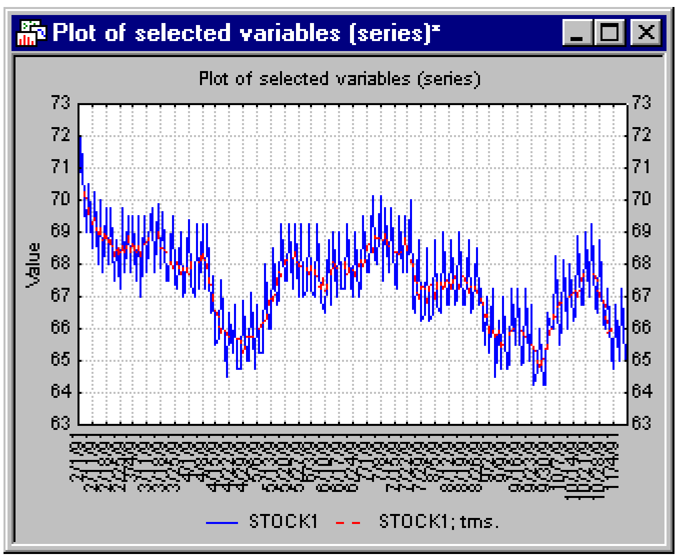


Following the naming conventions described earlier, the transformed variable has the same short name (*Stock1*) and will have the same long name, except that a brief description of the transformation (*5 pt. mov. aver.*) was added to the existing title. Note that if the original title had been much longer, so that the description of the transformation couldn't fit, then the original long variable name would have been deleted.

Further Processing of the Transformed Series. The transformed series in the work area has the same "status" as those series that were originally selected and read into the active work area from the file. For example, they can be plotted, saved, or used as input into further analyses. Now, compare the smoothed series with the original input series. On the *Review & plot tab*, click the *Plot* button next to the *Review multiple variables* button. By default, the original series and the smoothed series will be highlighted in the subsequent *Select variables for the Spreadsheet/plot* dialog.

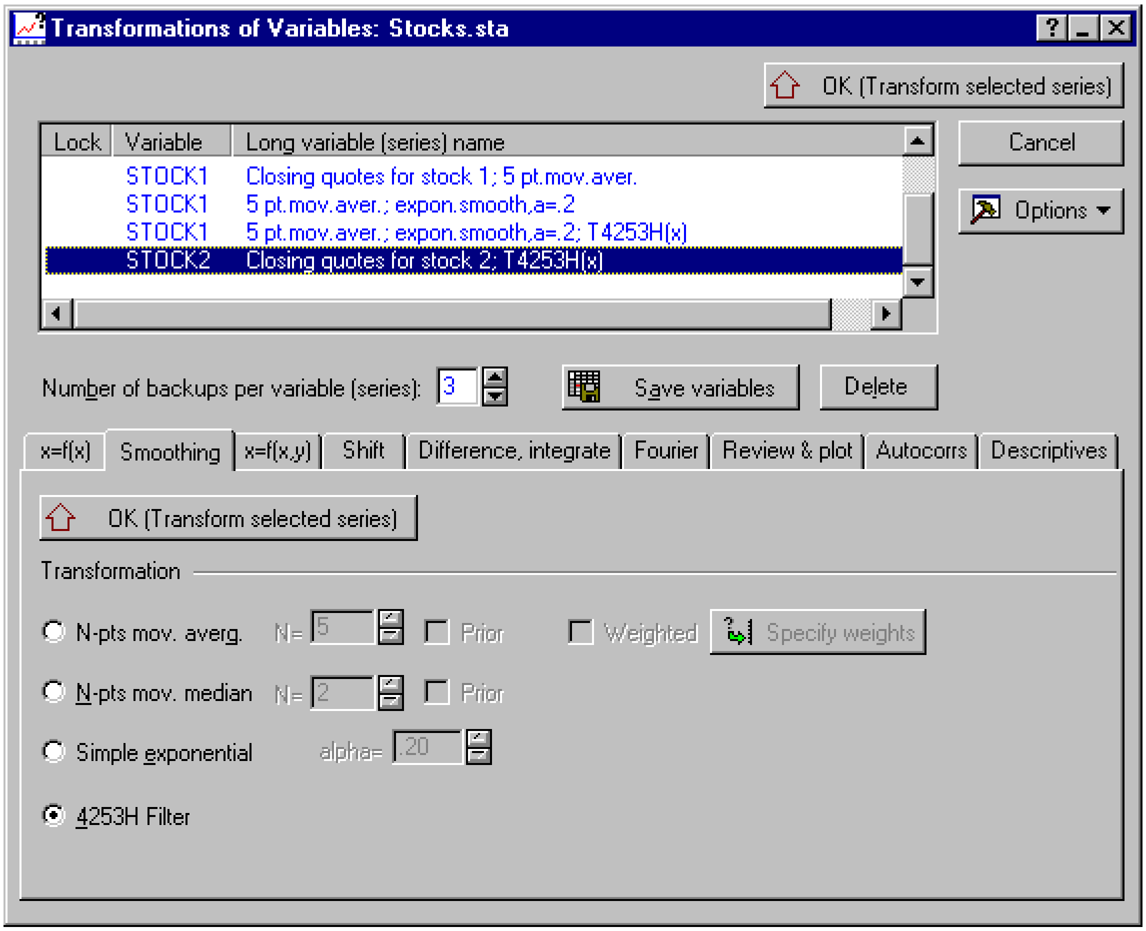


Simply accept this default selection and click the *OK* button. Shown below is the joint plot of the raw input series and the smoothed (transformed) series.

