436 stat

Time Series Analysis

Dr. Abdulhamid A. Alzaid

Office: 2B43

Office hours: Sun, Mon, Tues (9-10)

Textbooks

1- Time Series Analysis, by J. Cryer and k. Chan (2008). Springer

2- The Analysis of Time Series, by C. Chatfield (2003). Chapman and Hall.

Grading Scheme

Tutorial	10%
(attendance, participation, homework, short exams)	
Midterm I	10%
Midterm II	10%
Homework and data analysis reports	10%
Data analysis exam	20%
Final Exam	40%

Chapter 1 :Introduction

Q: What is a time series?

A time series is a collection of observations of some phenomenon collected sequentially over a period of time. For example, volume of rain over months of the year, number of daily accidents in Saudi Arabia, value of quarterly foreign remittances(لتحويلات المالية), and so on). This means that data have chronological (تسلسل زمني) order.

There are many examples of time series in many fields of knowledge it can be found in Agriculture - Medicine - Economics - Engineering - Education and others. Therefore, the methods used in time series analysis play an important role in the science of statistics.

Example 1: Figure 1.1 illustrates the profit gain of a company over a period of 50 years.

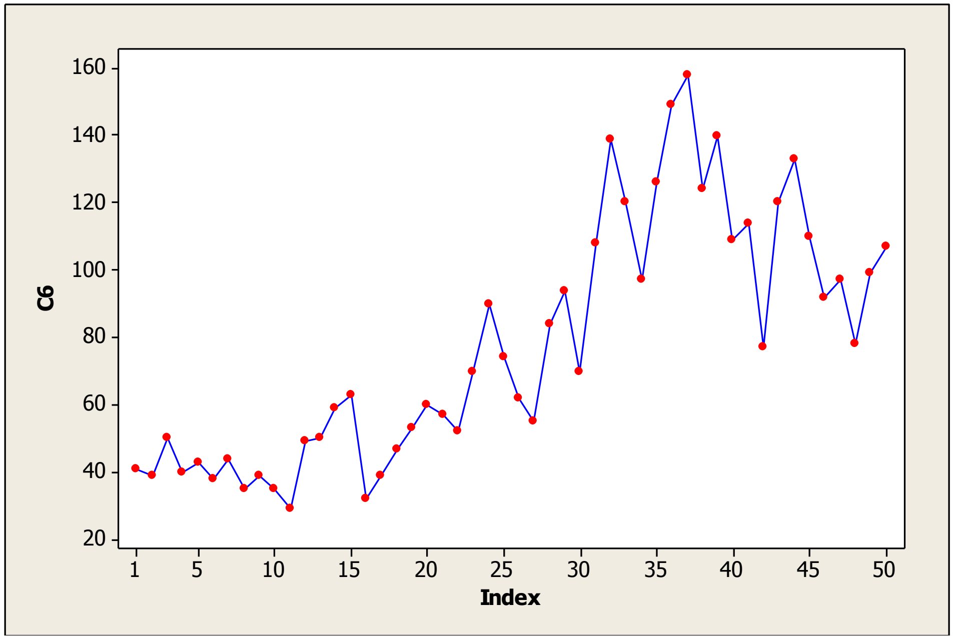


Figure 1.1 The profit gain of a company over a period of 50 years

Example 2: Figure 1.2 illustrates the average monthly temperatures in a city during a period of 6 years.

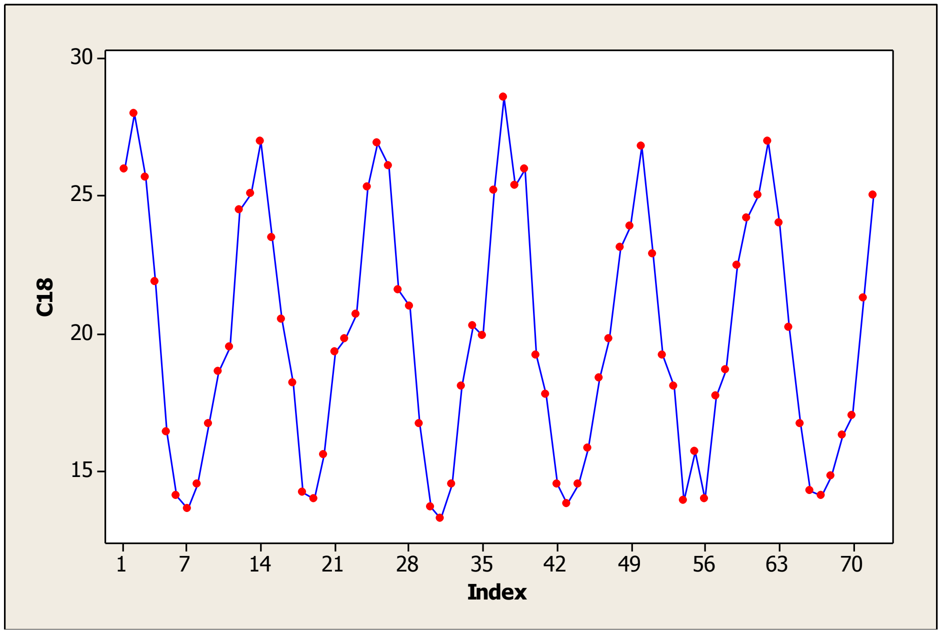


Figure (1.2): average monthly temperatures in a city during a period of 6 years

Example 3: Figure 1.3 illustrates the monthly sales for some industrial piece during a period of 15 years

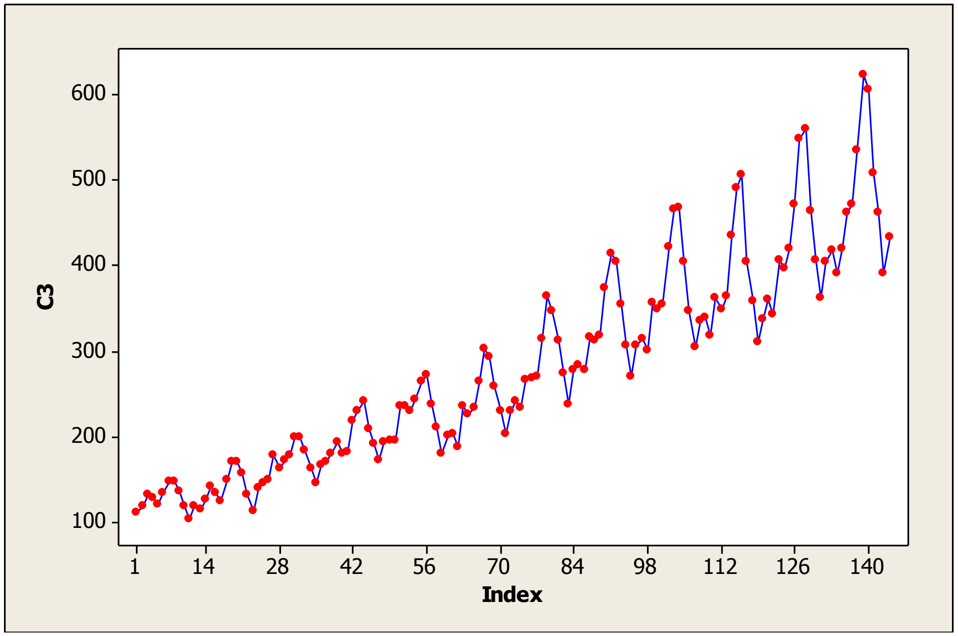


Figure (1.3): monthly sales for some industrial piece during a period of 15 years

Researchers might be interested for example in predicting the future sales so that proper planning of production can be accomplished, or even investigating the relation between sales series and any other series such as advertising and others.

1.2 Some used terminology

A time series is said to be continuous, when observations are taken in a continuous manner over time, and to be discrete when observations are taken at specific times (usually at equal intervals). In this course we will be interested in discrete time series.

As we know, most of the statistical theory, which we have already studied is interested in studying random samples that in which observations are independent. But as we have seen from the above examples, the nature of time series indicates that the observations are not independent. Therefore, statistical analysis to be used for the analysis must take into consideration the chronological (or spatial(المكانية)) order of the observations.

When observations are not independent of each other, then it is possible to predict future values of the series using the previous values.

If it is possible to predict the future with complete accuracy, then the series is called deterministic. However, most of the time series are stochastic and therefore completely accurate predictions are not possible.

Goals of time series analysis

There are several goals for the analysis of time series, some of which are:

1- Description

Time series analysis is used to describe and portray (تصوير) the available information that shows how the studied phenomenon evolve (تتطور) over time. That is, describe the main features of the time series, which will help in determining the best mathematical model that can be appropriate to achieve the other goals of the analysis, and get to know the upward and downward movements in the time series and to identify the major components such as trend and seasonal changes. So when analyzing any time series, the first step must be carried out is to plot the time series as we have seen in the previous examples and get some descriptive characteristics.

For example, in Figure (1.3), we notice the existence of strong seasonal effects, as sales increase in the middle of the year, and decreases in the ends. It also seems that annual sales increase from year to year (i.e. there is a growing trend, so for some series, description of the observations can be achieved through a simple model that includes trend component and seasonal component. However, some series may need a more complicated models.

2- Interpretation

Interpretation means explaining the changes occurring in the phenomenon using other time series that are related to it, or by using environmental factors affecting the phenomenon, for example,

one can study how the sea level is affected by temperature, or how sales are affected by advertising.

3- Control

In production lines (in the factories), one may get time series that designate the product quality in the manufacturing process, and the goal here might be to control product quality so that it does not go below a specified level.

4- Forecasting

Forecasting is considered one of the most important goals of time series analysis. As one might want to know or expect the future values of a time series.

Analysis of time series usually starts by identifying an appropriate model that explains the evolution pattern (نمط التطور) of the series, and then uses the model to extrapolate this pattern into the future.

The main assumption here is that **this pattern will continue in the near future**. It should be noted that any forecasting method will not give good forecasting results if the pattern did not continue in the future, so it is always advisable to restrict forecasting to the near future, and update the forecasts as new observations become available.

Measuring forecasting errors

Usually a time series is studied for the purpose of finding out the evolution pattern of the historical values of the phenomenon and then use this pattern to forecast the future values. However, any future forecast will contain a certain amount of uncertainty, this could be reflected by adding an error component in the forecasting model.

Error component is one representing factors that cannot be explained by the typical or regular components in the model. Of course, whenever the error component is small, this will increase our ability to forecast accurately, and vice versa.

If we assume that the value of the phenomenon at time t is y_t , and that our forecast at time t is $\hat{y}_t$, then forecast error at time t is defined as:

$$\varepsilon_t = \hat{y}_t - y_t, \quad t = 1,2,\ldots,n$$

Where n is the length of the series (i.e. no. of observations in the series).

Examining successive forecasting errors ε_t reveals how good is the forecasting model. As we know from regression analysis, a good model must produce errors that are random, i.e. errors that are free of any systematic changes, as shown in the following figure:

If these errors are acceptable, so that the forecasting method is considered appropriate then we should measure the size of these errors. There are some measures of error size, the most important are:

a. Mean Absolute Deviation (MAD):

It is defined as,

$$MAD = \frac{1}{k}\sum_{i=1}^{k}|\varepsilon_t|$$

$$= \frac{1}{k}\sum_{i=1}^{k}|y_{t-}\ \hat{y}_t|$$

MAD measures the deviations in the same units as the original data.

b. Mean Absolute Percentage Error (MAPE):

This measure finds out how accurate is the model fitted to the data, it is given as,

$$MAPE = \frac{100}{k} \sum_{i=1}^{k} \left| \frac{y_{t-} \hat{y}_t}{y_t} \right|$$

$$= \frac{100}{k}\sum_{i=1}^{k}\left|\frac{\varepsilon_t}{y_t}\right|$$

It gives the forecasting errors as a percentage, this provide us with a tool to compare different models, and their forecasting ability.

c. Mean Squared Deviation (MSD):

$$MSD = \frac{1}{k}\sum_{i=1}^{k}(\varepsilon_t)^2$$

$$= \frac{1}{k}\sum_{i=1}^{k}(y_{t-}\ \hat{y}_t)^2$$

This measure is similar to the usual measure MSE (mean squared error), but it is better in comparing the different models, because the MSE uses in the denominator $(n-r)$ degrees of freedom, where r represent the number of estimated parameters in the models, which change with the used model, whereas, MSD uses in the denominator (k) degrees of freedom (i.e. the number of obtained forecasts), which does not change with the model. Also note that MSD gives more weight for large errors as it squares them.