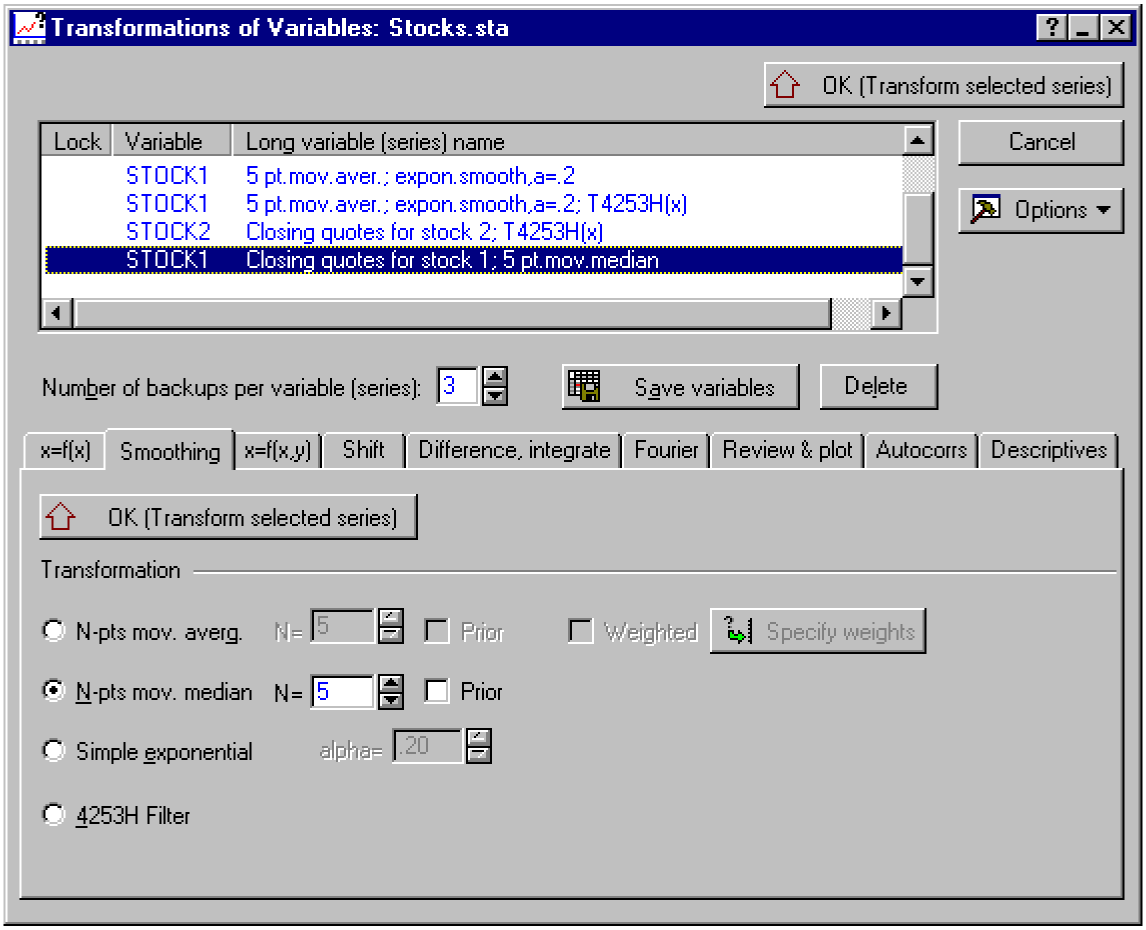


Multiple Successive Transformations. Now, continue to transform the transformed variable and observe how successive series are appended to the active work area. Highlight the transformed series (the one at the bottom of the list) by clicking on it. For now, turn off the automatic graphing option, that is, clear the *Plot variable (series) after each transformation* check box. Then, click on the Smoothing tab and select the *Simple exponential* smoothing option button with the default parameter *alpha* = *.20*. Now, click the *OK (Transform selected series)* button and then select the *4253H Filter* option button (this is a powerful smoothing/filtering technique that applies several moving average and moving median transformations in succession; refer to the Smoothing tab topic for details) and again click the *OK (Transform selected series)* button.

There are now 3 transformed series that were appended to the active work area. Assuming that you have not changed the *Number of backups per variable (series)* parameter from the default value of *3*, the active work area is now "full." The next transformation of any of the variables derived from *Stock1* will replace the "oldest" transformation for that variable. Thus, if you now transform variable *Stock2*, another (the first) backup of that variable will be added. Try this by applying the *4253H Filter* to *Stock2*. (Select the *Stock2* variable, make sure the *4253H Filter* option button is selected and then click the *OK (Transform selected series)* button.) After the transformation is complete, the active work area will look like this.



As you can see, all transformations of variable *Stock1* are still in place. However, now highlight the original variable *Stock1* again (scroll the edit window until *Stock1* is visible), and apply to it, for example, a 5-pt. mov. median transformation. (Select the *N-pts mov. median* option button on the and enter *5* in the corresponding *N=* box. Then click the *OK (Transform selected series)* button.) Now the "oldest" or first transformation that you performed on variable *Stock1* (the 5-point moving average transformation) will be replaced by the 5-point moving median transformation of series *Stock1*.