In all the measures above, we choose the model that produce the lowest values for MAD, MSD, MAPE.

Choosing the appropriate method for forecasting

Choosing the appropriate method of forecasting is one of the most important steps in the analysis of time series, which is not an easy task, and requires experience, skills, and employing the appropriate statistical methods for the data, but generally it depends on many factors including:

A) Minimizing forecasting errors, which is the first criteria analyst should pay attention to, these are measured through the three criteria mentioned above.

B) Quality of required forecast. If a point forecast is required, then using simple traditional methods will be enough to achieve the goal. Whereas, if we require to estimate interval forecast and to evaluate it through test of hypothesis, then more sophisticated methods should be employed, such as BOX-Jenkins methods.

C) Cost of used statistical methodology and availability of relevant statistical software.

D) Extent to which theoretical assumptions upon which forecasting model rely are satisfied. This is a very important consideration and should be checked.

Which means that the best forecasting method is not necessarily the method that achieves the highest accuracy or the smallest forecasting errors, but one method may be used because of type of the required forecast, another because of only small number of observations are available, a third because it has a low cost, and a fourth because its theoretical assumptions comply with the data set in hand.

Forecasting methods

It is possible to identify two main forecasting methods:

1- Regression approach

This approach is based on identifying the variable(s) that may have a causal relationship with the variable under study that we want to predict, this variable is called the dependent variable, then determine the appropriate statistical model or appropriate functional relationship which explains how the dependent variable is associated to the independent or explanatory variables. Using this model, we can predict the dependent variable under study. The main disadvantages of this approach are:

1- Difficulty of identifying all the explanatory variables that are related to the dependent variable.

2- Requires the availability of detailed historical information about all the explanatory variables, and the ability of knowing these variables or predicting them.

2- Time series approach

This approach relies on analyzing historical data of the variable under study in order to determine the pattern it follows. Assuming that this pattern will continue in the future, we use it to predict future values of the variable. Time series models are divided into three major types:

a) deterministic models

b) ad hoc methods

c) stochastic time series models

- **Deterministic models:**

As we know from our study in statistics that the mean model can be expressed in the following general form:

$$y_t = E(y_t) + \varepsilon_t,$$

where ε_t are uncorrelated random variables with mean equal to zero and a constant variance, this model is called deterministic if we are able to express $E(y_t)$ as a direct function of time t , and let it be $f(t, \boldsymbol{\beta})$, where the vector $\boldsymbol{\beta}$ denote the parameters of this function. In this case it is possible to express the observations of the time series y_t in the form:

$$y_t = f(t, \boldsymbol{\beta}) + \varepsilon_t , \qquad t = 1,2, \ldots , n$$

which means that future values of the series can be expressed in the form:

$$y_h = f(h, \boldsymbol{\beta}), \qquad h = t+1, t+2, \ldots$$

This indicate that future values of the series takes on a deterministic form, i.e. a non-random form $f(h, \boldsymbol{\beta})$. These models are based on two main assumptions:

1) The function $f(t, \boldsymbol{\beta})$ is a deterministic nonrandom function.
2) ε_t are uncorrelated random variables with mean zero and a constant variance.

These assumptions indicate that the variables $y_1, y_2, \dots, y_n$ are uncorrelated. Examples of mathematical functions used in these models are the polynomials, exponential functions, and trigonometric functions.

The deterministic models have some disadvantages:

1) These methods focus on mathematical logic in trying to find a suitable mathematical function that can be used to fit the data more than trying to discover the important statistical features of the series,

and the most important feature is their correlation structure. So they are just models to regenerate the observations $y_1, y_2, \ldots, y_n$.

2) These models assume that the long-term evolution of the series is systematic and regular so that it can be predicted very accurately.

3) These models also assume that the observations are not correlated, which is rarely true in different application areas.

Because of all these disadvantages, the deterministic models usually produce statistically less accurate forecasts.

- **Ad hoc methods**

These methods rely on expressing the forecast of the series at time t in terms of the current value y_t, and its past values $y_1, y_2, \dots, y_{t-1}$. So if we assume that t represents a certain origin point, and that we want to predict the value of the series after k time intervals, then this approach indicates using the following functional relationship:

$$\hat{y}_{t+k} = f(y_1, y_2, \dots, y_{t-1}, y_t)$$

Many ways exist to carry out such predictions, such as moving averages method, and exponential smoothing methods.

a) Simple Moving Average

This method uses the most recent k values of the series to

predict next value :

$$\hat{y}_{t+1} = \frac{1}{k}\left[y_t + \ y_{t-1} + \ \dots + \ y_{t-(k-2)} + y_{t-(k-1)}\right], \quad t = k, k+1, \dots, n$$

this means that:

$$\hat{y}_{t+2} = \frac{1}{k}\left[y_{t+1} + \ y_t + \ \dots + \ y_{t-(k-2)}\right]$$

That is, to find a simple moving average $\hat{y}_{t+2}$ we use the same values

used in finding the previous mean $\hat{y}_{t+1}$ after replacing the older value

$y_{t-(k-1)}$ with the most recent one y_{t+1} , and this what gave this procedure its name, moving average, because always the mean is updated by dropping the oldest observation and adding a new one. For example for $k = 3$, we can form a simple moving average as follows:

$$\hat{y}_4 = \frac{1}{3}[y_3 + \ y_2 + y_1]$$

$$\hat{y}_5 = \frac{1}{3}[y_4 + y_3 + \ y_2]$$

$$\hat{y}_6 = \frac{1}{3}[y_5 + \ y_4 + y_3]$$

$$\vdots$$

$$\hat{y}_n = \frac{1}{3}[y_{n-1} + \ y_{n-2} + y_{n-3}]$$

Choosing the right value for k depends on the experience of the researcher. Indeed, it is one of the difficulties of using simple moving average method. Another problem is in assigning equal weights for all observations, for example for $k = 8$, the weight given to the most recent value y_t is equal to the oldest value y_{t-7}, which contradicts with properties of time series, as it is more logical to assign larger weights to the most recent observations, that’s why it is preferred to use simple moving averages in forecasting when the observed time series is random in nature.

Example: For the following data, calculate a moving average of order $k = 3$:

355, 451, 435, 558, 556, 573, 565, 608

solution:

$$ma_1(3) = \frac{y_3 + y_2 + y_1}{3} = \frac{435 + 451 + 355}{3} = 419.68$$

$$ma_2(3) = \frac{y_4 + y_3 + y_2}{3} = \frac{558 + 435 + 451}{3} = 481.33$$

In the same manner, we get,

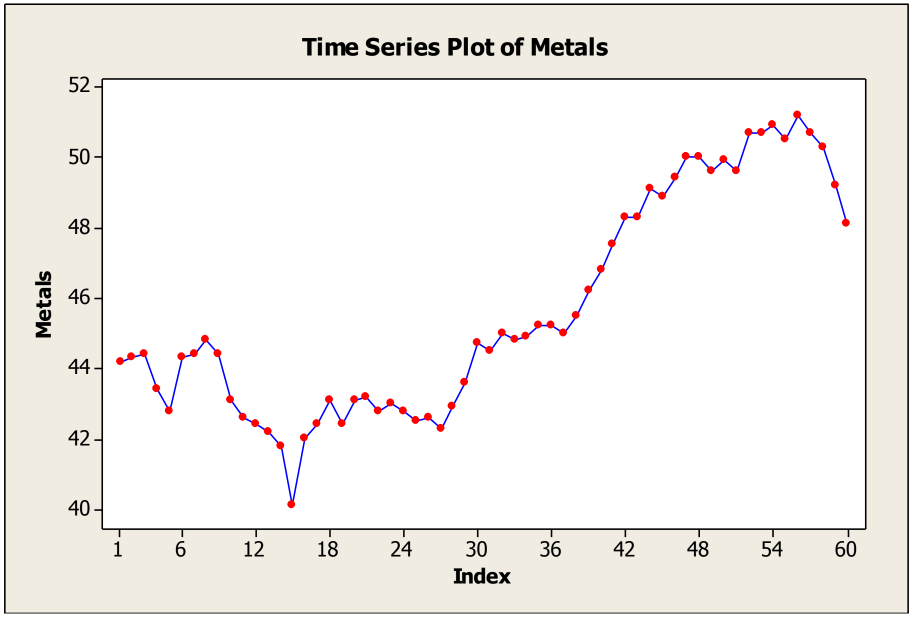
$$ma_3(3) = 516.33, ma_4(3) = 562.33, \quad ma_5(3) = 582,$$

$$ma_6(3) = 626.33$$

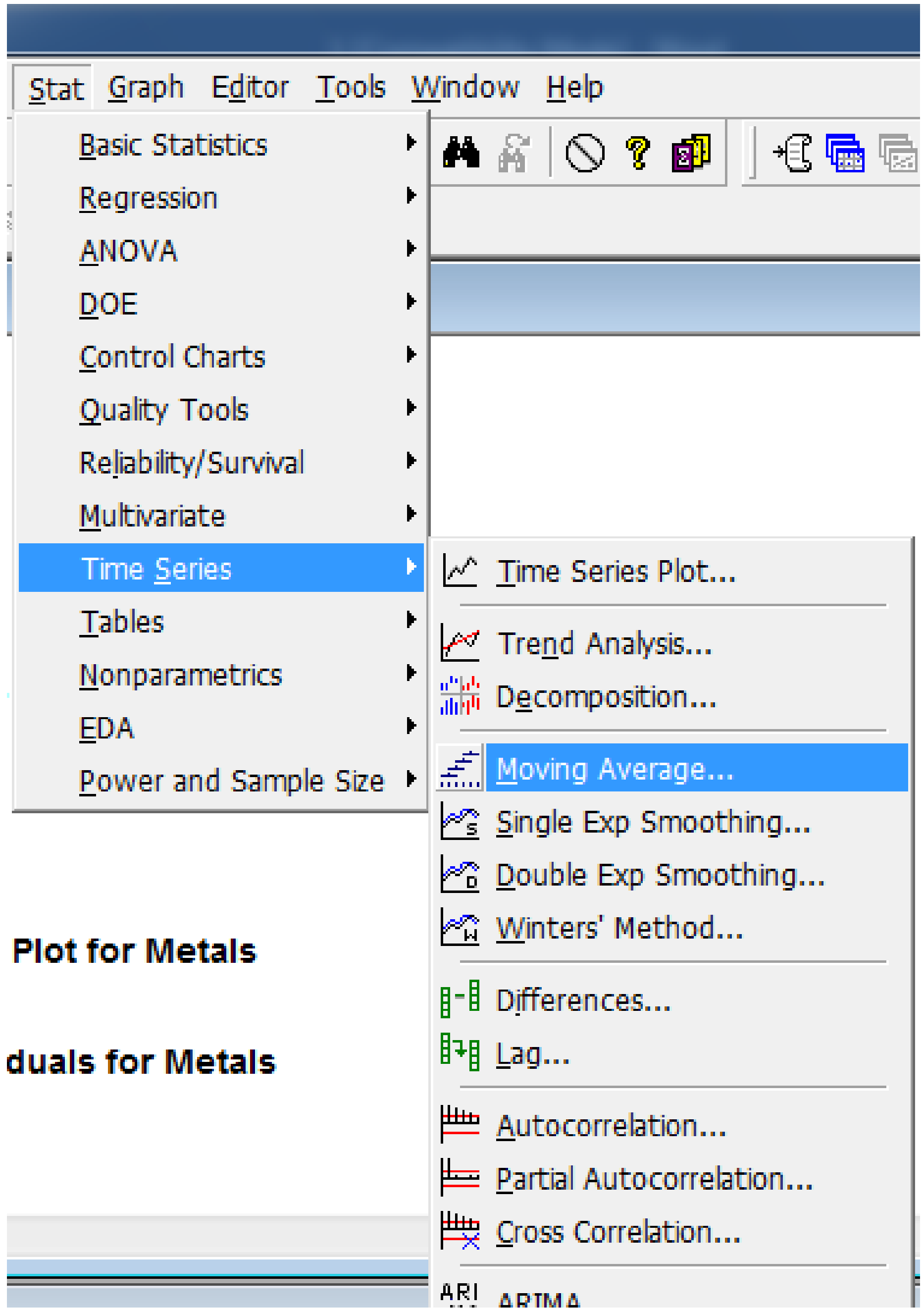
Example: In MINTAB program, open data file “EMPLOY.MTB”, Use data Variable (Metals):