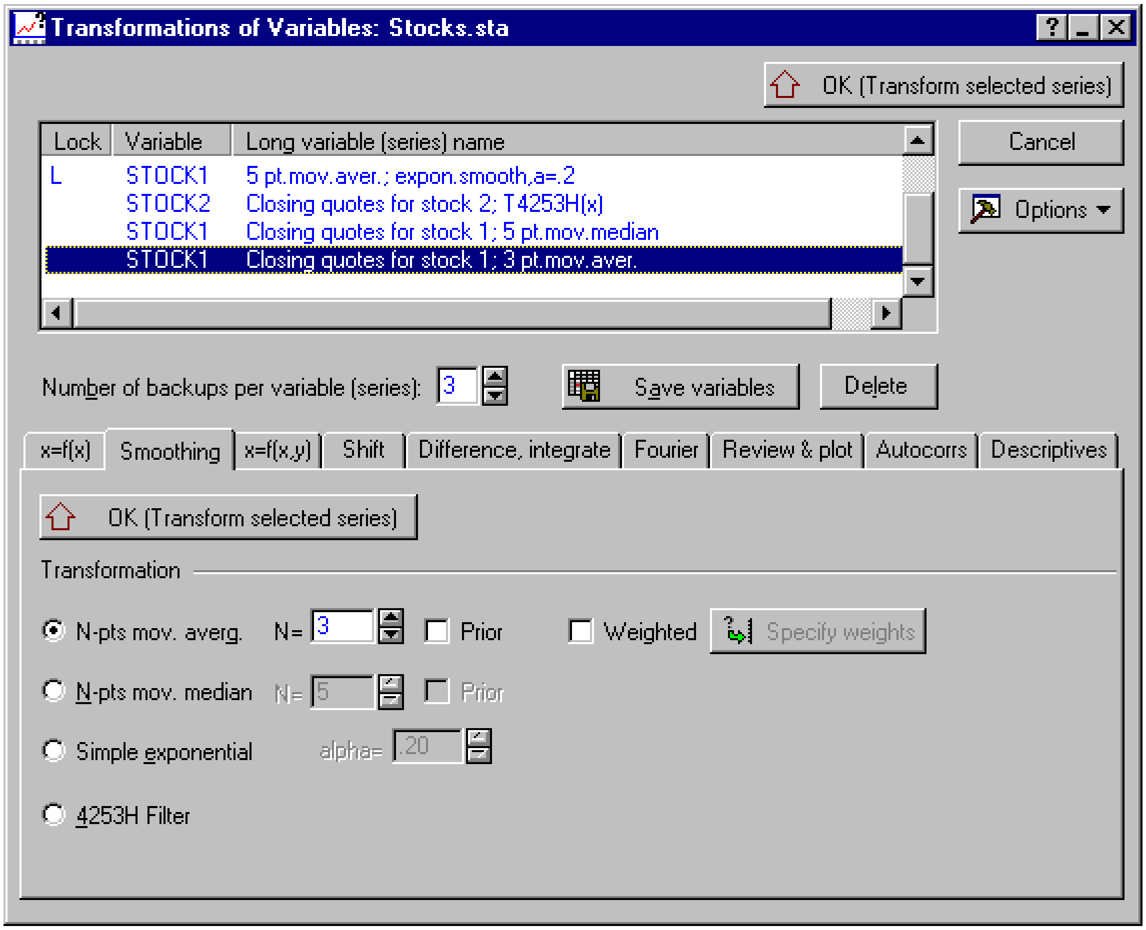
Perhaps an appropriate "mental model" of the way in which series are managed in the active work area is that of a carousel: Successive transformations of a series are placed in successive positions of the carousel. The number of places on the carousel is determined by the *Number of backups per variable (series)* parameter . Once all places have been taken, then the next transformation will replace the first transformation, as the carousel starts the next go-around.

Locking. Suppose you would like to keep a transformation in the active work area; that is, you would like to prevent it from being replaced by another transformation.

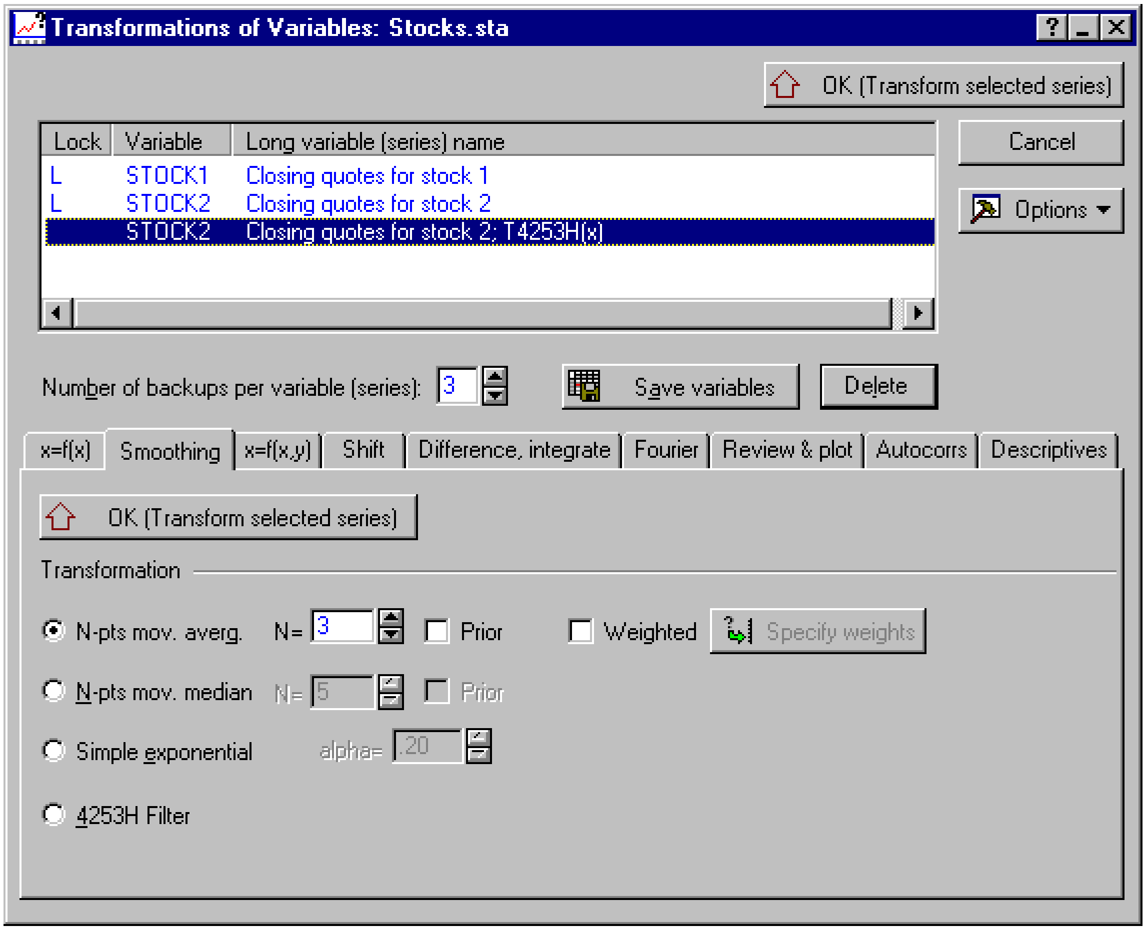
To accomplish this, the respective transformation should be *Locked*: Double-click on the respective series in the *Lock* column, and an *L* will appear in the respective row of that column. The respective series is now locked; that is, it will not be overwritten by successive transformations of the same variable, or, put another way, it will stay in the same place on the carousel.

For example, the next "oldest" transformation of *Stock1* that will be replaced is the *Exponential smoothing* transformation. Now, lock that series, then highlight the original *Stock1* series again and apply to it a 3-pts. moving average transformation. (Select the *N-pts mov. averg.* option button, enter a *3* in the corresponding *N=* box, and click the *OK*

(Transform selected series) button.) As you can see, the locked transformation was not overwritten.



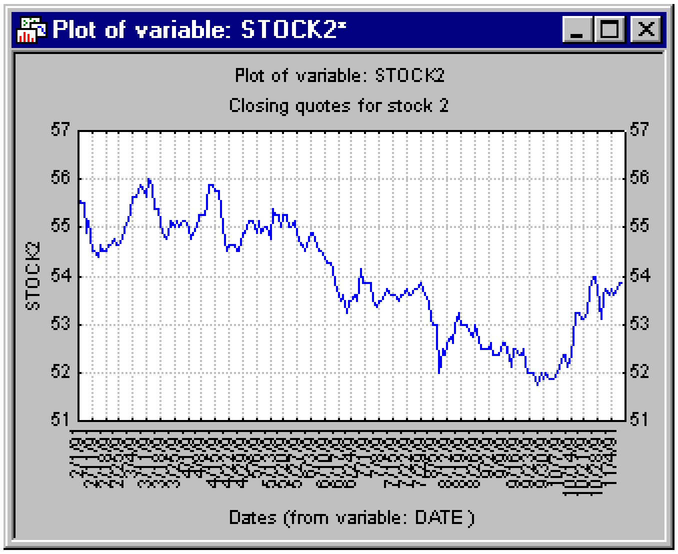
Saving the Series in the Active Work Area. Now save the transformations in the active work area. Suppose you would like to keep only the 4253H transformations of *Stock2* for further analysis. First, delete the series that you don't want to save from the active work area by highlighting them and then clicking the *Delete* button.



Then, save the variables in the active work area by clicking the *Save variables* button. You will be prompted for a file name under which *STATISTICA* will save all data currently in the active work area.

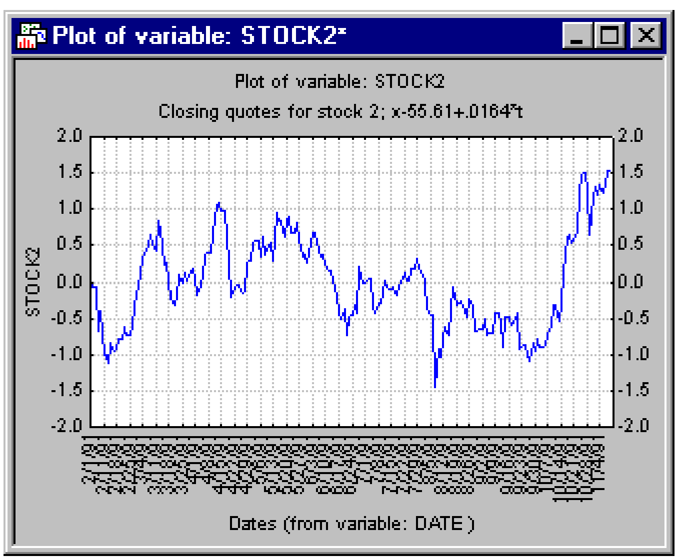
Autocorrelation Analysis. Thus far, the interpretation of the transformations has not been discussed. In general, the analysis of time series data requires a good deal of experience not only with the available techniques, but also with the nature of the data. For example, stock prices often follow what is called a random walk model. Simply stated, each observation is equal to the previous observation plus some random component. In a sense, the process behaves like "a drunken man whose position at time t is his position at time t-1 plus a step in a random direction at time t" (Wei, 1990, page 71). If so, you may expect that the simple autocorrelation is highest for a lag of 1, next highest for lag 2, etc., that is, that the autocorrelation function will show a slow decay. Put another way, the "drunken man" will be closest to where he was immediately before, a bit farther away from where he was before that, and so on. Technically, this process can be expressed as an auto-regressive process, with the autoregressive parameter (Φ in ARIMA terminology) approaching 1.0.

Plotting the Autocorrelation Function. Now examine whether the closing quotes stored in *Stock2* follow this simple model. First, plot *Stock2* (highlight *Stock2* and click the *Plot* button next to the *Review highlighted variable* button on the *Review & plot* tab).

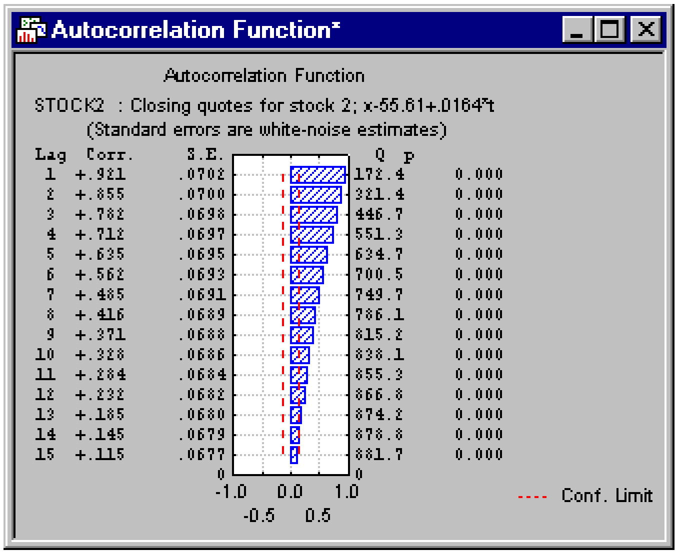


It appears that *Stock2* shows a downward trend. Such trends will bias the autocorrelation function; that is, if the stock is generally going down, then obviously, each quote will be more similar to the adjacent quotes as compared to those that are farther away.

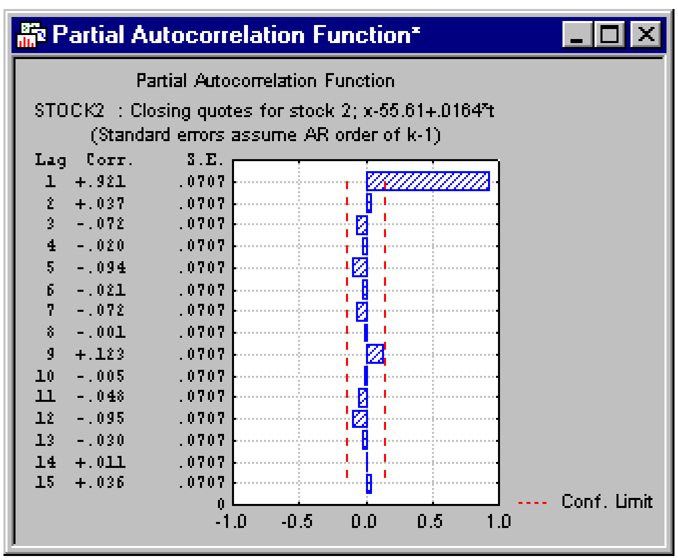
Therefore, you can detrend the series by clicking on the *x=f(x)* tab, selecting the *Trend subtract* option button, and then clicking the *OK (Transform selected series)* button. Now, if you plot the transformed variable again (by clicking the *Plot* button next to the *Review highlighted variable* button on the *Review & plot* tab), you can see that the trend was removed.



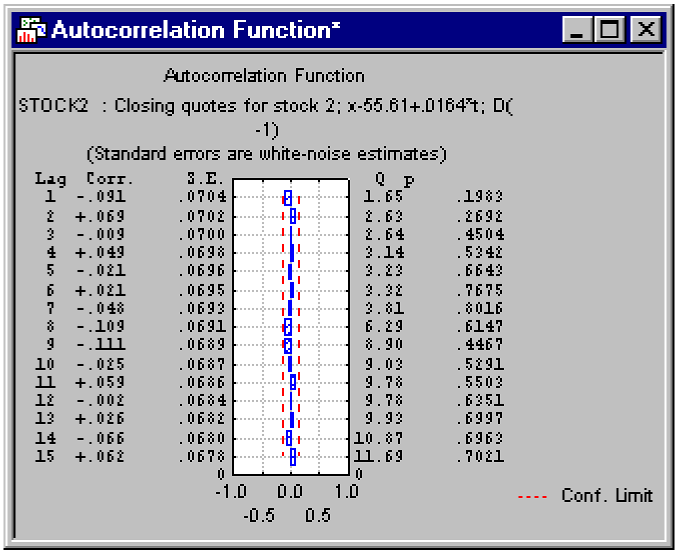
Now click the *Autocorrelations* button on the *Autocorrs tab* to display a spreadsheet and plot of the autocorrelation function.



The correlation for lag 1 is large, and decays slowly thereafter; the plot of the partial autocorrelation function also supports the random walk model. (Click the *Partial autocorrelations* button to produce this plot.)



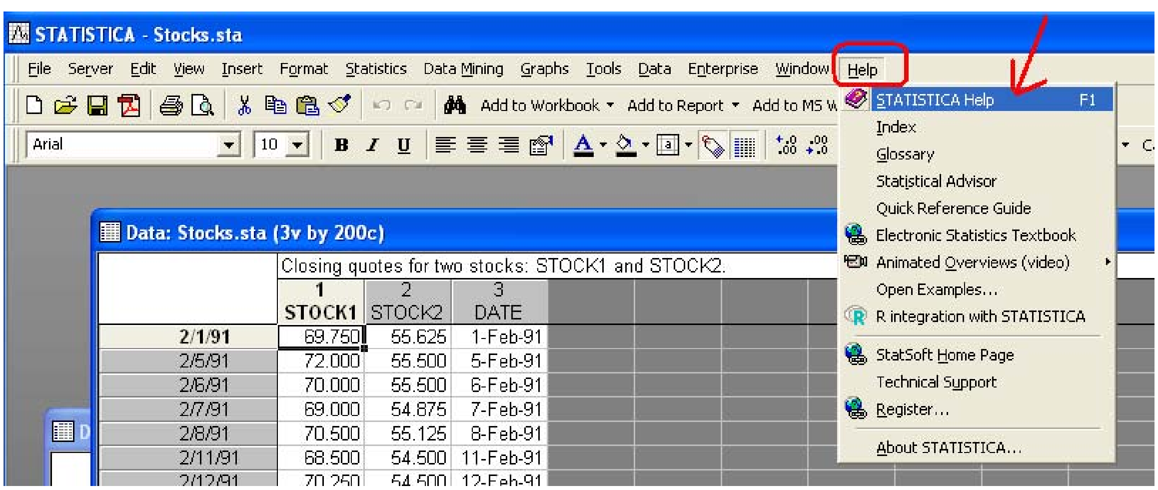
Above and beyond the very strong autocorrelation at lag 1, none of the partial autocorrelations are significant. Put into words, each observations is mostly similar to the previous observation, plus some random shock -- which represents the random walk model. You can "remove" the strong single autocorrelation by differencing the series. To do this, click on the *Difference, integrate tab*, select the simple *Differencing (x = x-x(lag))* option button, and click the *OK (Transform selected series)* button. Then click the *Autocorrelations* button on the *Autocorrs tab* to produce the plot for the differenced series.



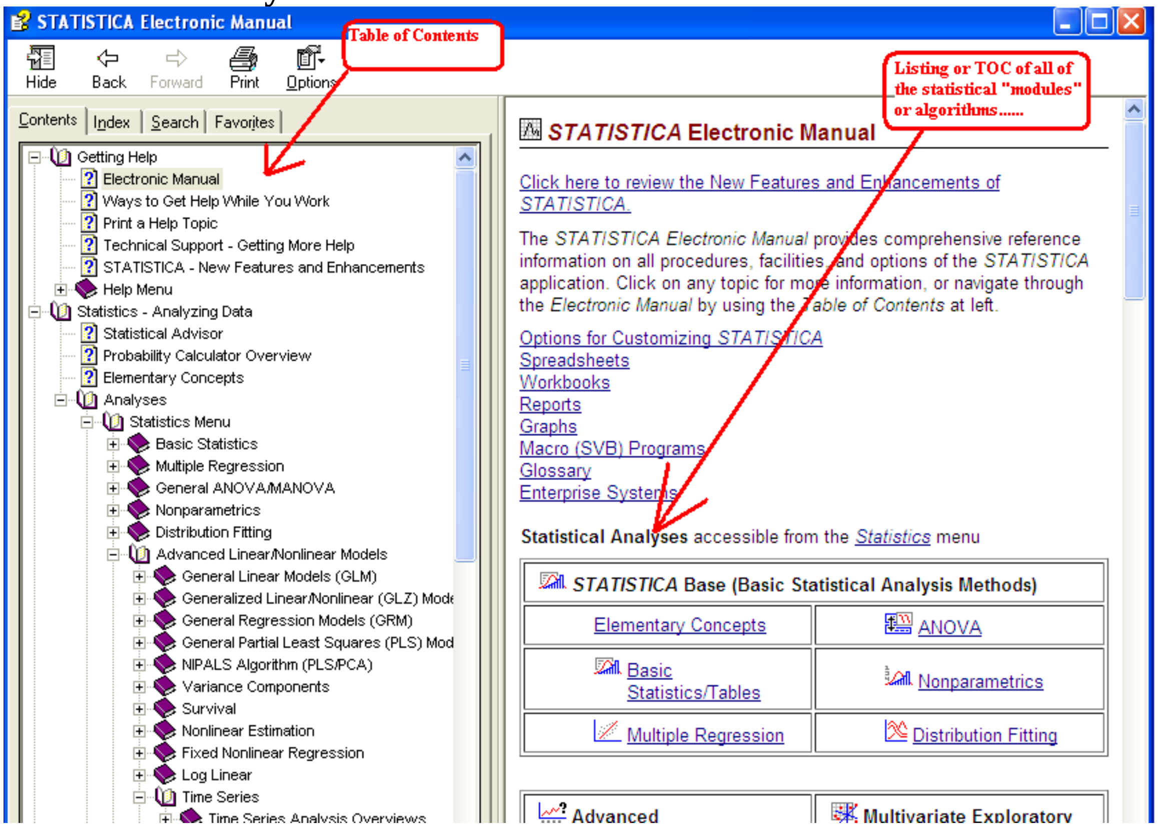
As you can see, none of the autocorrelations is significant. Thus, your initial guess from prior experience with the "behavior" of stock prices has been confirmed; *Stock2* indeed follows the random walk model, and, unfortunately, given a particular quote at a particular time, there is no way to predict whether the stock will go up or down.

End of Example 1 [Handbook Tutorial No. 43] in the HANDBOOK OF STATISTICAL ANALYSIS & DATA MINING APPLICATIONS Companion Web Page set of tutorials.

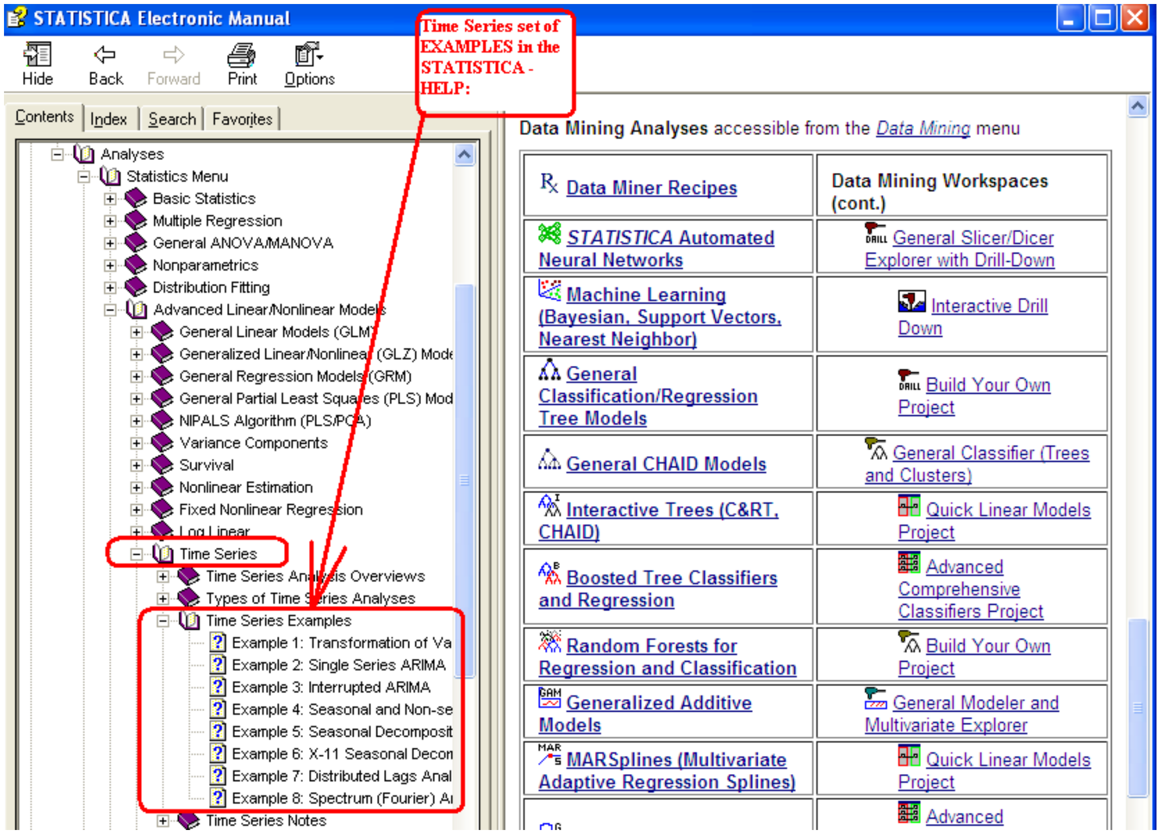
Now, you the reader, can go to the STATISTICA – HELP files [open STATISTICA on your computer, and go to HELP pull down menu and release on "STATISTICA help":



Which will take you to:



Scrolling down both the "left side panel" and the "right side panel" in the above, takes us to the full set of 8 Time Series Examples [on the left panel], and also the Data Mining module listing [on the right panel]:



Now, let's look at just the beginning paragraph of each of these remaining Time Series Examples [*e.g. Examples No 2 through No 8*]:

Example No. 2 synopsis [go to the STATISTICA HELP on your installed STATISTICA software to find the full example]:

Example 2: Single Series ARIMA

This example is based on Chapter 9 from the classic book on ARIMA by Box and Jenkins (1976). The data are the monthly passenger totals (measured in thousands) in international air travel, for twelve consecutive years: 1949-1960 (see Box and Jenkins, 1976, page 531, "Series G").

The data are partially listed below; they are also included with *STATISTICA* in example data file *Series_G.sta*. Open this data file via the *File* - *Open Examples* menu; it is in the *Datasets* folder.

Data: Series_G.sta* (1v by 144c)

	Monthly passenger totals (in 1000's) 1949-1960; Box & Jenkins, 1976; series G. 1 SERIES_G
JAN 1949	112
FEB 1949	118
MAR 1949	132
APR 1949	129
MAY 1949	121
JUN 1949	135
JUL 1949	148
AUG 1949	148
SEP 1949	136
OCT 1949	119
NOV 1949	104
DEC 1949	118
JAN 1950	115
FEB 1950	126
MAR 1950	141

The series shows a clear growing trend over the years but at the same time there is strong seasonality present in the data (e.g., March figures are usually higher than February and April)…………………

Please open STATISTICA software on your computer, and complete this example in the HELP………….[The software is supplied on the DVD bound with this 'Handbook' – see inside back cover of book]…..

Example No. 3 synopsis [go to the STATISTICA HELP on your installed STATISTICA software to find the full example]:

Example 3: Interrupted ARIMA

This example is based on data presented in McDowall, McCleary, Meidinger, and Hay (1980), who provide an excellent applications-oriented introduction to interrupted time series analysis (see also the Introductory Overview for a discussion).

The data in the example file *Director.sta* will be used. This file contains the number of monthly calls (in 100's) to Cincinnati Directory Assistance over the period from January, 1962, through December, 1976. In March 1974 (the 147'th month in the series), Cincinnati Bell initiated a new 20 cent charge for calls to Directory Assistance. This caused a marked drop-off in the number of requests for assistance, and the purpose of this analysis is to fit a model to the series that takes this abrupt change into account. Open this data file via the *File* - *Open Examples* menu. Shown below is a listing of the first few cases in the data file *Director.sta*.

Data: Director.sta (2v by...

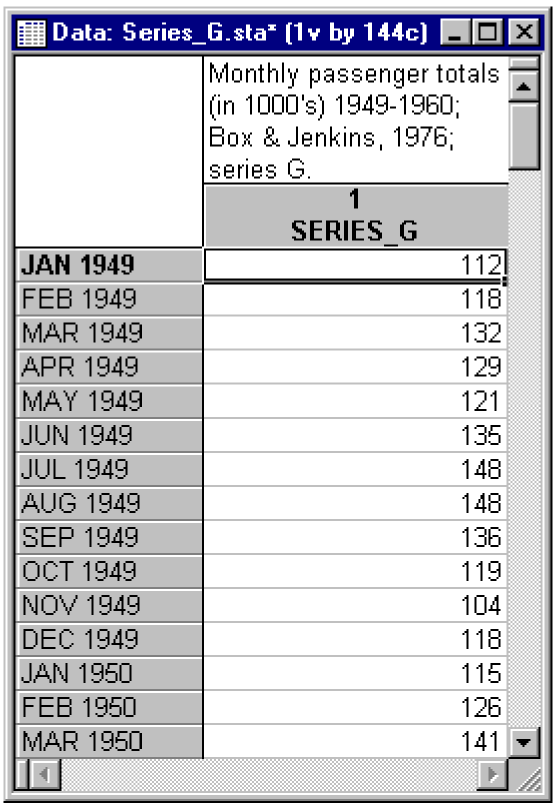
	Calls to directory a	
	1 MONTH	2 CALLS
1	Jan-1962	350
2	Feb-1962	339
3	Mar-1962	351
4	Apr-1962	364
5	May-1962	369
6	Jun-1962	331
7	Jul-1962	331
8	Aug-1962	340
9	Sep-1962	346
10	Oct-1962	341
11	Nov-1962	357
12	Dec-1962	398
13	Jan-1963	381
14	Feb-1963	367
15	Mar-1963	383
16	Apr-1963	375
17	May-1963	353

Please open STATISTICA software on your computer, and complete this example in the HELP…………[The software is supplied on the DVD bound with this 'Handbook' – see inside back cover of book]…..

Example No. 4 synopsis [go to the STATISTICA HELP on your installed STATISTICA software to find the full example]:

Example 4: Seasonal and Non-seasonal Exponential Smoothing

Example 2 discusses the analysis of a data set from the classic book on ARIMA by Box and Jenkins (1976). The data are monthly passenger totals (measured in thousands) in international air travel, for twelve consecutive years: 1949-1960 (see Box and Jenkins, 1976, page 531, "Series G"). The *Series_G.sta* data file is partially listed below. Open this data file via the *File* - *Open Examples* menu; it is in the *Datasets* folder.



The ARIMA analysis required a good deal of preparatory work during the identification stage. In fact, it usually requires a lot of experience and familiarity not only with ARIMA but also with the nature of the data, in order to identify satisfactory models. Often, the purpose of ARIMA is mostly to derive forecasts, and the interpretation of the nature of the model (i.e., the number and types of parameters) is only of secondary interest. In those cases, exponential smoothing provides a much easier alternative, one that usually produces forecasts of equal or better quality (see the Introductory Overview for a discussion of this point).

In this example, exponential smoothing will be performed on the same series used in Example 2 and the forecasts derived by the two methods will be compared.

Please open STATISTICA software on your computer, and complete this example in the HELP…………[The software is supplied on the DVD bound with this 'Handbook' – see inside back cover of book]…..

Example No. 5 synopsis [go to the STATISTICA HELP on your installed STATISTICA software to find the full example]:

Example 5: Seasonal Decomposition (Census Method I)

This example is based on a series reporting the monthly US total retail sales from 1953 to 1964. The data set is reported in Shiskin, Young, and Musgrave (1967) to illustrate the results of the X-11 (Census method II) seasonal adjustment procedure. In this example, the older Census method I seasonal decomposition method will be used to analyze this series; refer to the X-11 Seasonal Decomposition Example to see the results of the Census method II decomposition of this series.

Data File. Shown below is a partial listing of the file *Retail.sta* that contains this series (note that, as reported in Shiskin, Young, and Musgrave, 1967, these numbers are not directly comparable to the official published retail sales figures). Open this data file via the *File* - *Open Examples* menu; it is in the *Datasets* folder.

Data: Retail.sta* ...

U.S. Total retail sale, 1953 to 1964

	1 SALES	2 DATE
1	$12,903	Jan-1953
2	$12,198	Feb-1953
3	$13,711	Mar-1953
4	$14,115	Apr-1953
5	$14,520	May-1953
6	$14,442	Jun-1953
7	$14,250	Jul-1953
8	$14,045	Aug-1953
9	$13,952	Sep-1953
10	$14,819	Oct-1953
11	$13,828	Nov-1953
12	$16,314	Dec-1953
13	$12,213	Jan-1954
14	$11,948	Feb-1954
15	$13,576	Mar-1954
16	$14,025	Apr-1954
17	$14,116	May-1954
18	$14,533	Jun-1954
19	$14,259	Jul-1954

Please open STATISTICA software on your computer, and complete this example in the HELP…………[The software is supplied on the DVD bound with this 'Handbook' – see inside back cover of book]…..

Example No. 6 synopsis [go to the STATISTICA HELP on your installed STATISTICA software to find the full example]:

Example 6: X-11 Seasonal Decomposition (Census Method II)

This example is based on a series reporting the monthly US total retail sales from 1953 to 1964. The data set is reported in Shiskin, Young, and Musgrave (1967) to illustrate the results of the X-11 (Census method II) seasonal adjustment procedure. Note that the same example series is used in the Seasonal Decomposition Example; there, the older Census method I seasonal decomposition method was used to analyze this series.

Data File. Shown below is a partial listing of the file *Retail.sta* that contains this series (note that, as reported in Shiskin, Young, and Musgrave, 1967, these numbers are not directly comparable to the official published retail sales figures). Open this data file via the *File* - *Open Examples* menu; it is in the *Datasets* folder.

Data: Retail.sta* ...

U.S. Total retail sale, 1953 to 1964

	1 SALES	2 DATE
1	$12,903	Jan-1953
2	$12,198	Feb-1953
3	$13,711	Mar-1953
4	$14,115	Apr-1953
5	$14,520	May-1953
6	$14,442	Jun-1953
7	$14,250	Jul-1953
8	$14,045	Aug-1953
9	$13,952	Sep-1953
10	$14,819	Oct-1953
11	$13,828	Nov-1953
12	$16,314	Dec-1953
13	$12,213	Jan-1954
14	$11,948	Feb-1954
15	$13,576	Mar-1954
16	$14,025	Apr-1954
17	$14,116	May-1954
18	$14,533	Jun-1954
19	$14,259	Jul-1954

Please open STATISTICA software on your computer, and complete this example in the HELP…………[The software is supplied on the DVD bound with this 'Handbook' – see inside back cover of book]…..

Example No. 7 synopsis [go to the STATISTICA HELP on your installed STATISTICA software to find the full example]:

Example 7: Distributed Lags Analysis

Overview and Data File. This example is based on data published by the US Education Department. The file *Teachers.sta* contains data describing;

1. The number of students enrolled in public schools (variable *Children*),
2. The number of public school teachers (*Teachers*), and
3. The average salary of public school teachers (*Salary*).

These data are presented for the period from 1900 through 1980, in ten-year intervals. Open this data file via the *File* - *Open Examples* menu; it is in the *Datasets* folder.

Data: Teachers.sta* (3v by 9c)

Number of public school teachers and their salary as a function of enrollment

	1 CHILDREN	2 TEACHERS	3 SALARY
1900	15503110	423062	$325
1910	17813852	523210	$485
1920	21578316	679302	$871
1930	25678015	854263	$1,420
1940	25433542	875477	$1,441
1950	25111427	913671	$3,010
1960	36086771	1355000	$5,174
1970	45909088	2061115	$9,570
1980	40984093	21835000	$17,600

It is reasonable to assume that the number of teachers employed in the public schools will be a function of the number of students that are in the schools. However, you may expect some lag in the relationship. When, due to demographic changes, there are many students, there will be more hiring of teachers; however, it will take some time to "produce" those teachers. Greater demand for teachers should also drive up the salaries, again, probably with some lag.

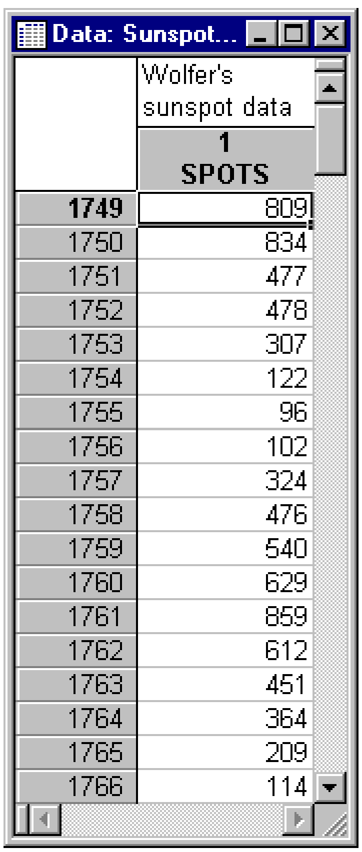
Please open STATISTICA software on your computer, and complete this example in the HELP…………[The software is supplied on the DVD bound with this 'Handbook' – see inside back cover of book]…..

Example No. 8 synopsis [go to the STATISTICA HELP on your installed STATISTICA software to find the full example]:

Example 8: Spectrum (Fourier) Analysis

The Introductory Overview discusses two very simple examples (based on Shumway, 1988) to illustrate the nature of spectrum analysis and the interpretation of results. If you are not familiar with this technique, it is recommended that you first review that section of the Introductory Overview.

Overview and Data File. File *Sunspot.sta* contains part of the famous Wolfer sunspot numbers for the years 1749 through 1924 (Anderson, 1971). Open this data file via the *File* - *Open Examples* menu; it is in the *Datasets* folder. Shown below is a listing of the first few cases in the example file.



	Wolfer's sunspot data 1 SPOTS
1749	809
1750	834
1751	477
1752	478
1753	307
1754	122
1755	96
1756	102
1757	324
1758	476
1759	540
1760	629
1761	859
1762	612
1763	451
1764	364
1765	209
1766	114

The number of sunspots are believed to affect the weather on earth, and thus human activities such as agriculture, telecommunications, etc. In this analysis, you will try to find out whether sunspot activity is cyclical in nature (which it is, this data set is widely discussed in the literature; see, for example, Bloomfield, 1976, or Shumway, 1988).

Please open STATISTICA software on your computer, and complete this example in the HELP…………[The software is supplied on the DVD bound with this 'Handbook' – see inside back cover of book]…..