44.2 44.3 44.4 43.4 42.8 44.3 44.4
44.8 44.4 43.1 42.6 42.4 42.2 41.8
40.1 42.0 42.4 43.1 42.4 43.1 43.2
42.8 43.0 42.8 42.5 42.6 42.3 42.9
43.6 44.7 44.5 45.0 44.8 44.9 45.2
45.2 45.0 45.5 46.2 46.8 47.5 48.3
48.3 49.1 48.9 49.4 50.0 50.0 49.6
49.9 49.6 50.7 50.7 50.9 50.5 51.2
50.7 50.3 49.2 48.1

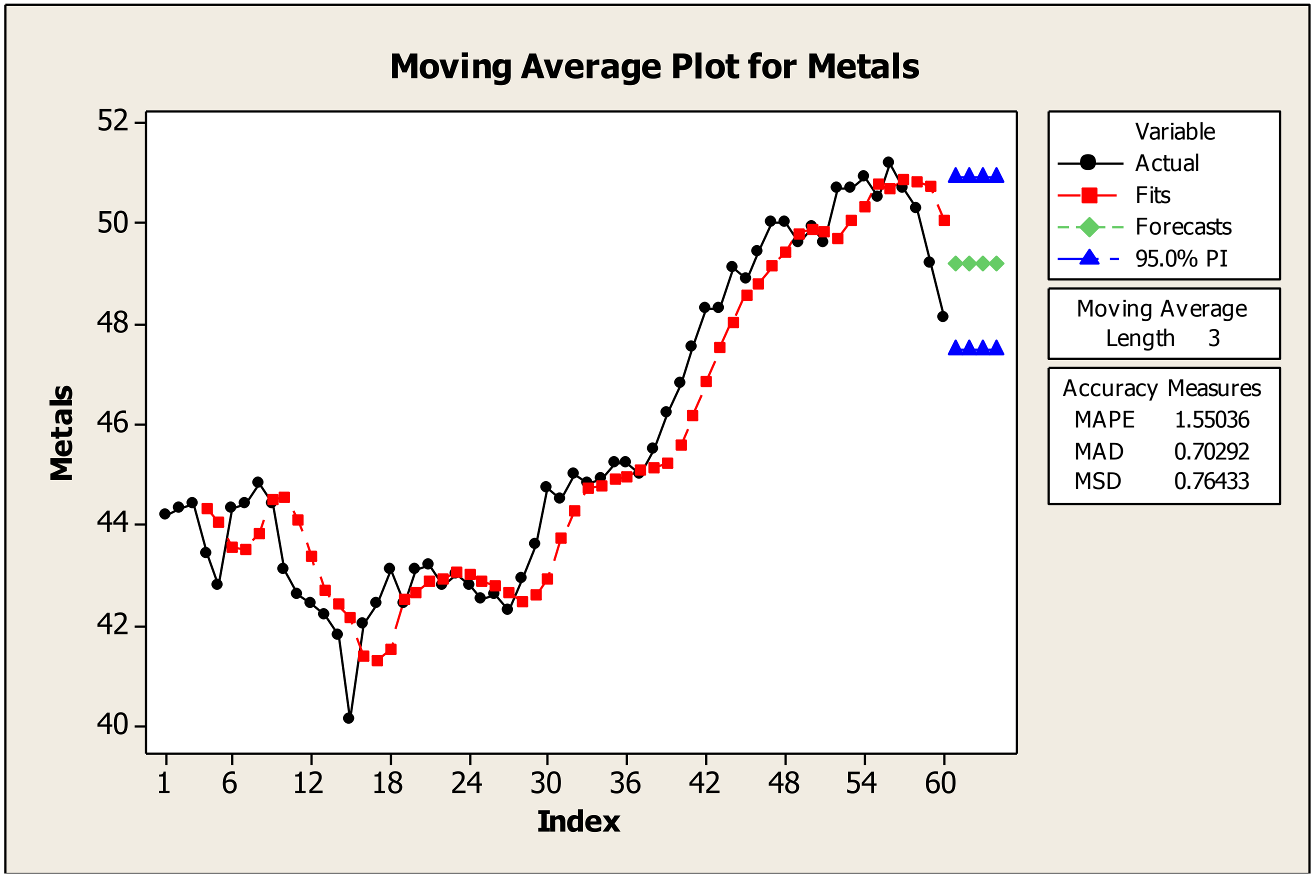
Plotting the data, we get:



And we can apply the moving average with order k = 3 as Follows:


Stat Graph Editor Tools Window Help
Basic Statistics
Regression
ANOVA
DOE
Control Charts
Quality Tools
Reliability/Survival
Multivariate
Time Series
Tables
Nonparametrics
EDA
Power and Sample Size
Time Series Plot...
Trend Analysis...
Decomposition...
Moving Average...
Single Exp Smoothing...
Double Exp Smoothing...
Winters' Method...
Differences...
Lag...
Autocorrelation...
Partial Autocorrelation...
Cross Correlation...
ARIMA
Plot for Metals
duals for Metals

And get the following:



- **single exponential smoothing**

As we have seen, simple moving average assigns the same weight to all observations, that is, it gives both old and recent observations the same importance in smoothing, but real life applications dictate that most recent observations should have more influence on the smoothing than older ones.

AS previously seen, for the time series $y_1, y_2, \dots, y_t$, the simple moving average (SMA) of order k has the form:

$$\hat{y}_t = \frac{1}{k}\left(y_t + y_{t-1} + \cdots + y_{t-k+1}\right) \ ,$$

Or,

$$\hat{y}_t = \frac{1}{k}y_t + \frac{1}{k}y_{t-1} + \cdots + \frac{1}{k}y_{t-k+1}$$

Or,

$$\hat{y}_t = \alpha y_t + \alpha y_{t-1} + \cdots + \alpha y_{t-k+1}$$

This means that SMA gives all observations the same weight α.

This problem can be avoided by giving the old observations weights that decrease exponentially, which is called the simple exponential smoothing (SES),

$$S_t = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} \ldots,$$

$$t = 1, \ldots n, \quad 0 < \alpha < 1$$

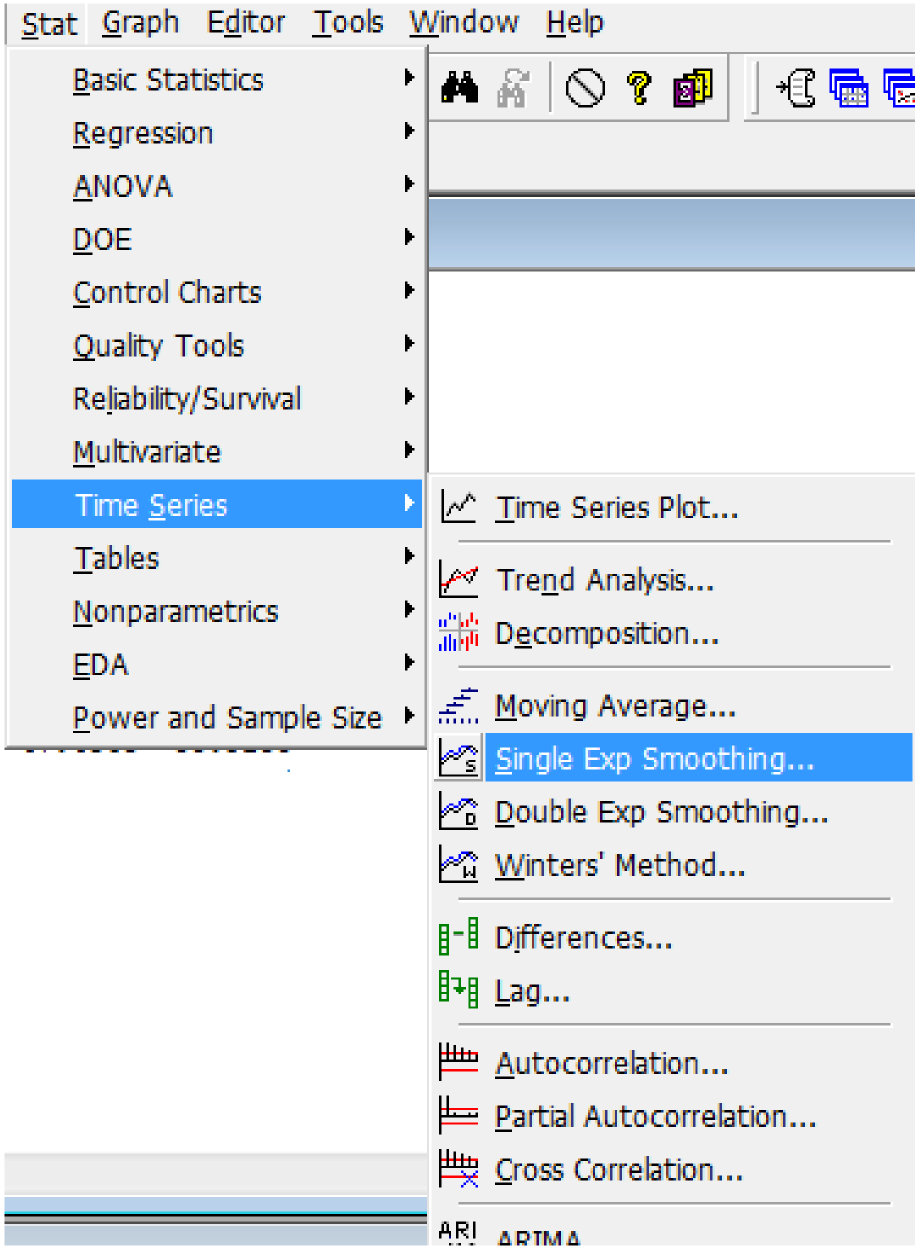
the value S_t is a weighted average that decreases exponentially, it can be written in an recursive manner as follows:

$$S_t = \alpha y_t + (1-\alpha)S_{t-1} \quad , t = 1, \ldots n; \qquad S_0 = \bar{y}\,, \qquad 0 < \alpha < 1$$

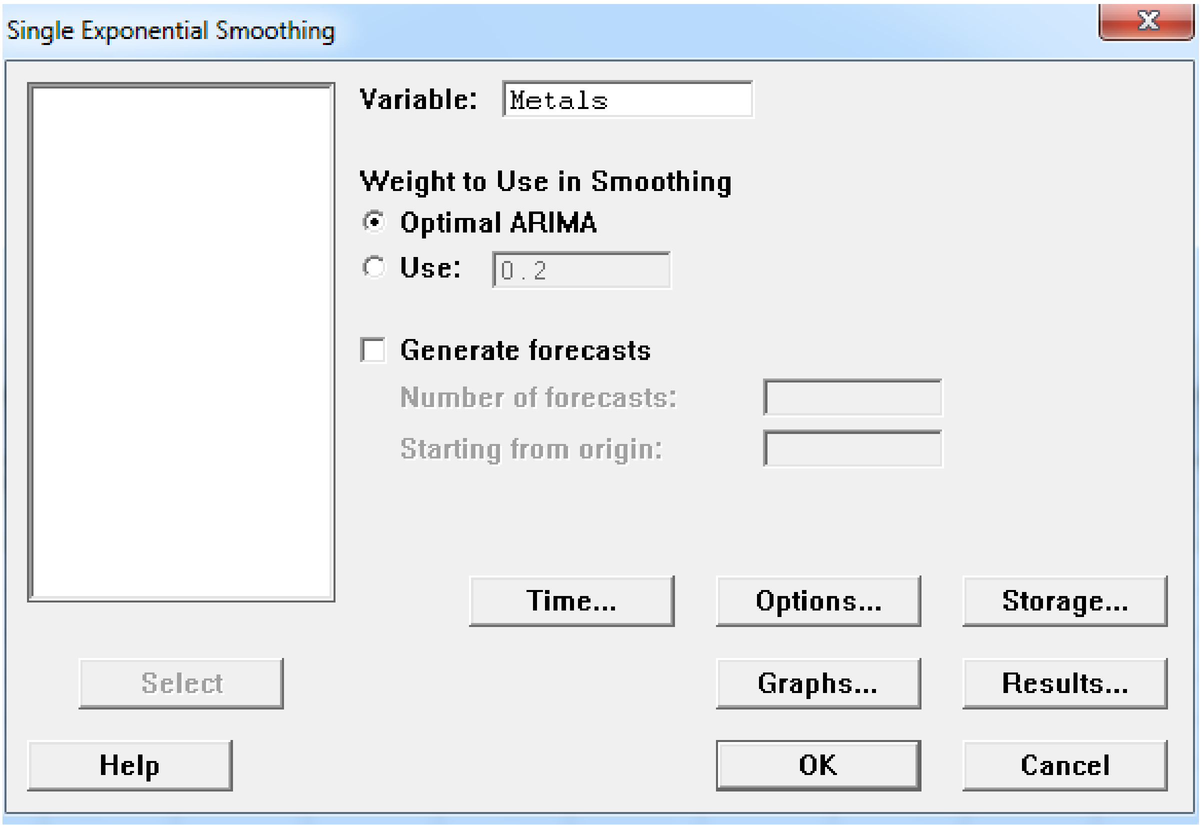
Example: Open data file "EMPLOY.MTB" , use data variable (Metals), smooth the data using single exponential smoothing.

Solution:

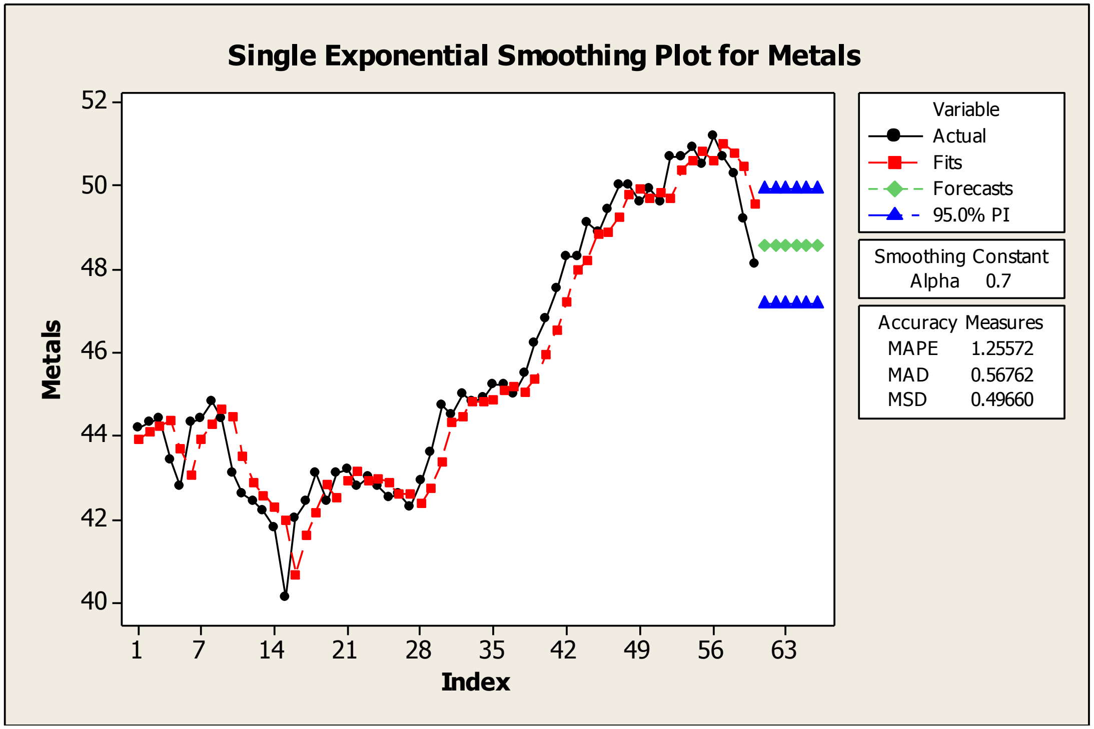
From Minitab, we have:


Stat
Graph
Editor
Tools
Window
Help
Basic Statistics
Regression
ANOVA
DOE
Control Charts
Quality Tools
Reliability/Survival
Multivariate
Time Series
Tables
Nonparametrics
EDA
Power and Sample Size
Time Series Plot...
Trend Analysis...
Decomposition...
Moving Average...
Single Exp Smoothing...
Double Exp Smoothing...
Winters' Method...
Differences...
Lag...
Autocorrelation...
Partial Autocorrelation...
Cross Correlation...
ARIMA

we get the following window:



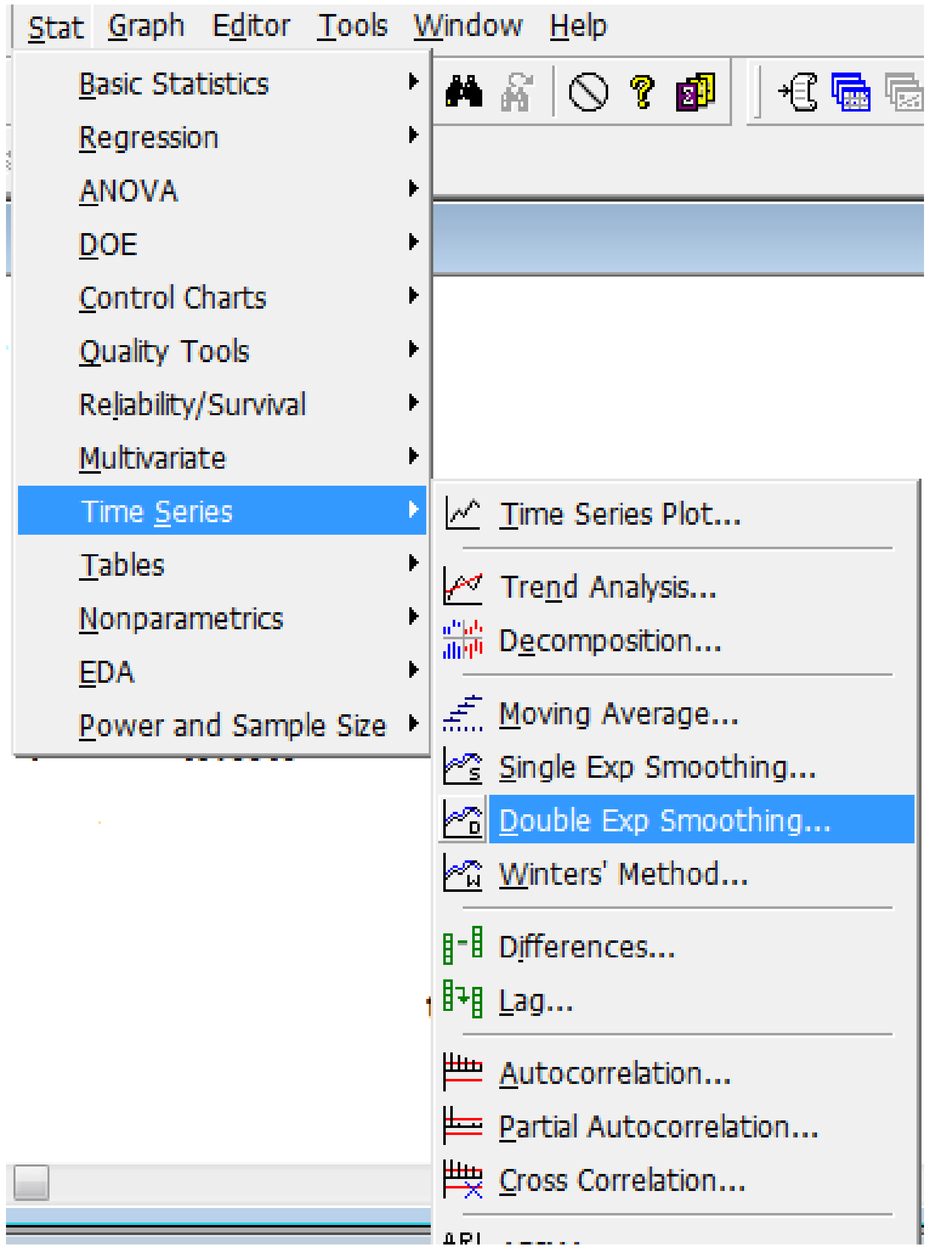
And the result is:



Where we note that the smoothing is better than that obtained from SMA .

Note also the difference between giving a small value for α and larger values. If the value is large then we give recent values larger effect, while older values has little effect in forecasting. For small values for α, the resulting series will be smoother, and vice versa for large values of α. This means that in case the series has lots of fluctuations then we use a small value for α. Usually, we try several values for α and choose the value that gives the best value of the accuracy measures we have seen before.

Note: SES does not provide good forecasts if the series contains trend component (see forecasts in the above figure), and therefore there are other ways of exponential smoothing that provide better forecasts in this case. For example, the so- called double exponential smoothing method, which is a generalization to SES, where in a first stage the original data is smoothed by single exponential smoothing, and in the second stage the smoothed data is smoothed again. Note that in this case we have two smoothing parameters, one for the level of the series, and the other for trend. The following figure shows the result of using this method to data from the previous example:


Stat
Graph
Editor
Tools
Window
Help
Basic Statistics
Regression
ANOVA
DOE
Control Charts
Quality Tools
Reliability/Survival
Multivariate
Time Series
Tables
Nonparametrics
EDA
Power and Sample Size
Time Series Plot...
Trend Analysis...
Decomposition...
Moving Average...
Single Exp Smoothing...
Double Exp Smoothing...
Winters' Method...
Differences...
Lag...
Autocorrelation...
Partial Autocorrelation...
Cross Correlation...

we get the following: