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, etc.. . This means that data have chronological 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.

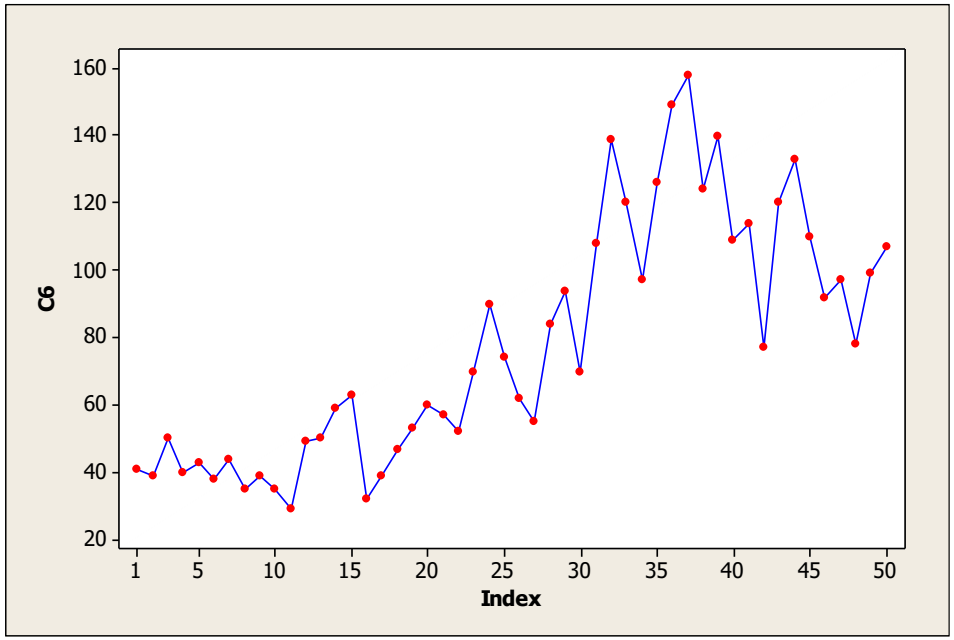


Figure 1.1 The profit gain of a company over a period of 50 years

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

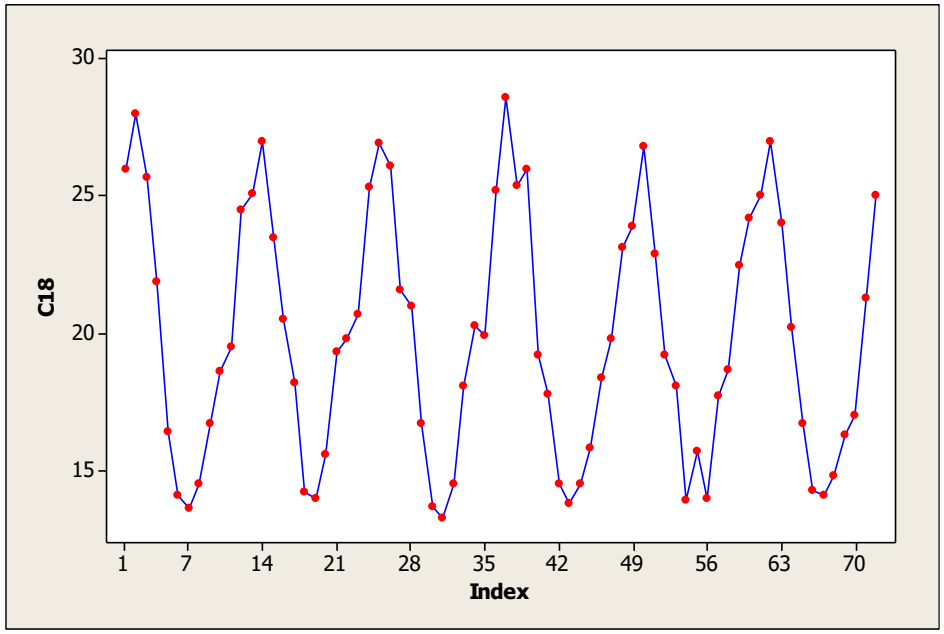


Figure (1.2): average monthly temperatures in a city during a period of 6 years

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

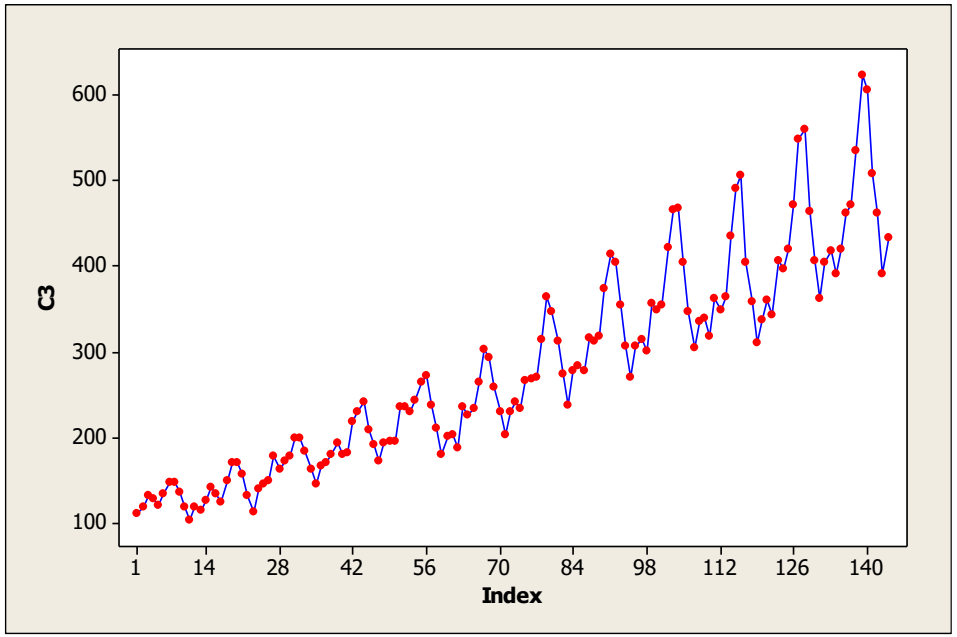


Figure (1.3): monthly sales for some industrial piece during a period of 15 years

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 indicate that the observations are not independent. Therefore, statistical analysis to be used for the analysis must take into consideration the chronological (or spatial) 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

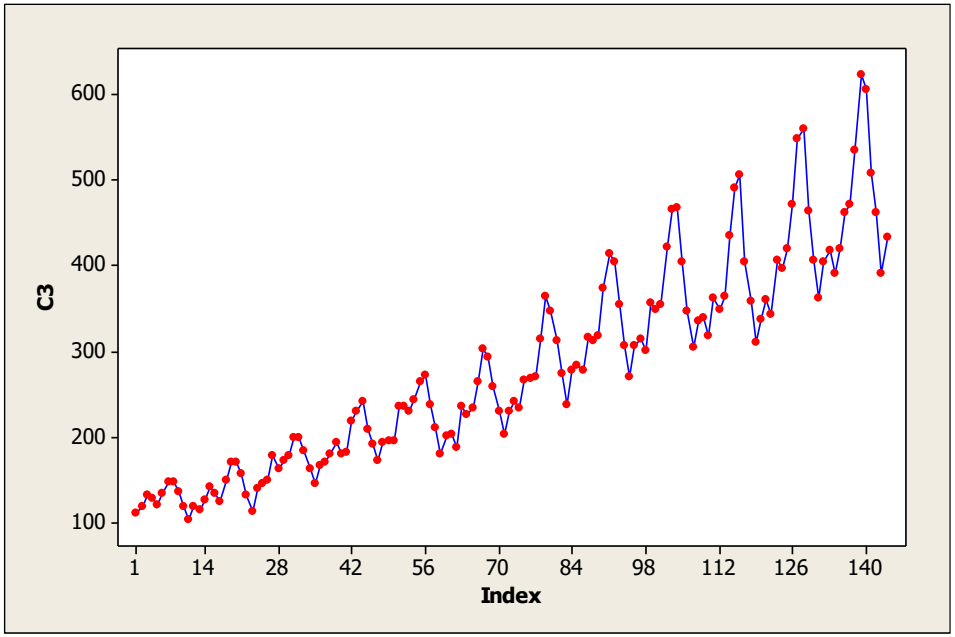
There are several goals for the analysis of time series, some of which are:

1- Description

Time series analysis is used to describe and portray the available information that shows how the studied phenomenon evolve 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 at 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 exceed 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 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:


Random errors
0.10
0.05
0.00
-0.05
C5
t
0
10
20
30
40
50
60
C6

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.

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:

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

b- Requires the availability of detailed historical information about all the explanatory variables, and the ability of knowing

these variables or predicting them.

c- 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 $f(t.\beta)$, where the vector β 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:

$$y_t = f(t.\beta) + \varepsilon_t . \qquad t = 1.2. \ldots . n$$

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

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

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

1) The function $f(t.\beta)$ 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 $y_1.y_2.\ldots.y_n$ 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 represent a certain origin point, and that we want to predict the value of the series after k time intervals, then this approach indicate 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 :

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

this means that:

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

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 it this that 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,

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

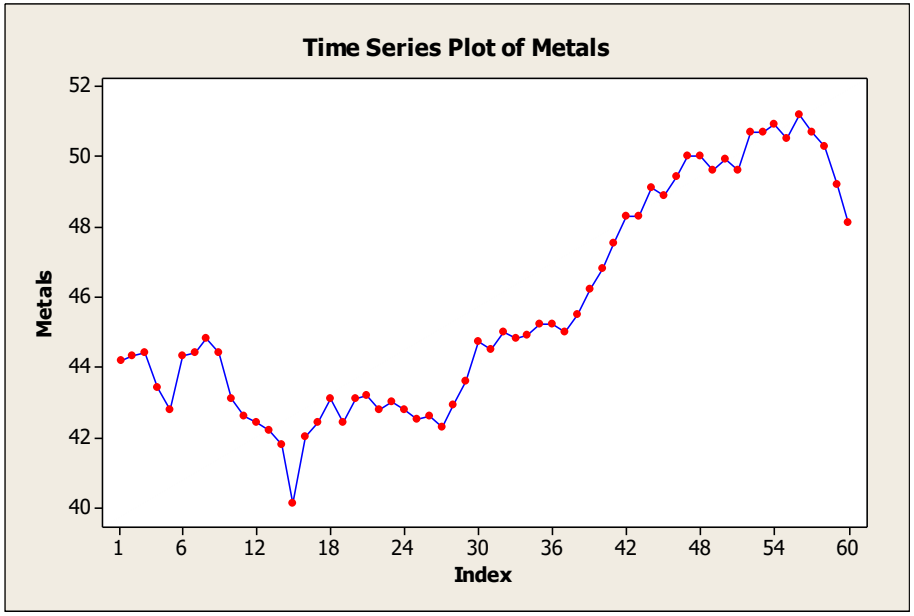
$$ma_6(3) = 626.33$$

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

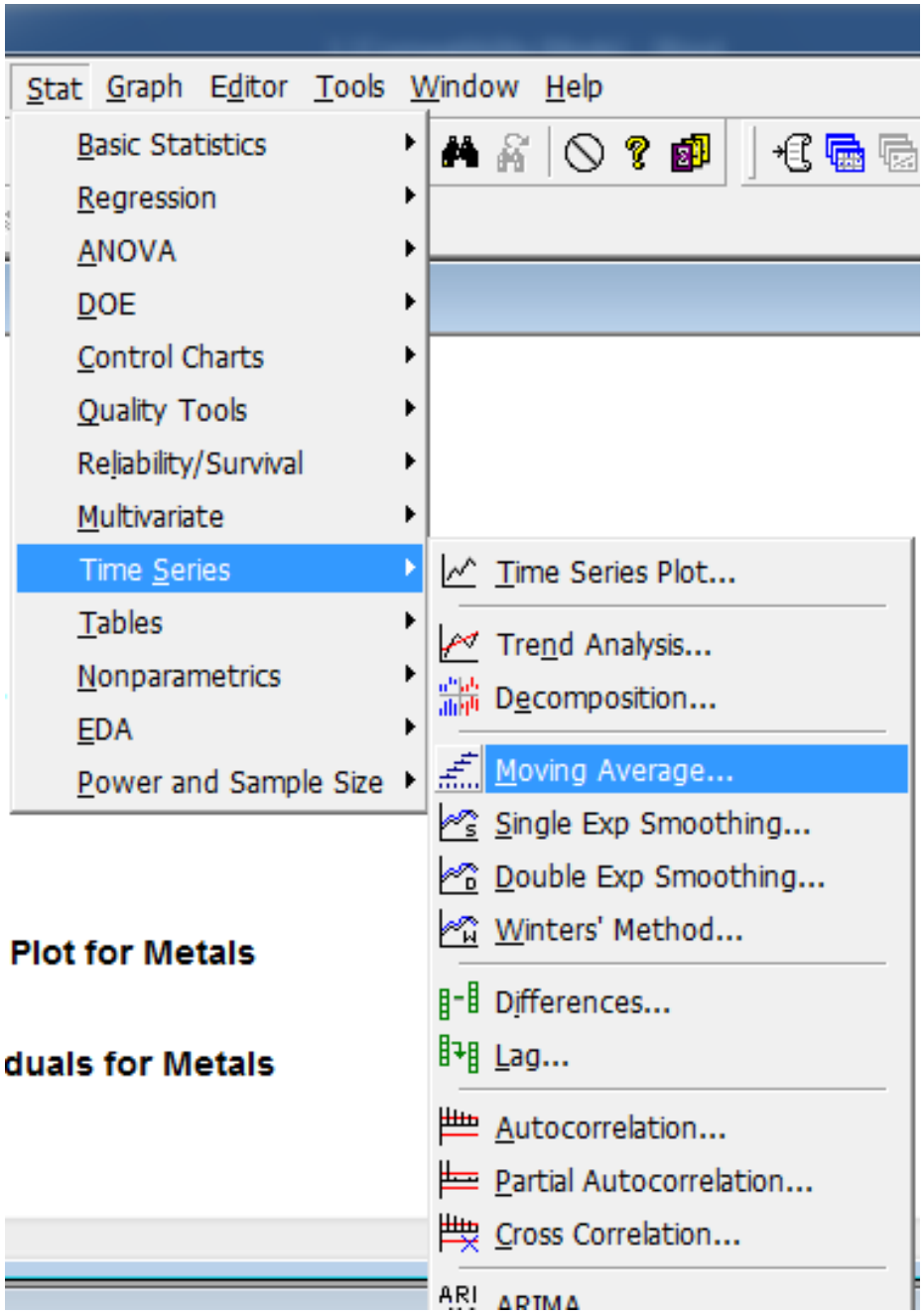
```
44.2 44.3 44.4 43.4 42.8 44.3 44.4
44.8 44.4 43.1 42.6 42.4 42.2 41.8
40.1 42.0 42.4 43.1 42.4 43.1 43.2
42.8 43.0 42.8 42.5 42.6 42.3 42.9
```

```
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
```

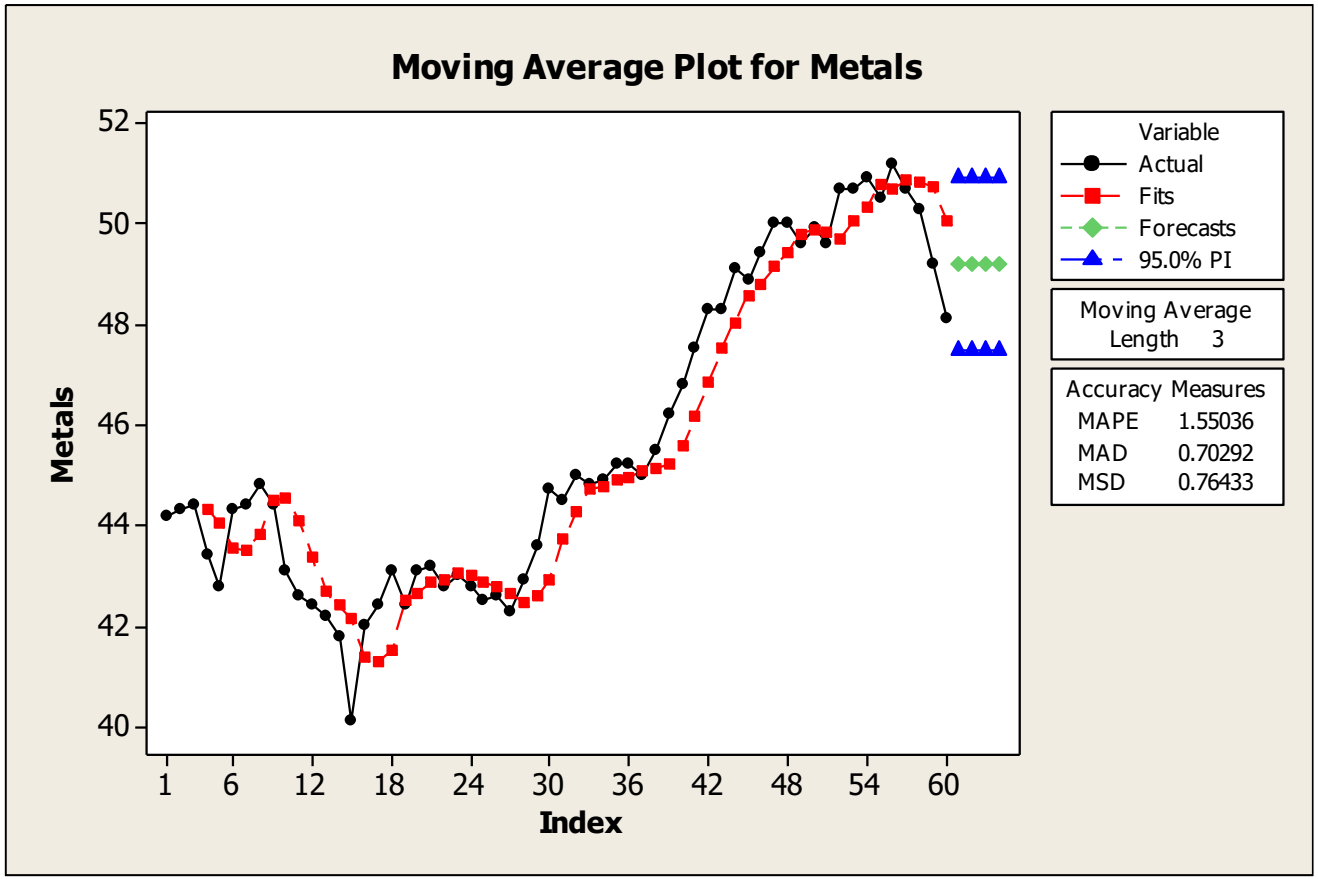
Plotting the data, we get:



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 $y_1 . y_2 . \ldots . y_t$, the simple moving average (SMA) of order *k* has the form:

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

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),

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

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

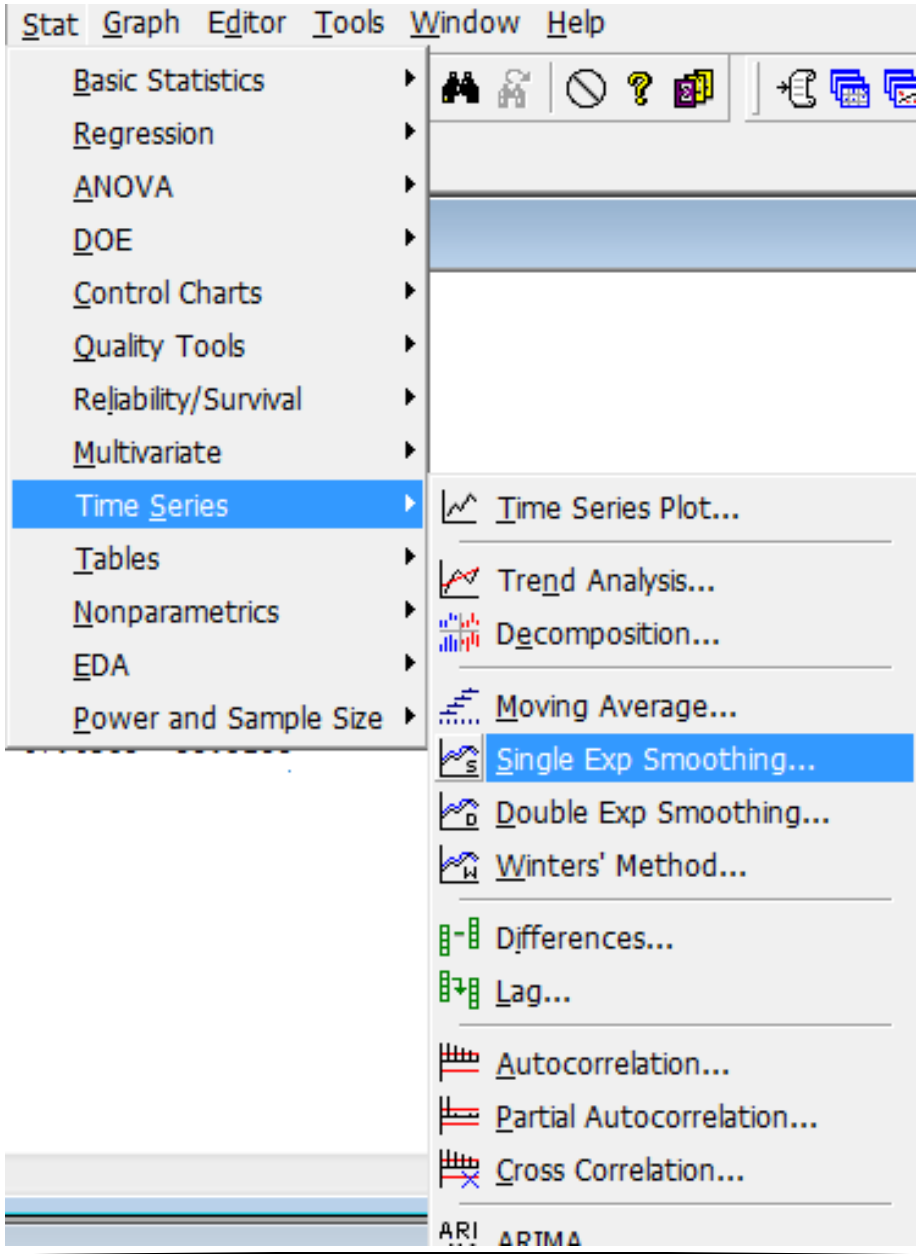
the value S_t is a weighted average that decreases exponentially, it can be written in an recursive manner as follows:

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

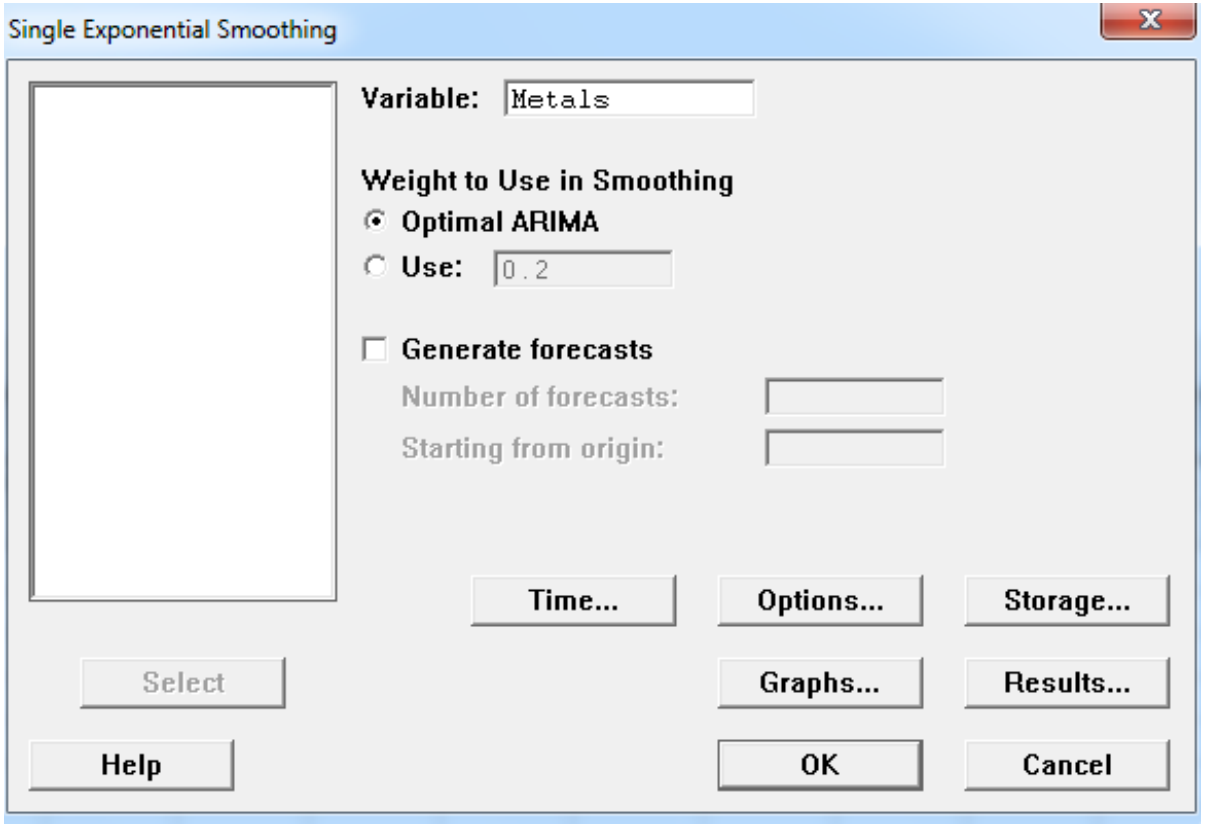
Example: Open data file "EMPLOY.MTB" , use data variable (Metals), smooth the data using single exponential smoothing.

Solution:

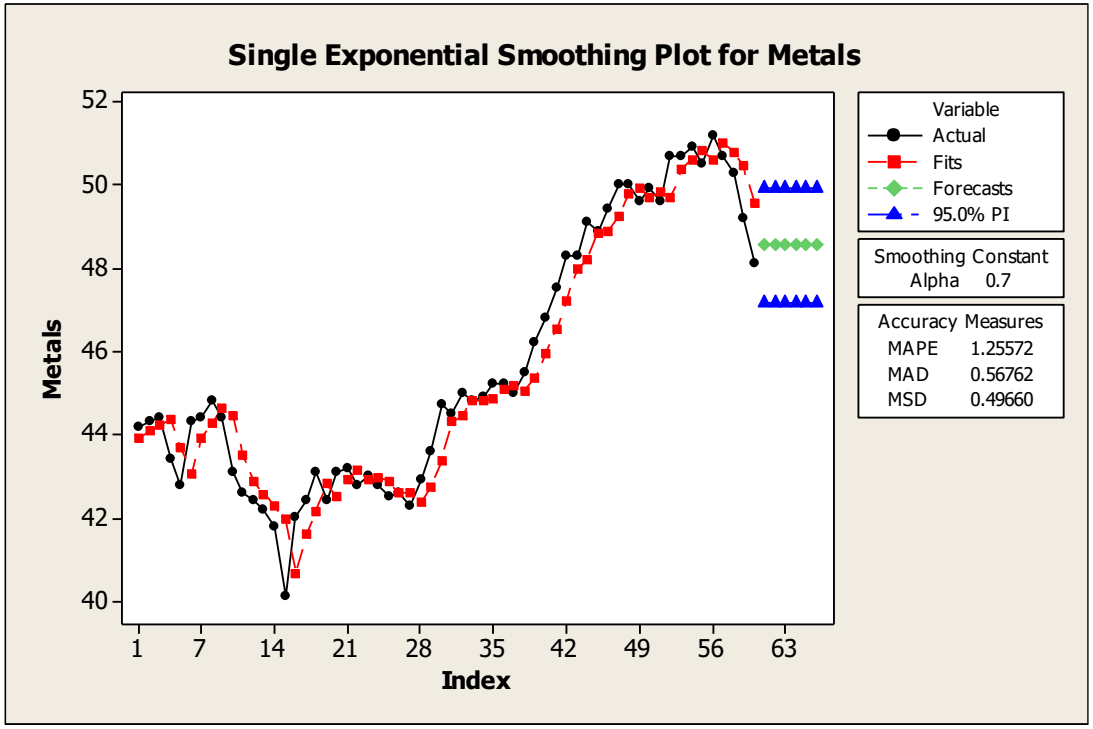
From Minitab, we have:


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:

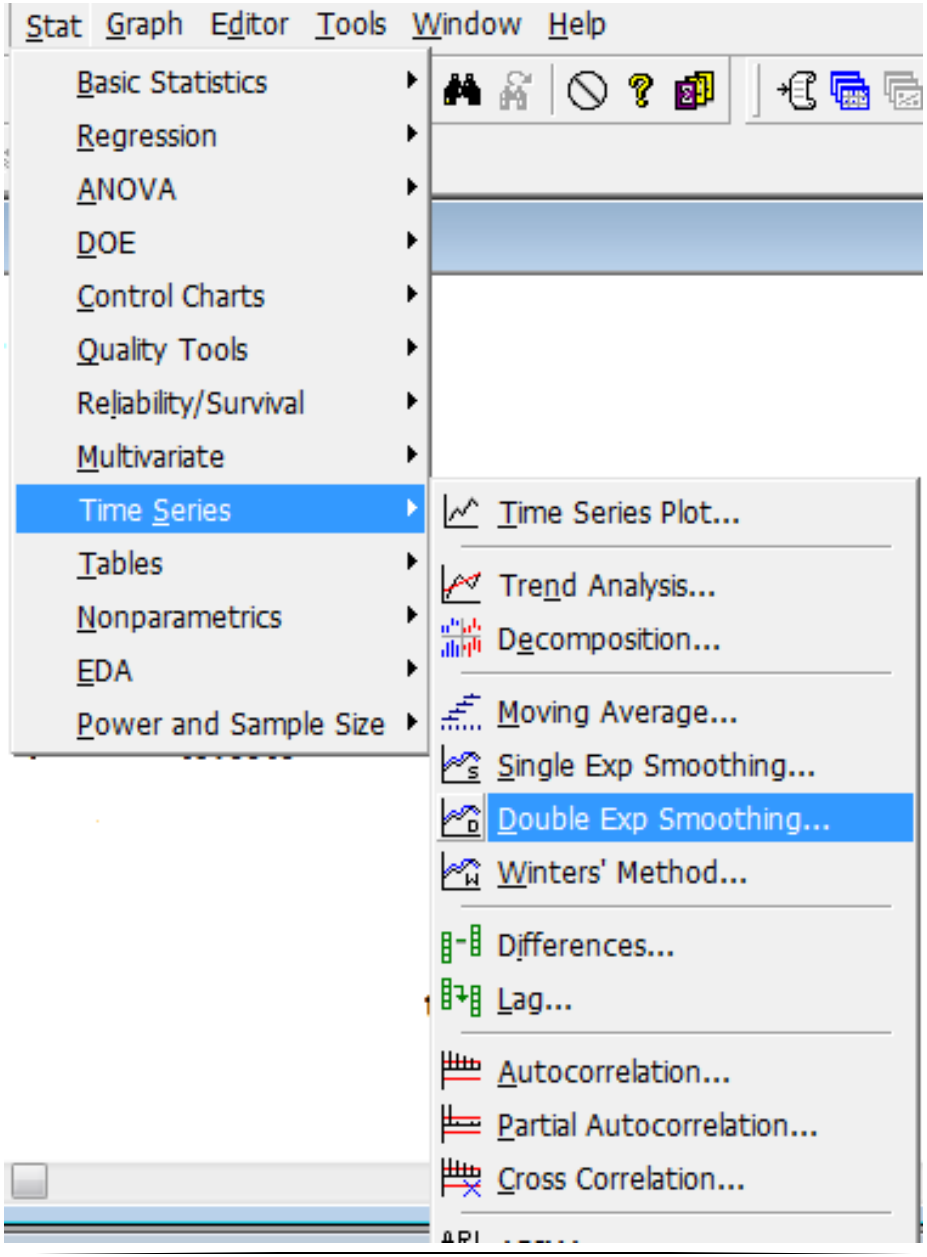


Where we note that the smoothing is better than that obtained from SMA .

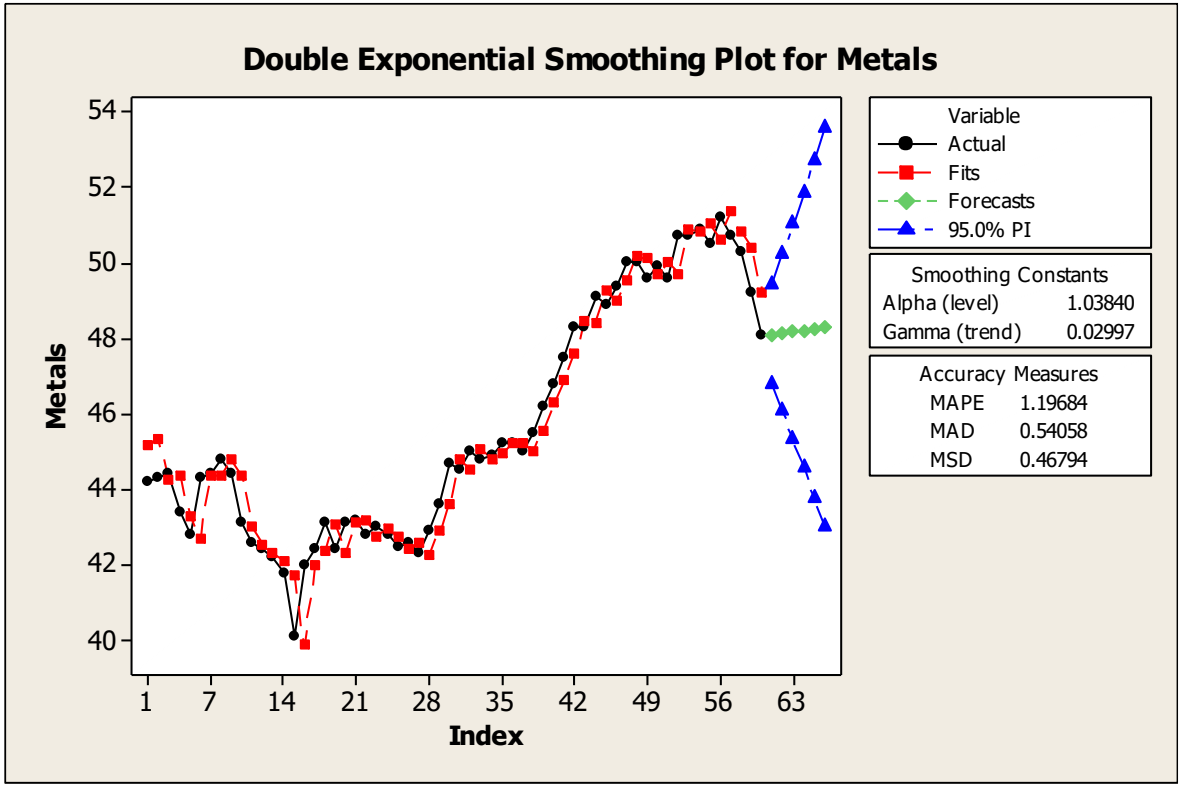
Note also the difference between giving a small value for α and larger values. If the value is large then we give recent values larger effect, while older values has little effect in forecasting. For small values for α, the resulting series will be smoother, and vice versa for large values of α. This 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 step the original data is smoothed by single exponential smoothing, and in the second step 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:



1.6.2.3 Stochastic time series models

The techniques discussed in the previous lecture are simple and traditional, and none of them can be considered to be statistically structured methodology for the analysis of time series. The Stochastic time series analysis provide more sophisticated methods of forecasting. The random model always assumes the existence of a theoretical stochastic process able to generate the time series at our hands. If it is assumed theoretically that such a process is used to

produce large group of series on the same time interval under study, then every series will be different from the others, however, all group of series will follow same probability rules. This is exactly the same case as the population and the sample, where we can select many different samples from the same population, however these samples will follow same probability rules as the population.

Therefore, the proposed method suggested here, assumes that the observations of the time series $(y_1 . y_2 . \ldots . y_n)$ that are observed in the

time interval $(1.2.\ldots.n)$ is a realization drawn from multivariate random vector $(Y_1.Y_2.\ldots.Y_n)$ that have cumulative distribution function $F(y_1.y_2.\ldots.y_n)$ which is used to make inferences about the future of the stochastic process. It is well known in statistical science, that knowing or determining such a cumulative distribution function is a very difficult task, but it is the norm to create a model to describe the behavior of the series efficiently, this efficiency depend on how such model can reflect properties of the true probability distribution.

We will present in this course a modern statistical methodology for the analysis of time series called Box-Jenkins methodology denoted shortly as ARIMA models.

1.7 Types of change in time series

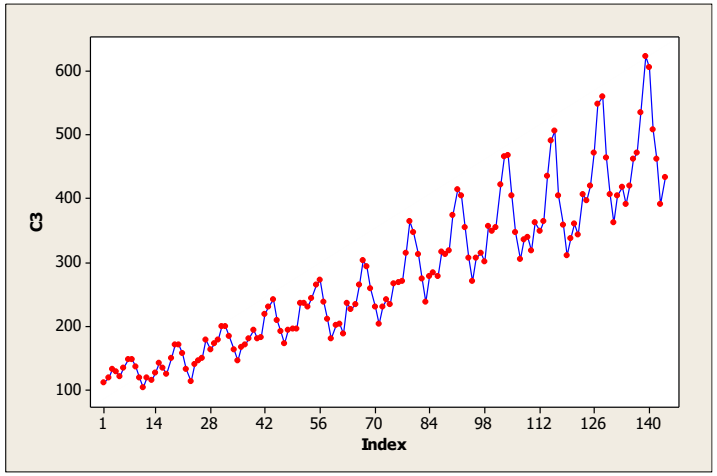
Traditional methods of time series analysis rely on dismantling the change in a time series into four different components:

- trend component

- seasonal component
- cyclical component
- random component

1.7.1 trend component

If there exist a long term increase (or decrease) in the level of the series, then we say there exist a trend component in the series, see figure 1.3 for an example.



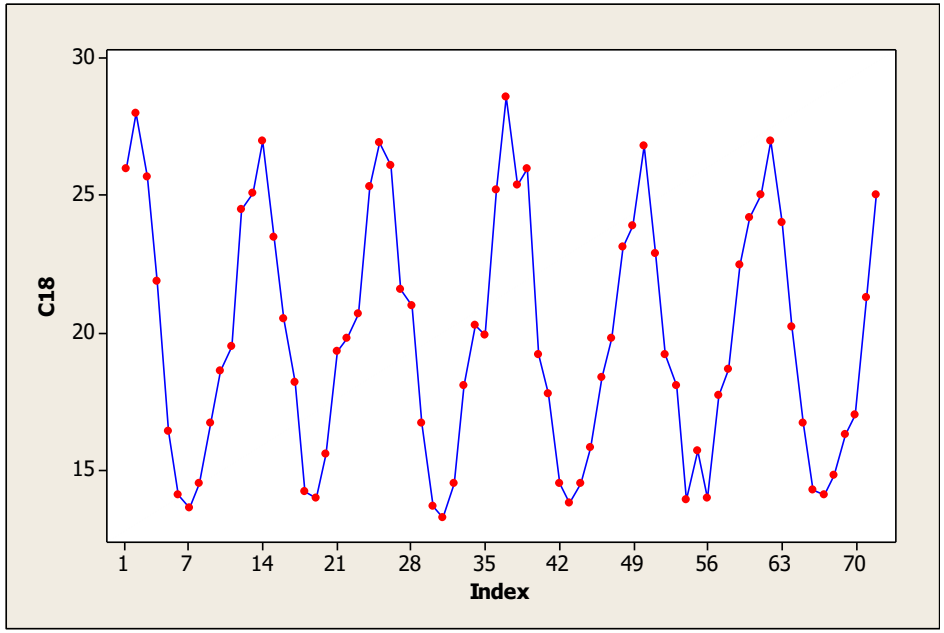
So when examining the time series plot, often we notice the presence of a slow and gradual changes in the short term (increase or decrease), and a general tendency to increase in the long term, as it happens, for example, in time series of the number of births, or

the number of pilgrims, or prices of goods annually. On the other hand, we may find a general tendency to decrease in the long term, as for example, in the series of the number of deaths, or oil stocks, or for a particular disease.

1.7.2 seasonal component

Many time series in practice can be affected by what is called seasonal pattern changes, by which we mean the series repeats its

behavior at certain periods of the year, for example, the electric power consumption reaches its peak in summer and fall in winter, see figure (1.2) for the time series of daily temperature as an example.



Seasonal changes occur at specific periods less than a year, such as hour, day, week, month, quarter, etc.

1.7.3 cyclical variation

These changes are similar to seasonal variation, but they appear in long periods of time (more than one year), and to discover the cyclical variation one need a very long annual series, for example, climate changes needs data of thirty years or more to discover its

cycle. Also, economic cycles need a long periods of time, for example five or ten Years, to appear.

1.7.4 Random variation

After getting rid of seasonal, trend, or cyclical components from the data, we are left with a residual series, which represent the irregular changes. These changes differ from the other components, as they

can't be predicted, and they do not occur according to any law or system.

Chapter 2: Basic Concepts

As we mentioned earlier, the modern time series analysis presented by **Box and Jenkins** in the year (1971), is based on examining the random nature of the time series. This methodology assumes that there is always a theoretical random process (Stochastic process)

capable of generating infinite number of time series of a certain length n, and that the observed series we are studying (called sometimes a sample) is just one of them. We study this sample for the purpose of understanding and describing the nature of the random stochastic process that generated it.

Box-Jenkins methodology is popularly used in the scientific community of theoretical and applied sciences. It has proven to be highly efficient in modeling and forecasting time series that arise in

various fields of knowledge such as economics, business administration, environment, chemistry and engineering, among others. The method of Box-Jenkins has several advantages including:

1- It is a comprehensive approach, in the sense that it offers good solutions for all stages of analysis in the form of a more scientific and rational scheme than other methods through building models,

diagnosis and estimating the parameters and forecasting future values.

2 - Richness of the stochastic models that this methodology is capable of dealing with, enables Box-Jenkins methodology to reflect the probabilistic mechanism for a lot of stochastic processes that appear in various areas of application. These models are known as ***Autoregressive Moving Average*** models or **ARMA** models in short.

3 - It does not assume independence between the observations of the time Series but, in fact, it takes advantage of the dependence structure between the observations in the modeling and forecasting process, which usually lead to a more accurate and credible forecasts than the ones we get through the conventional methods.

4- It gives more credible confidence intervals for future values when compared to other conventional methods such as exponential smoothing.

However, the method of Box-Jenkins has some disadvantage, the most important one is that it requires availability of a large number of observations (at least 50 observations), to be able to get a good model.

2.1 Stationarity

Modern time series analysis assumes that any observation y_{t_1} at certain point of time t_1 is just a single observation randomly chosen

from a random variable Y_{t_1} (which represents all observations that can be observed at time t_1) and has a cumulative distribution function $F(Y_{t_1})$.

Similarly, it assumes that any two observations $(y_{t_1}.y_{t_2})$ at any two different time points $(t_1.t_2)$ represents a single point drawn from bivariate random variable $(Y_{t_1}.Y_{t_2})$ (which represents all

observations that can be observed at the two time points $(t_1 . t_2)$ and has a cumulative distribution function $F(Y_{t_1} . Y_{t_2})$.

In general modern time series analysis assumes the existence of a (theoretical) stochastic process capable of generating an infinite number of time series, and that the observed time series at hand is just one of them, and that there is a probabilistic distribution for the random variables $(\mathrm{Y}_1 . \mathrm{Y}_2 . \ldots . \mathrm{Y}_\mathrm{n})$.

2.1.1 Strict Stationarity

We say that a time series is strictly stationary if the joint cumulative probability distribution of any subset of the variables that make up the series is not affected by displacing the time forward or backward any number of time units. So, if $(t_1. t_2. \ldots . t_m)$ is any subset of time units, where $m = 1.2.3. \ldots$ and $k = \pm 1. \pm 2. \ldots$, then we say the series is strictly stationary if the joint cumulative probability distribution for the variables $(Y_{t_1}. Y_{t_2}. \ldots . Y_{t_m})$ is the same as the joint cumulative

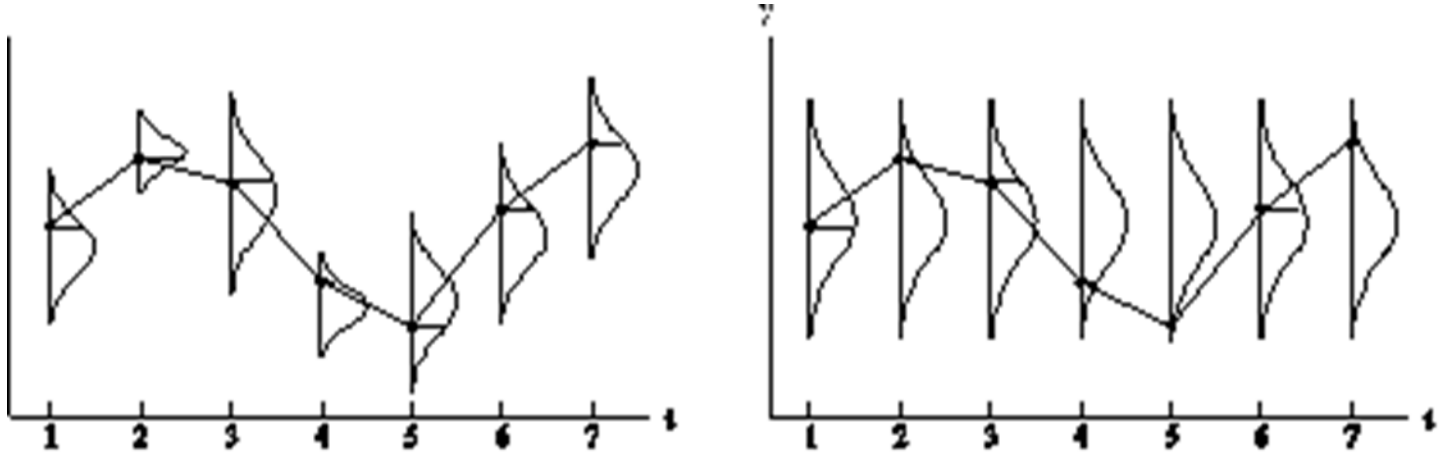
probability distribution for the variables $(Y_{t_1+k}. Y_{t_2+k}. \ldots . Y_{t_m+k})$ for any time point t and any time shift k . Mathematically we can write the condition of strict stationarity as:

$$F(Y_{t_1}. Y_{t_2}. \ldots . Y_{t_m}) = F(Y_{t_1+k}. Y_{t_2+k}. \ldots . Y_{t_m+k})$$

$$\Rightarrow \quad P(Y_{t_1} \leq c_1. Y_{t_2} \leq c_2. \ldots . Y_{t_m} \leq c_m)$$

$$= P(Y_{t_1+k} \leq c_1. Y_{t_2+k} \leq c_2. \ldots . Y_{t_m+k} \leq c_m)$$

Strict stationarity simply means that the mechanism of generating the observations for the stochastic process under consideration is constant through time, so that the shape of the model and the parameter estimates do not change with time shift.



Stochastic processes and realized time series.

From this definition we can see that strict stationarity necessarily leads to the fact that the mean and the variance of the stochastic process are constant (of course provided they exist). Also the covariance between

any two variables Y_t and Y_s depend only on time lag (i.e. the time distance between them).

So strict stationarity leads to the following:

i) $\mu_t = E(Y_t) = \mu\,. \quad t = 0.\pm1.\pm2.\ldots$

ii) $\sigma_t^2 = Var(Y_t) = \sigma^2\,. \quad t = 0.\pm1.\pm2.\ldots$

iii) $\gamma(s.t) = Cov(Y_s.Y_t) = E[(Y_s - \mu)(Y_t - \mu)] = \gamma(s-t)$

that is the covariance between $(y_s . y_t)$ will be a function in the time lag $(s - t)$ only, so:

$$\gamma(t . t - k) = Cov(Y_t . Y_{t-k}) = \gamma(k)$$

As we know, the variance could be considered as a special case of the covariance function $\gamma(s . t)$ if $s = t$, i.e.

$$Var(Y_t) = \gamma(t . t)$$

and if the series is stationary then,

$$Var(Y_t) = \gamma(t.t) = \gamma(0). \quad t = 0.\pm1.\pm2.\ldots$$

2.1.2 Weak Stationarity

We say that a series is weakly stationary if the moments up to second order exist, and:

1- The expected value or the mean of the process μ_t does not depend on time t , i.e. :

$$\mu_t = E(Y_t) = \mu \,.\, t = 0. \pm 1. \pm 2. \ldots$$

2- The variance σ_t^2 does not depend on time t , i.e.

$$\sigma_t^2 = Var(Y_t) = \sigma^2 \,.\, t = 0. \pm 1. \pm 2. \ldots$$

3- Covariance between any two variables depend only on the time lag between them, i.e.,

$$Cov(Y_{t-k}. Y_t) = \gamma(k). \quad t = 0. \pm 1. \pm 2. \ldots ; k = \pm 1. \pm 2. \ldots$$

From the above we can see that strict stationarity always leads to weak stationarity, the vice versa is only correct in the case that the joint cumulative distribution of the variables $(Y_{t_1}.Y_{t_2}.\ldots.Y_{t_m})$ is the multivariate normal distribution since this distribution is completely defined by its first two moments, in this case only if the stochastic process is weakly stationary then it is strictly stationary.

From now on, if we mention stationarity from now on, then we mean weak stationarity.

2.1.3 The importance of stationarity

If the statistical characteristics of the stochastic process that generated the time series is not stationarity, we will face many difficulties. The most important is the large number of parameters, such as expectations, variances and covariance's and the difficulty of interpreting these parameters.

- Reducing the number of parameters:

If we assume that the process y_t is stationary and that one observation is available at every time point, which is the case in most real life time series, so that we have the following observed series $(y_1 . y_2 . \ldots . y_n)$, then the major parameters of the theoretical process are :

$$E(\boldsymbol{Y}) = [E(Y_1)\, E(Y_2) \ldots E(Y_n)]^t = [\mu_1 \;\; \mu_2 \ldots \mu_n]^t$$

$$Var(Y) = \gamma(s.t) = \begin{bmatrix} \gamma(1.1) & \gamma(1.2) \dots & \gamma(1.n) \\ \gamma(2.1) & \gamma(2.2) \dots & \gamma(2.n) \\ \vdots & \vdots & \vdots \\ \gamma(n.1) & \gamma(n.2) \dots & \gamma(n.n) \end{bmatrix}$$

Where we interpret the mean of the stochastic process at time t , i.e. μ_t as the mean for all values that this process can generate at time t, also, we interpret the variance of the stochastic process at time t , i.e. $\gamma(t.t)$ as the variance for all these values. Whereas, the covariance

$\gamma(s.t)$ measures the linear dependence between all values that this process can generate at time s and time t.

Now notice that number of expectations is n, and the number of parameters of the variance and covariance matrix is $n(n+1)/2$. Thus, the total number of main parameters to be estimated if the process is not stationary are $n(n+1)/2 + n = n(n+3)/2$ which is a large number especially if the number of

observations n is large. However, in the case of stationarity, number of parameters will be $(n + 2)$ which are:

$$\mu.\gamma(0).\gamma(1).\ldots.\gamma(n)$$

Where in case of stationarity, μ represent level of the series. Also the variance $\gamma(0)$ measures variability of the process around μ. In the same manner we can interpret the auto-covariance at time lag k (i.e. $\gamma(k)$), so $\gamma(1)$ represent the auto-covariance between variables one

period of time apart, $\gamma(2)$ represent the auto-covariance between variables two period of times apart, etc.

Preliminary Stationarity tests

There are several ways to test the stationarity of the series, some of these methods are accurate others are approximate. If the series follows a known theoretical model then we can test its stationarity by calculating its expectation, variance and covariance functions. If both the expectation and variance does not depend on time, and the auto-

covariance function depend only on time lag between any two variables, then stationarity of the series can be decided.

Example: If the series follow the following model:

$$y_t = \beta_0 + \varepsilon_t. \quad t = 1.2. \ldots . n$$

Where β_0 is a fixed constant, and the variables $\varepsilon_1. \varepsilon_2. \ldots$ are uncorrelated random variables with mean zero and contrast variance σ^2. Is the series stationary?

solution:

Calculate the expectation, variance and covariance of the process:

$E(Y_t) = \beta_0 \ . \ \ t = 0. \pm 1. \pm 2. \ldots$

$V(Y_t) = \ V(\beta_0 + \varepsilon_t) = V(\varepsilon_t) = \sigma^2$

$$Cov(Y_t . Y_{t-k}) = Cov(\beta_0 + \varepsilon_t . \beta_0 + \varepsilon_{t-k}) = 0 \ . \ \ k = \pm 1. \pm 2. \ldots$$

Therefore, we note that all the weak stationarity conditions are fulfilled here.

Example: If the series follow the following model:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t. \quad t = 1.2. \ldots . n$$

Where $\beta_0 . \beta_1$ are fixed constants, and the variables $\varepsilon_1 . \varepsilon_2 . \ldots$ are uncorrelated random variables with mean zero and contrast variance σ^2. Is the series stationary?

solution:

We calculate the expectation of the process:

$$E(y_t) = \beta_0 + \beta_1 t \ . \ t = 1.2. \ldots$$

This means that the expected value of the series is not constant but increasing (decreasing) by a constant value if $\beta_1 > 0$, $(\beta_1 < 0)$ i.e. the series has a trend component in case $\beta_1 \neq 0$, and hence it is not stationary.

Example: If the series $\{y_t\}$ follow the following model:

$$y_t = y_{t-1} + \varepsilon_t. \quad t = 1.2. \ldots . n$$

where $\{\varepsilon_t\}$ is a random process as defined in the previous example. Is the process stationary?

solution:

$$E(y_t) = E(y_{t-1}) + E(\varepsilon_t) = E(y_{t-1}) + 0 = E(y_{t-1})\,.\quad t = 1.2.\ldots n$$

Which means that the mean of the series is constant, and does not depend on time t. Now we look at the variance,

$$\mathrm{Var}(y_t) = \mathrm{Var}(y_{t-1}) + \sigma^2 + 2Cov(y_{t-1}.\varepsilon_t)$$

$$= \mathrm{Var}(y_{t-1}) + \sigma^2$$

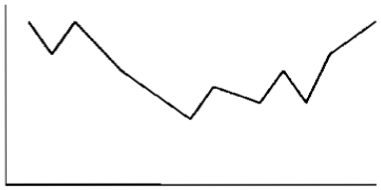
So that $\mathrm{Var}(y_t) \neq \mathrm{Var}(y_{t-1})$, i.e. the variance is not constant , and hence the process is not stationary.

Previous examples have shown how to check stationarity of a time series if the mathematical model that explains the behavior of the random process generated it is known. But in practical applications often this is not the case, and we will mention later some methods for investigating stationarity of the series. But as a general guideline is to check the plot of time series, and if we notice the observations to

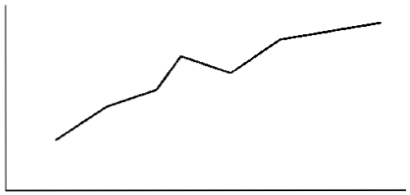
oscillate around a constant line that pass through the middle of the series, then we might be able to believe that the series is stationary. However, if we notice existence of a trend component and/or that the dispersion of the data change over time then we find this an indication of non-stationarity of the series, see figure bellow:

not Stationary in variance

Stationary series



not Stationary in mean



not Stationary in mean

If the series is not stationary, then sometimes some mathematical transformations might be able to transform it to stationarity, we will see this in section 2.5.

2.2 Auto-Correlation function (ACF)

For any stationary process $\{Y_t\}$, the auto-covariance function between Y_t and Y_{t-k} is defined as:

$$\gamma_k = Cov(Y_t . Y_{t+k}) = E[(Y_t - \mu)(Y_{t+k} - \mu)]$$

This function measure the degree of linear association between any two variables of the same time series, for example, $\gamma(1.2)$ measures linear association between all values that could be generated by the stochastic process at time point 1, and those at time point 2.

Notes:

1 - If $\gamma(s.t) = 0$, this means that the two variables Y_t and Y_s are linearly uncorrelated, however, they might still be nonlinearly correlated.

2 – If $\gamma(s.t) = 0$, and the two variables Y_t , Y_s have bivariate normal distribution then this lead to the fact that they are independent.

3 - Sample variance can be regarded as a special case of auto-covariance function $\gamma(s.t)$, by letting $s = t$, this means that $var(Y_t) = \gamma(t.t)$.

4 - If the series is stationary, then auto-covariance function $\gamma(s.t)$ is a function of the time lag $k = |s - t|$ only, and usually we denote it as $\gamma(|s - t|)$, or $\gamma(k)$.

2.2.1 what is Autocorrelation

It is known that the use of covariance function to measure the degree of linear dependence between two variables raises some practical problems.

The first: being the lack of reference boundaries (low, high) that can be referenced to determine the strength or weakness of the linear relationship. Secondly: the covariance depends on the measurement units of the data, so it is always preferable to calibrate the covariance

by dividing by the product of standard deviation of the variables $\boldsymbol{Y_t}$ and $\boldsymbol{Y_s}$ to get what is known as auto-correlation function.

Definition:

The correlation coefficient $\rho(s.t)$ is defined as the correlation coefficient between the variables Y_t and Y_s and is given by the form:

$$\rho(s.t) = \frac{\gamma(s.t)}{\sqrt{Var(Y_s)\,Var(Y_t)}}$$

$$= \frac{E[(Y_s - \mu_s)(Y_t - \mu_t)]}{\sqrt{E(Y_s - \mu_s)^2\,E(Y_t - \mu_t)^2}} \;;$$

$$Where \;\; s.t = 0.\pm 1.\pm 2. \ldots$$

Since it measures the linear correlation between the same random variable data but at different time points, so usually the term "autocorrelation function" is used, and in short written as ACF.

2.2.2 Characteristics of the autocorrelation function

1 - Autocorrelation between the variable Y_t and itself equal one, that is $\rho(t.t) = 1$.

2 - $\rho(t.s) = \rho(s.t)$ because $\gamma(t.s) = \gamma(s.t)$.

3 - Value of $\rho(t.s)$ always lies in the interval $[-1\ .\ 1]$.

4- If $\gamma(s.t) = 0$, then this indicate that the variables Y_t and Y_s are linearly uncorrelated, however, they might still be nonlinearly correlated.

If the stochastic process that generated the time series is stationary, then we redefine the auto-correlation coefficient as:

$$\rho(k) = \frac{E[(Y_t - \mu)(Y_{t-k} - \mu)]}{\sqrt{E(Y_t - \mu)^2}}$$

$$= \frac{\gamma(k)}{\gamma(0)}; \quad k = 0. \pm 1. \pm 2. ..$$

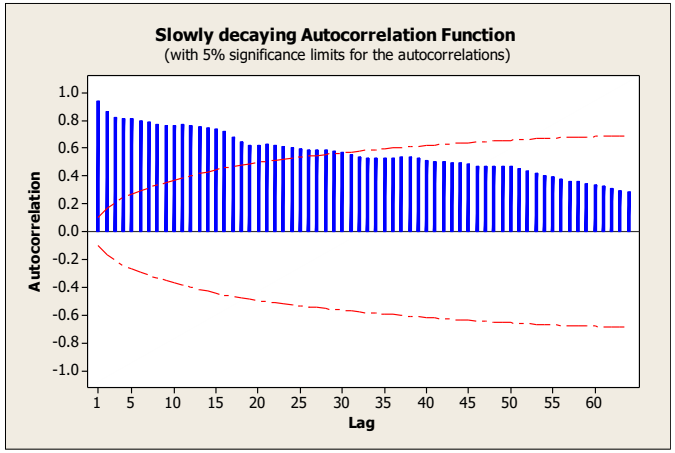
Where $\gamma(0)$ denote the variance of the stationary process, and $\gamma(k)$ denote its auto-covariance at time lag k. For example, $\rho(1)$ measures degree of linear correlation between any two variables that are one time period apart, i.e. between Y_1 and Y_2, or Y_{99} and

Y_{100}, in general between Y_t and Y_{t-1}. In the same manner, $\rho(3)$ measures degree of linear correlation between any two variables that are 3 time periods apart, i.e. between Y_1 and Y_4, or Y_{10} and Y_{13}, in general between Y_t and Y_{t-3}.

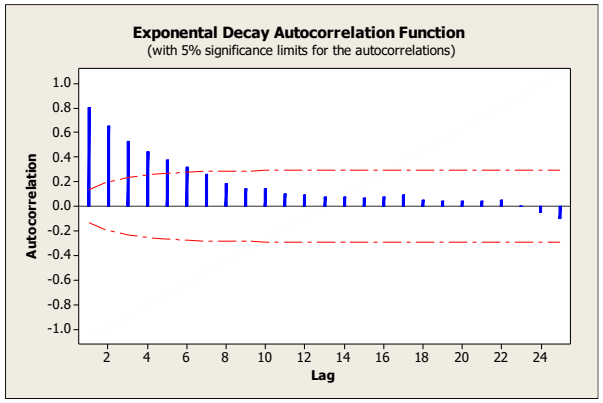
2.2.3 The importance of the autocorrelation function

When analyzing time series, we might face many forms of autocorrelation functions, for example:

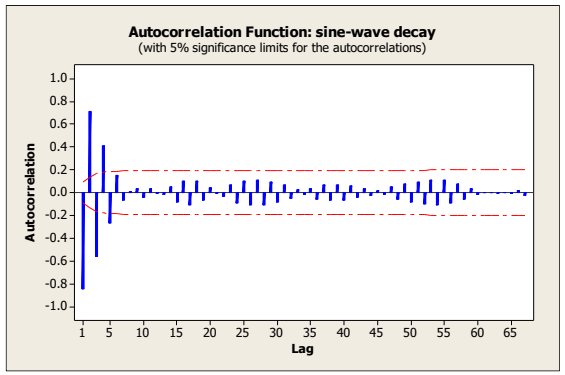
- we might find it decaying slowly.



- or, decaying very quickly in an exponential form.



- or, decaying in sine function form.



- Sometimes it cut off suddenly (i.e. equal zero) after a certain number of time lags.

Autocorrelation function $\rho(k)$, plays an important and essential role when using Box - Jenkins methodology for analyzing time series. As the form of the ACF can determine the initial appropriate model for the data. It is also one of the important tools in diagnostic tests of the residuals of the initial model in order to improve it.

Example*:* Let the random process $\{\varepsilon_t\}$ be uncorrelated random variables with mean zero and constant variance σ^2, find autocorrelation function of the process $\{\varepsilon_t\}$.

Note: $\{\varepsilon_t\}$ is called the "**white noise process**" , and it will be used frequently in this course.

solution:

According to the definition of the process, then:

$$E(\varepsilon_t) = 0.\ \ t = 0.\pm1.\pm2.\ldots$$

$$Var(\varepsilon_t) = \sigma^2.\ \ t = 0.\pm1.\pm2.\ldots$$

$$\gamma(k) = Cov(\varepsilon_t.\varepsilon_{t-k}) = 0.\ \ k \neq 0;\ \ t = 0.\pm1.\pm2.\ldots$$

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = 0.\ \ k \neq 0$$

This means that:

$$\rho(k) = \begin{cases} 1. & k = 0 \\ 0. & k \neq 0 \end{cases}$$

Example:

If the series y_t have the following model:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t. \quad t = 1.2. \ldots . n$$

Where $\{\varepsilon_t\}$ is the white noise process as defined in the previous example. Find autocorrelation function of the series y_t.

solution:

$$Var(y_t) = Var(\beta_0 + \beta_1 t + \varepsilon_t) = Var(\varepsilon_t) = \sigma^2$$

This is because $(\beta_0 + \beta_1 t)$ is not a random variable, but it is a deterministic function.

and,

$$\gamma(s.t) = Cov(\beta_0 + \beta_1 s + \varepsilon_s.\ \ \beta_0 + \beta_1 t + \varepsilon_t) = 0.\ s \neq t$$

$$= Cov(\varepsilon_s.\ \ \varepsilon_t) = 0. \qquad\qquad s \neq t$$

So that,

$$\rho(k) = \begin{cases} 1. & k = 0 \\ 0. & k \neq 0 \end{cases}$$

Example:

If the process $\{y_t\}$ have the following model:

$$y_t = \varepsilon_t - \theta\varepsilon_{t-1}. \quad t = 1.2. \ldots . n$$

Where $\{\varepsilon_t\}$ is the white noise process as defined in the previous example. Find the autocorrelation function of the process $\{y_t\}$.

solution:

$$E(y_t) = 0. \quad t = 1.2. \ldots . n$$

$$Var(Y_t) = Var(\varepsilon_t - \theta\varepsilon_{t-1})$$

$$= Var(\varepsilon_t) + \theta^2 Var(\varepsilon_{t-1}) - 2Cov(\varepsilon_t . \varepsilon_{t-1})$$

$$= \sigma^2 + \theta^2\sigma^2 - 0 = \sigma^2(1 + \theta^2)\,; t = 1.2. \ldots$$

Now, we find the auto-covariance function for observations that are

one time lag apart i.e. $\gamma(1)$:

$$\gamma(t . t + 1) = Cov(y_t . y_{t+1})$$

$$= Cov(\varepsilon_t - \theta\varepsilon_{t-1}.\ \ \varepsilon_{t+1} - \theta\varepsilon_t) = -\theta\sigma^2$$

In the same manner, we find the auto-covariance function for observations that are two time lags apart i.e. $\gamma(2)$:

$$\gamma(t.t+2) = Cov(y_t.Y_{t+2})$$

$$= Cov(\varepsilon_t - \theta\varepsilon_{t-1}.\ \ \varepsilon_{t+2} - \theta\varepsilon_{t+1}) = 0$$

in the same manner, it can also be shown that $\gamma(3) = \gamma(4) = \cdots = 0$

So the auto-covariance function has the form:

$$\gamma(k) = \begin{cases} \sigma^2(1+\theta^2) & k = 0 \\ -\theta\sigma^2 & k = 1 \\ 0. & k \geq 2 \end{cases}$$

thus the auto-correlation function for this process is:

$$\rho(k) = \begin{cases} 1. & k = 0 \\ \dfrac{-\theta}{1+\theta^2}. & k = 1 \\ 0 & k \geq 2 \end{cases}$$

2.2.4 Estimating the Autocorrelation Function

As stated previously the importance of imposing stationarity conditions on the stochastic process that generated the observed time series. The most important was, reduction of the number of major parameters of the process (first and second moments), and easiness of their interpretation, and the possibility of estimating these parameters using the available observations $y_1 . y_2 . \ldots . y_n$ of the time

series. Based on these estimates, we can estimate the sample auto-correlation function for the stationary process as follows:

$$r_k = \hat{\rho}(k) = \frac{\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

It can be shown that if the random process $\{y_t\}$ is stationary and linear, and the fourth moment $E(y_t^4)$ is bounded, then the estimate r_k of the auto-correlation function follow asymptotically a normal distribution with mean ρ_k and a known variance that also depend

on ρ_k. Then it is possible to perform testing of hypothesis for the significance of various auto-correlation coefficients at different time lags.

- Bartlett 1946, has proven that if observations q time lags apart are not correlated, that is,

$$\rho_k = 0. \quad k > q$$

then the sample variance of the statistic r_k can be approximated by:

$$V(r_k) \cong \frac{1}{n}\,(1 + 2\sum_{j=1}^{q} \rho_j^2). \quad k > q$$

Then one can get approximate estimates of standard errors (SE) of the estimators r_k by replacing ρ_k by r_k and taking the square root in the previous form:

$$SE(r_k) \cong \sqrt{\frac{1}{n}\,(1 + 2\sum_{j=1}^{q} r_k^2)} \quad .k > q$$

- In the special case when all observations are uncorrelated, that is $\rho_k = 0.$ for $k > 0$ then this equation simplifies to:

$$SE(r_k) \cong \sqrt{\frac{1}{n}} \; . \; k > q$$

So if we assume that the process $\{y_t\}$ is completely random, that is a white noise process then, for large sample size the distribution of the estimator r_k (according to central limit theorem) is normal distribution with mean ρ_k and variance $\frac{1}{n}$ i.e.,

$$r_k \sim N\left(\rho_k . \frac{1}{n}\right)$$

This means that if the series at hand is completely random, then we can find a 95% Confidence interval for ρ_k, which is:

$$r_k - 1.96\sqrt{var(r_k)} < \rho_k < r_k + 1.96\sqrt{var(r_k)}$$

That is:

$$r_k - 1.96\sqrt{1/n} < \rho_k < r_k + 1.96\sqrt{1/n}$$

- Anderson in 1942 have shown that for a sample of moderate size and assuming that the estimator $\rho_k = 0.$ then the sample estimator r_k follows approximately the normal distribution, and thus the statistic:

$$z = \frac{r_k - 0}{SE(r_k)}$$

follows approximately standard normal distribution under the hypothesis $\rho_k = 0$, thus it can be used to test the hypothesis:

$$H_0: \rho_k = 0 \quad \text{vs} \quad H_1: \rho_k \neq 0 \quad \text{for } k > q$$

We reject the null hypothesis, at significance level α if $|z| > z_{\alpha/2}$.

Note:

It has been the norm in practical applications to reject the null hypothesis:

$\rho_k = 0$, if $|z| > 2$ assuming that $\alpha = 0.05$,

but it should be noted that it is not always preferable to fix α at a certain value to test the significance of the autocorrelation coefficients for all time lags. Some recent studies have concluded that it is preferable to use larger values for α at lower time lags, and then use smaller values for α at larger time lags. Choosing the right value of α, depends actually more on the expertize of the researcher, and how he reads the different graphs of the data.

Example:

The following data represents the number of sold units (percentage) yearly at a large department stores:

Year	1992	1993	1994	1995	1996	1997	1998	1999
y_t	1	3	2	4	3	2	3	2

Calculate the autocorrelation coefficients, and draw the estimated autocorrelation function.

solution:

One can easily calculate:

$$\bar{y} = \frac{20}{8} = 2.5 \quad ; \quad \sum_{t=1}^{8} (y_t - 2.5)^2 = 6$$

Also we can find the pairs $(y_t - 2.5)$:

Year	1992	1993	1994	1995	1996	1997	1998	1999
$(y_t - 2.5)$	-1.5	0.5	-0.5	1.5	0.5	-0.5	0.5	-0.5

According to the definition of autocorrelation function r_k, then:

$$r_1 = \hat{\rho}(1) = \frac{\sum_{t=1}^{7}(y_t - 2.5)(y_{t+1} - 2.5)}{6}$$

$$r_1 = \frac{1}{6}[(-1.5)(0.5) + (0.5)(-0.5) + (-0.5)(1.5) + (1.5)(0.5)$$

$$+ \ (0.5)(-0.5) + (-0.5)(0.5) + (0.5)(-0.5)] = -0.29$$

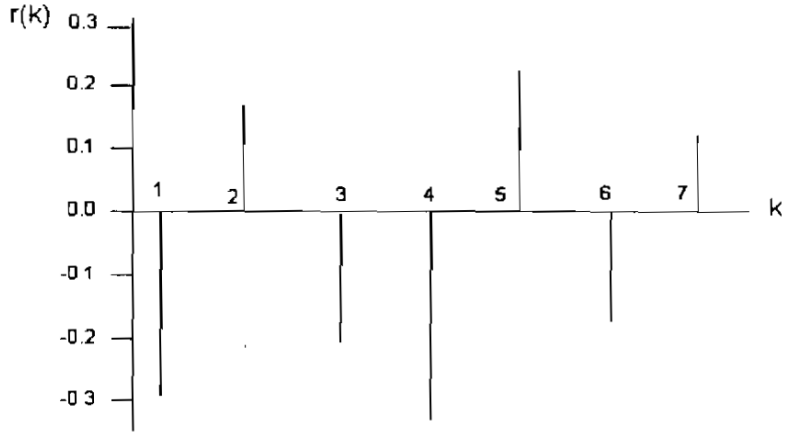
Also,

$$r_2 = \hat{\rho}(2) = \frac{\sum_{t=1}^{6}(y_t - 2.5)(y_{t+2} - 2.5)}{6} = 0.17$$

Similarly, the rest of the values are calculated:

$r_3 = -0.21.\ r_4 = -0.33.\ r_5 = 0.21.\ r_6 = -0.17.\ r_7 = 0.13$

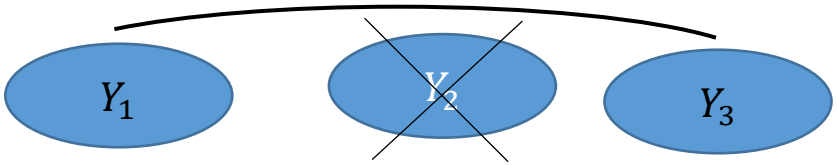
The auto-correlation function can be drawn such that, on the horizontal axis the time lags, k, and on the vertical axis auto-correlation coefficients, this figure is called the *correlogram*.



2.3 Partial autocorrelation function

The idea of this correlation arise as follows:

If two variables, say, Y_1 and Y_3 are found to be correlated , then this might be because of correlation between them and a third



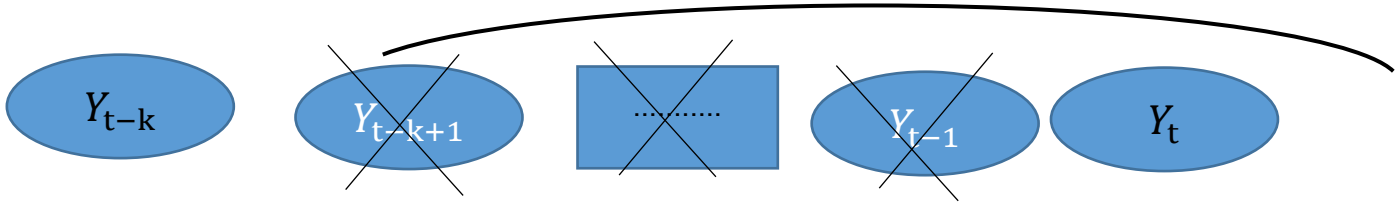
variable, Y_2 , so if we can calculate correlation between Y_1 and Y_2 ,

and correlation between Y_3 and Y_2, and remove or control this correlation, then the resulting correlation is called *partial auto-correlation* .

The autocorrelation between Y_1 and Y_3 where the effect of Y_2 has been removed or controlled is called the *partial auto-correlation* between Y_1 and Y_3 .

This idea can be applied to any number of variables, such that the correlation between any two variables with the removal of the effect of variables that falls between them.

One can calculate the auto-correlation between the two variables Y_t and Y_{t-k}, and removing or controlling the effect of all the variables that fall between them, i.e. $(Y_{t-k+1}. \ldots . Y_{t-1})$, this is called the *partial auto-correlation* between Y_t and Y_{t-k}.



The basic idea behind the partial auto-correlation is calculating the linear correlation coefficient between $[Y_t - E(Y_t|Y_{t-1}.\ldots.Y_{t-k+1})]$ and $[Y_{t-k} - E(Y_{t-k}|Y_{t-1}.\ldots.Y_{t-k+1})]$

Where $E(Y_t|Y_{t-1}.\ldots.Y_{t-k+1})$ and $E(Y_{t-k}|Y_{t-1}.\ldots.Y_{t-k+1})$ are calculated from the corresponding conditional probability distributions .

2.3.1 Yule-Walker system of equations

Assuming that we have a stationary process with mean equal to zero, we can write a multiple regression model of order p as Follows:

$$Y_t = \phi_{11} Y_{t-1} + \phi_{22} Y_{t-2} + \cdots + \phi_{kk} Y_{t-p} + \varepsilon_t$$

Where ε_t is the white noise process, multiplying both sides by Y_{t-k}, and taking expectations, we find:

$$E(Y_t Y_{t-k})$$

$$= \phi_{11} E(Y_{t-1} Y_{t-k}) + \phi_{22} E(Y_{t-2} Y_{t-k}) + \cdots + \phi_{kk} E(Y_{t-p} Y_{t-k})$$

$$+ E(\varepsilon_t Y_{t-k})$$

i.e,

$$\gamma_k = \phi_{11}\gamma_{k-1} + \phi_{22}\gamma_{k-2} + \cdots + \phi_{kk}\gamma_{k-p}$$

And dividing both sides by γ_0 , we find:

$$\rho_k = \phi_{11}\rho_{k-1} + \phi_{22}\rho_{k-2} + \cdots + \phi_{kk}\rho_{k-p} \ , k \geq 1$$

This is called the Yule-Walker system of equations, and consists of a k linear equation in the unknowns $\phi_{11}, \phi_{22}, \ldots, \phi_{kk}$. We can solve this system by the determinants to get ϕ_{kk} (The

mathematical derivation details for this is not the concern of this course) :

$$\phi_{kk} = \begin{cases} 1 & .k = 0 \\ \rho_1 & .k = 1 \\ \dfrac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & 1 \end{vmatrix}} & .k = 2.3.\ldots \end{cases}$$

Where | | denote the determinant.

We note that for large values of k, the above solution is difficult to find, thus another approach that uses recurrence relations is proposed in the literature, as follow:

$$\phi_{00} = 1$$

$$\phi_{11} = \rho_1$$

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_j}$$

Where,

$$\phi_{kj} = \phi_{k-1.j} - \phi_{kk}\phi_{k-1.k-j} \quad .j=1.2.....k-1$$

2.3.2 Properties of partial autocorrelation function (PACF)

This function has several properties, including:

1- partial autocorrelation coefficient at time lag zero is equal to one, that is, $\phi_{00} = 1$.

2- The value of ϕ_{kk} always fall in the closed interval $[-1,1]$.

3- $\phi_{11} = \rho_1$, this is because there are no observations fall between Y_{t-1} and Y_t .

4- If $\phi_{kk} = 0$, then this means there is no linear partial autocorrelation between Y_{t-k} and Y_t , however, there might be a nonlinear partial autocorrelation between them.

2.3.3 Estimating the partial autocorrelation function

One can get the sample partial autocorrelation function from the previous equations by replacing ϕ_{kk} by r_{kk} , and ρ_k by r_k.

The statistic r_{kk} is an estimator for ϕ_{kk} i.e.:

$$\hat{\phi}_{kk} = r_{kk} \quad .k = 0.1. \ldots$$

To function r_{kk} has the following properties:

1- Anderson and Quenouille (1949) have found that if the partial correlation coefficient $\phi_{kk} = 0$, and for a large sample size, then the estimated sample partial autocorrelation coefficients r_{kk} follow the normal distribution with estimated standard error:

$$se(r_{kk}) \cong \sqrt{\frac{1}{n}}, \quad k>0$$

2- For large sample size n, we can carry out the following test:

$$H_0: \phi_{kk} = 0$$

$$H_1: \phi_{kk} \neq 0$$

Where we use the statistic:

$$Z = \frac{|r_{kk}| - 0}{\sqrt{\frac{1}{n}}} = \sqrt{n}\ |r_{kk}|$$

and reject H_0 at significance level α, if $|Z| > z_{\alpha/2}$

Example:

The following data represent the daily demand of a particular product:

158 222 248, 216 226 239, 206 178 169

Calculate the autocorrelation function and partial autocorrelation function and draw them.

solution:

1- Finding the autocorrelation function r_k:

First we calculate the mean of the series:

$$\bar{y} = \frac{1}{9}\sum Z_i = \frac{1}{9}[158 + \cdots + 169] = 206.89$$

sample autocorrelation function has the form:

$$r_k = \frac{\sum_{t=k+1}^{9}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{9}(y_t - \bar{y})^2} . k = 0.1. \ldots.$$

We need to find the quantities:

$$r_1 = \frac{\sum_{t=2}^{9}(y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^{9}(y_t - \bar{y})^2}, \ldots\ldots\ldots\ldots\ldots\ldots,$$

$$r_8 = \frac{\sum_{t=9}^{9}(y_t - \bar{y})(y_{t-8} - \bar{y})}{\sum_{t=1}^{9}(y_t - \bar{y})^2}$$

Which means that if we have n observations, then we need to calculate $(n-1)$ coefficients of r_k. To simplify calculations, we will find first the following pairs, $(y_t - \bar{y}) = (y_t - 206.89)$ as follow:

$$(158 - 206.89).(222 - 206.89).\ldots.(169 - 206.89)$$

$$\Longrightarrow (-48.89).(15.11).(41.11).(9.11)\ \ldots.(-37.89)$$

Then we get the required r_k coefficients as follow:

$$r_1 = \frac{(-48.89 \times 15.11) + (15.11 \times 41.11) + \cdots + (-28.89 \times -37.88)}{(-48.89)^2 + (15.11)^2 + \cdots + (-37.89)^2} = 0.2651$$

$$r_2 = \frac{(-48.89 \times 41.11) + (15.11 \times 9.11) + \cdots + (-0.89 \times -37.88)}{(-48.89)^2 + (15.11)^2 + \cdots + (-37.89)^2} = -0.212$$

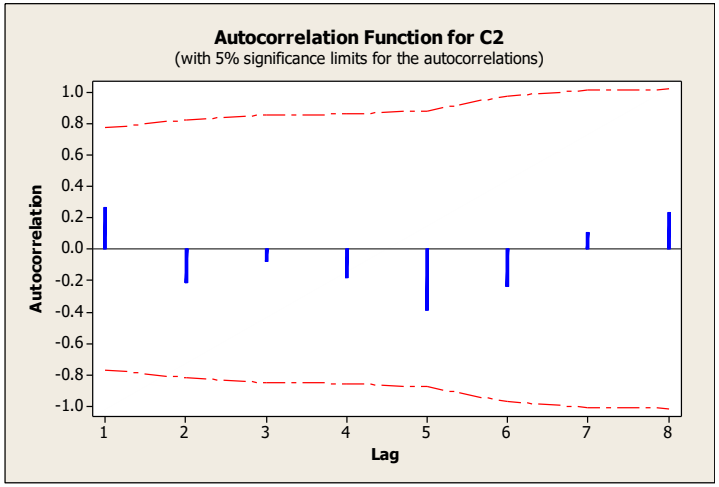
And the same for other coefficients,

$$r_3 = -0.076\,.\ \ r_4 = -0.183\,.\ r_5 = -0.387.$$

$$r_6 = -0.242,$$

$$r_7 = 0.104.\ r_8 = 0.230$$

Drawing the correlogram , we have:



The following table shows the result of calculations in the *Minitab* :

Autocorrelation Function: C2

Lag	ACF	T
1	0.265116	0.80
2	-0.211557	-0.59
3	-0.076111	-0.21
4	-0.182772	-0.49
5	-0.386675	-1.01
6	-0.242061	-0.57

```
7   0.104208   0.24
8   0.229851   0.52
```

We can also estimate the variance of r_k from relationship:

$$\hat{V}(r_k) \cong \frac{1}{n}\left(1 + 2\sum_{j=1}^{q} r_j^2\right). \quad q < k$$

Then:

$$\hat{V}(r_1) \cong \frac{1}{9}\left(1 + 2\sum_{j=1}^{0} r_j^2\right). \quad q < 1$$

$$\cong \frac{1}{9}(1 + 2r_0^2) = \frac{1}{9}(1 + 2(0)^2) = \frac{1}{9} = 0.11$$

$$\hat{V}(r_2) \cong \frac{1}{9}\,(1 + 2\sum_{j=1}^{1} r_j^2). \quad q < 2$$

$$\cong \frac{1}{9}(1 + 2r_1^2) = \frac{1}{9}(1 + 2(0.2651)^2) = 0.12$$

and the same for the rest of the values we get:

$$\hat{V}(r_3) \cong \frac{1}{9}(1 + 2r_1^2 + 2r_2^2) \cong 0.1367$$

$\hat{V}(r_4) \cong 0.138$, $\hat{V}(r_5) \cong 0.1454$, $\hat{V}(r_6) \cong 0.1787.$

$\hat{V}(r_7) \cong 0.1931.\ \hat{V}(r_8) \cong 0.2013.$

We note that the as time lag between the variables increase, then the variance of the estimated correlation coefficients increases.

2- Finding the partial autocorrelation r_{kk}:

$r_{00} = 1$

$r_{11} = r_1 = 0.265.$

And the rest of the coefficients are found through the recurrence relation:

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1.j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1.j} r_j}. \quad k = 2.3. \ldots$$

Where,

$$r_{kj} = r_{k-1.j} - r_{kk} r_{k-1.k-j} \quad .j = 1.2. \ldots . k - 1$$

So,

$$r_{22} = \frac{r_2 - \sum_{j=1}^{1} r_{1.j}\ r_{2-j}}{1 - \sum_{j=1}^{1} r_{1.j}\ r_j} = \frac{r_2 - r_{11}r_1}{1 - r_{11}r_1}$$

$$= \frac{(-0.212)-(-0.265)(0.265)}{1-(-0.265)(0.265)} = -0.304$$

$$r_{33} = \frac{r_3 - \sum_{j=1}^{2} r_{2.j}\ r_{3-j}}{1 - \sum_{j=1}^{2} r_{2.j}\ r_j} = \frac{r_{3-}[r_{21}r_2 + r_{22}r_1]}{1 - [r_{21}r_1 + r_{22}r_2]}$$

So we need the value of r_{21}:

$$r_{21} = r_{11} - r_{22}r_{11} = 0.345$$

Thus,

$$r_{33} = \frac{-0.076 - [(0.345)(-0.212) + (-0.304)(0.265)]}{1 - [(0.345)(0.265) + (-0.304)(-0.212)]} = 0.092$$

The same calculations for the other values:

$r_{44} = -0.298$

$r_{55} = -0.294$

$r_{66} = -0.207$

$r_{77} = 0.013$

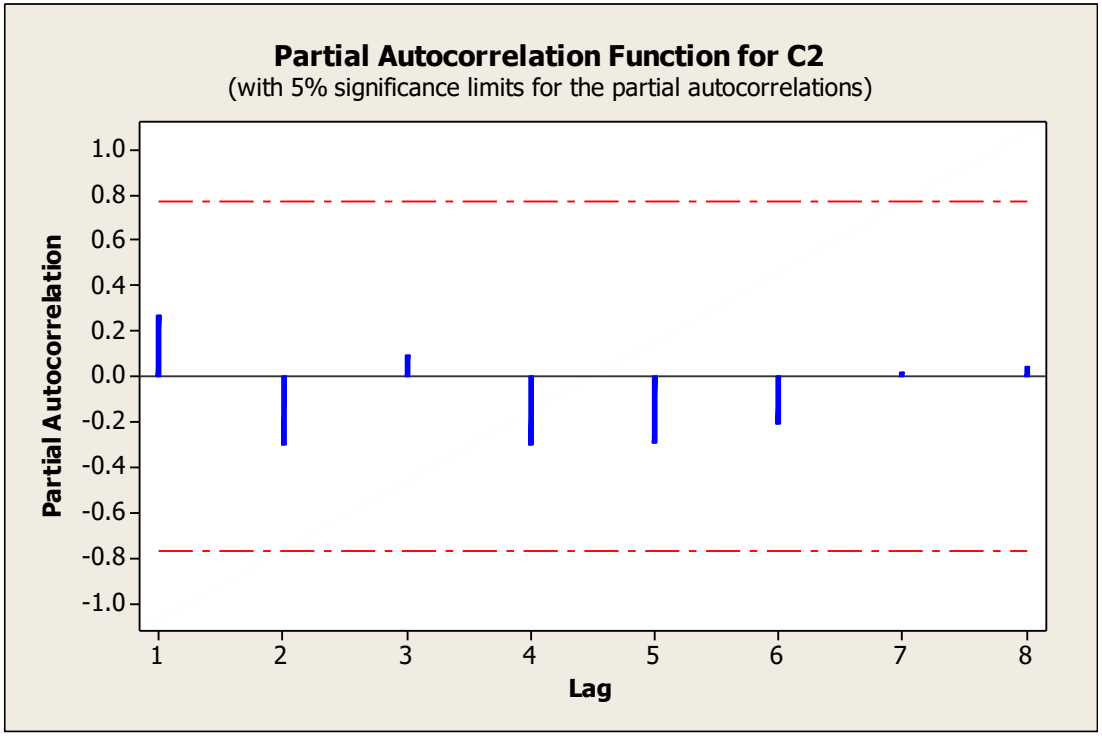
$r_{88} = 0.042$

The variance of these coefficients is estimated by:

$$\hat{V}(r_{kk}) \cong \frac{1}{n} = \frac{1}{9}$$

The following table shows the result of calculations in the Minitab:

Partial Autocorrelation Function: C2		
Lag	ACF	T
1	0.265116	0.80
2	-0.303151	-0.91
3	0.091617	0.27
4	-0.298000	-0.89
5	-0.294454	-0.88
6	-0.206605	-0.62
7	0.013411	0.04
8	0.042363	0.13



2.4 Time series operators

Proper understanding of Box-Jenkins methodology (which will be discussed later) depends on understanding how some important

operators work, such as difference operator, and backshift operator.

2.4.1 Backshift operator

If the value of the series at time t is y_t, and at time r is y_r, then the backshift operator B, is defined as follow:

$$By_t = y_{t-1}$$

$$B^2 y_t = By_{t-1} = y_{t-2}$$

$$\vdots$$

$$B^r y_t = y_{t-r} \,.\quad r = 1.2. \ldots$$

For example, for the model:

$$y_t = y_{t-1} + e_t$$

It can be rewritten using the backshift operator as follows:

$$y_t - y_{t-1} = e_t \Rightarrow y_t - By_t = e_t \Rightarrow (1 - B)y_t = e_t$$

The backshift operator plays an important role in the algebraic manipulations when working with Box-Jenkins methodology, where it is used in polynomial forms, such as:

1- Autoregressive operator

This is defined as:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

Where $\phi(B)$ is a polynomial of order p in the operator B, and $\phi_1.\phi_2.\ldots.\phi_p$ are constants.

The polynomial $\phi(B)$ is used with values of the time series y_t as follows:

$$\phi(B)y_t = y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_p y_{t-p}$$

2- Moving Averages operator

This is defined as:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

Where $\theta(B)$ is a polynomial of order q in the operator B, and $\theta_1.\theta_2.\ldots.\theta_q$ are constants.

The polynomial $\theta(B)$ is used with values of the white noise process ε_t as follows:

$$\theta(B)\varepsilon_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q}$$

2.4.2 Difference operator

This operator is denoted as ∇, an is defined as follows:

If we have a time series y_t, then the difference operator is defined as:

$$\nabla y_t = y_t - y_{t-1}$$

$$\nabla^2 y_t = \nabla\nabla y_t = \nabla(y_t - y_{t-1})$$

$$= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$$

$$= y_t - 2y_{t-1} + y_{t-2}$$

The relationship between the backshift and difference operators can be noted from the following relation:

$$\nabla = (1 - \mathrm{B})$$

and in general,

$$\nabla^r y_t = (1 - \mathrm{B})^r y_t$$

For example if we applied this relation to find $\nabla^2 y_t$, we get:

$$\nabla^2 y_t = (1 - \mathrm{B})^2 y_t$$

$$= (1 - 2\mathrm{B} + \mathrm{B}^2) y_t$$

$$= y_t - 2y_{t-1} + y_{t-2}$$

which is the same result that we found previously.

2.5 Transformations for non-stationary time series

Time series in many applications are often not stationary in the mean, where we find the level of the series is either increasing or decreasing with time. It is also possible to find some series that have variance changing with time, and it is possible to have both forms of non-stationarity to exist in a time series. However, luckily, in many situations it is possible to transform the time series into a stationary series through simple transformations. In this case we call the time series as homogeneous stationary time series.

In what follows we will cast some light on some of the most important mathematical transformations used to transform the nonstationary stochastic models into stationary ones.

2.5.1 Differences of the series

If the observed time series y_t shows some trend component –either deterministic or stochastic- then taking the first differences of y_t usually succeeds in transforming the series into a stationary series, so if we denote the resulting series as z_t, then:

$$z_t = \nabla y_t = y_t - y_{t-1} \,.\, t = 2.3. \ldots . n$$

Where n denote the number of observations available, or what is called the length of the series. So, if the observations of the nonstationary series are $y_1 . y_2 . \ldots . y_n$, then the first differences are found as follows:

y_t	y_{t-1}	$z_t = \nabla y_t = y_t - y_{t-1}$
y_1	-	-
y_2	y_1	$z_2 = y_2 - y_1$
y_3	y_2	$z_3 = y_3 - y_2$
$\vdots$	$\vdots$	$\vdots$
y_n	y_{n-1}	$z_n = y_n - y_{n-1}$

And as we note that taking differences of first order, we lose one observation, and taking difference of order two, we lose two observations, etc.

Example:

If the series y_t follow the following model:

$$y_t = \beta_0 + \beta_1 t + W_t. \quad t = 1.2. \dots n$$

Where $\{W_t\}$ is a stationary process having a mean μ, variance σ^2, and covariance function γ_k, prove that the series $y_t$is not stationary. How would you transform it to a stationary series?

Solution:

$$E(y_t) = \beta_0 + \beta_1 t + \mu$$

It is clear that the mean changes with time, Therefore the series is not stationary in the mean. Now we find the following:

$$y_{t-1} = \beta_0 + \beta_1 (t-1) + W_{t-1}$$

From which we can find the first differences:

And we could create the first series of the differences ∇y_t:

$$z_t = \nabla y_t = y_t - y_{t-1}$$

$$= [\beta_0 + \beta_1 t + W_t] - [\beta_0 + \beta_1 (t-1) + W_{t-1}]$$

$$= \beta_1 + W_t - W_{t-1} \ . t = 2.3. \ldots . n$$

Now we can see the effect of taking the first differences transformation on the series:

$$E(z_t) = \beta_1 + \mu - \mu = \beta_1$$

This means that the series z_t is stationary in the mean.

Also, its variance is:

$$Var(z_t) = var(\beta_1 + W_t - W_{t-1}) = Var(W_t - W_{t-1})$$
$$= Var(W_t) + Var(W_{t-1}) - 2Cov(W_t . W_{t-1})$$
$$= 2\sigma^2 - 2\gamma_1$$

Which is free of time t, so z_t is stationary in the variance. We can also, see the effect of difference operator on the auto-covariance function, lets denote the auto-covariance function for transformed series z_t as $\gamma_z(k)$, then:

$$\gamma_z(1) = Cov(z_t . z_{t-1})$$

$$= Cov(\,[\beta_1 + W_t - W_{t-1}].[\beta_1 + W_{t-1} - W_{t-2}])$$

$$= Cov(W_t . W_{t-1}) - Cov(W_t . W_{t-2}) - Cov(W_{t-1} . W_{t-1})$$

$$+ Cov(W_{t-1} . W_{t-2})$$

$$= \gamma_1 - \gamma_2 - \sigma^2 + \gamma_1 = 2\gamma_1 - \gamma_2 - \sigma^2$$

Which means that $\gamma_z(1)$ does not depend on time t.

Similarly, we can find $\gamma_z(2)$:

$$\gamma_z(2) = Cov(z_t . z_{t-2})$$

$$= Cov(\,[\beta_1 + W_t - W_{t-1}].[\beta_1 + W_{t-2} - W_{t-3}])$$

$$= Cov(W_t . W_{t-2}) - Cov(W_t . W_{t-3}) - Cov(W_{t-1} . W_{t-2}) +$$

$$Cov(W_{t-1} . W_{t-3})$$

$$= \gamma_2 - \gamma_3 - \gamma_1 + \gamma_2 = 2\gamma_2 - \gamma_3 - \gamma_1$$

Which means that $\gamma_z(2)$ does not depend on time t. Generally, we can show that:

$$\gamma_z(k) = 2\gamma_k - \gamma_{k+1} - \gamma_{k-1}$$

Hence, since the auto-covariance function depend on time lag k, and not on time t, so the series z_t is stationary.

Example:

If the series y_t can be modeled as:

$$y_t = y_{t-1} + \varepsilon_t \ . t = 1.2. \ldots . n$$

Where ε_t is the white noise process. Show that the series y_t is not stationary. How can you transform it to a stationary process?

Solution:

$$E(y_t) = E(y_{t-1}) + 0$$

So the series is stationary in the mean, since the mean function does not depend on time t.

And for the variance:

$$Var(y_t) = Var(y_{t-1}) + Var(\varepsilon_t) + 2Cov(y_t . \varepsilon_t)$$

$$= Var(y_{t-1}) + \sigma^2 + 0$$

Which indicate that $Var(y_t) \neq Var(y_{t-1})$, so the series is not stationary in the variance. Now, we can try to apply the first differences operator to try to stabilize it:

Subtracting y_{t-1} from both sides of the model equation, we get

$$y_t - y_{t-1} = y_{t-1} + \varepsilon_t \ - y_{t-1}$$

i.e.

$$\nabla y_t = \varepsilon_t$$

so the first difference operator transformed the series into white noise series, which is stationary series by definition.

But this is not always the case, as sometimes the variance might increase or decrease with time, in this case we might need a different tool for stabilizing the series. Some of common transformations for stabilizing the variance are mentioned in the following section.

2.5.2 Variance stabilizing transformations

- Logarithmic

- Square root
- Reciprocal

The logarithmic transformation is used if the variance of the series is increasing or decreasing with time, and the mean is almost constant. It is assumed that the values of the observations are all positive (since the logarithm is only defined for positive numbers). It is also possible to use the square root or reciprocal transformation or any other transformation from the Box-Cox family of transformations. However, the logarithmic transformation is the most commonly used one in such cases.

The most important case of non-stationarity, is the one in which lack of stationarity happens in both mean and variance together. Many examples in economic, social and demographic fields can have their values at time t are greater than their value at time $t-1$ with a constant rate, plus a component of random errors. In such cases we can represent the series approximately in the form:

$$y_t = \alpha\, y_{t-1} + y_{t-1}\,.\, 0 < \alpha < 1$$

This kind of series features a growing trend in both mean and variance, and almost constant growing rate of the phenomenon. To use the logarithmic transformation, we rewrite the model as:

$$y_t = (1 + \alpha)\, y_{t-1}$$

Taking logarithm of both sides, we find,

$$ln(y_t) = ln(1 + \alpha) + ln(\, y_{t-1})$$

Subtracting $ln(\, y_{t-1})$ from both sides, we find,

$$ln(y_t) - l\, n(y_{t-1}) = ln(1 + \alpha) = \delta$$

Where δ is a constant quantity, this means that :

$$z_t = \nabla ln(y_{t)} = ln(y_t) - l\, n(y_{t-1}) = \delta$$

so the first differences of the logarithm of the data turned it into a stationary process.

Notes: It is recommended not to use this type of transformation before the use of the normal differences of the data, and if the

normal differences failed to stabilize the variance, then we resort to logarithmic transformation.

- You must make sure that all the values of the series are positive before using this transformation. In case there were negative values in the data, then you can, for example, address this problem by adding a certain constant term for each value so that all values become positive.

Note that the addition of a constant term to a variable does not affect the variance and the autocorrelation function of this variable, and therefore this process will not affect the

autocorrelation structure of the series while helping to ensure stationary.

The transformed series can be studied and analyzed, and after the completion of the analysis, researcher should reverse the transformation process so that the results be consistent with the data he wanted to analyze in the first place.

- In some cases, the first difference transformation may still be not stationary, and therefore we may need to take the second differences of logarithms to stabilize the series.

2.5.3 Box-Cox transformations

This family of transformations are common in the field of design of experiments, it takes the following form:

$$g(x) = \begin{cases} \dfrac{x^{\lambda} - 1}{\lambda}. & \lambda \neq 0 \\ \ln(x)\ . & \lambda = 0 \end{cases}$$

Note: Subtracting 1, and dividing by λ, makes the function $g(x)$ change smoothly as λ approach zero. As we know from calculus that $\lim_{\lambda \longrightarrow 0} \frac{x^{\lambda}-1}{\lambda} = \ln(x)$. Also, note that choosing $\lambda = 0.5$, impose a square root transformation, this is useful if the data are

count data that follows Poisson distribution, and $\lambda = -1$ is the reciprocal of the data.

Chapter 3: Random Time Series Models

3.1 Meaning of linearity in regression models and in time series models

As we know in regular regression models of the form:

$$y = f\left(x_1 . x_p . \ldots . x_p + \beta_0 . \beta_1 . \ldots \beta_p\right) + \varepsilon$$

We mean by linearity, the linearity in coefficients, *or main parameters* $\boldsymbol{\beta} = (\beta_0 . \beta_1 . \ldots \beta_p)^T$ regardless of the shape of the explanatory variables $\boldsymbol{x} = (x_1 . x_p . \ldots . x_p)^T$. For instance, the simple linear regression $y = \beta_0 + \beta_1 x_1 + \varepsilon$ is linear regression model

because it is linear in the parameters, the same is true for the models $y = \beta_0 + \beta_1 x_1^2 + \varepsilon$, and $y = \beta_0 + \beta_1 \ln(x) + \varepsilon$ they are all linear in the parameters, note we can redefine the explanatory variables as $w = x_1^2$ for the first model, then it takes the form of a liner model, and the estimating equations for the parameters are the same and will not be affected,

$\hat{\beta} = (w^T w)^{-1} w^T y$. Whereas, the model $y = \beta_0 + \beta_1^2 x_1 + \varepsilon$ is not linear, because it is not linear in the parameter β_1 and thus general regression rules can't be applied here.

On the other hand, in time series context there exist many form of functions that relate the values of the variable under study y_t with its previous values $y_{t-1}. y_{t-2}. \dots$ and the values of a completely random variables we called them white noise $\varepsilon_t. \varepsilon_{t-1}. \varepsilon_{t-2}. \dots$. We will only study the linear time series models in this course.

Linearity in the time series context is completely different than linearity in regression context, as it mean here linearity in the

explanatory variables $y_{t-1}. y_{t-2}. \ldots$ but not in the model parameters. It is interesting to know that most of the time series models are not linear in the parameters! and this is one of the difficulties in studying time series.

3.2 Static and dynamic models

Traditional regression models of the form:

$y_t = f(x_1. x_p. \ldots . x_p. \beta_0. \beta_1. \ldots \beta_p) + \varepsilon_t$ applied to time series data is considered static models , that is they are not dynamic. Since the

model $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$ depend on the variable ε_t which represent the disturbance that affect the system at time t, but its effect does not extend to time period $(t + 1)$ because the system at time $(t + 1)$ is affected by ε_{t+1} only , and this variable is not correlated with ε_t (this result from the definition of the white noise process), so such systems have no memory, in the sense that it completely "forget" disturbances that occurred in the past , so such systems are called "*static systems*".

Time series random models, on the other hand, depend on the history of the series $y_{t-1}. y_{t-2}. \ldots$ or on the disturbances occurred in the past $\varepsilon_{t-1}. \varepsilon_{t-2}. \ldots$, or on both of them as explanatory variables. Thus, these models consist of three main groups of models. The first is known as the autoregressive models, and it is such a models where the variables $y_{t-1}. y_{t-2}. \ldots$ plays the role of explanatory variables that affect the dependent variable y_t.

The simplest of these models is the autoregressive model of order one, which takes the form:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \ ; t = 1.2. \ldots . n$$

It might be thought at first glance of the model that the system at time t depends on the variable ε_t only, and not on previous disturbance ε_{t-1}, but checking the model carefully, then we would notice that the model depend on ε_{t-1} through y_{t-1}, since this variable (according to the model formula) can be written as:

$$y_{t-1} = \beta_0 + \beta_1 y_{t-2} + \varepsilon_{t-1}$$

Thus the system actually does not forget the random variable ε_{t-1}, in fact it does not forget all the disturbances $\varepsilon_{t-1}. \varepsilon_{t-2}. \ldots$(by continue substituting in the model). Thus the autoregressive model belongs to the dynamic systems.

The second group of the random time series models is called the moving average models, and it is a more complicated models than the autoregressive models, where the system at time t is related directly to the disturbances $\varepsilon_{t-1}. \varepsilon_{t-2}. \ldots$ that occurred in the past. Hence these models have memory, and belong to the dynamic systems.

The simplest of these models is the moving average model of order one, which takes the form:

$$y_t = \beta_0 + \varepsilon_t + \beta_1 \varepsilon_{t-1} \; ; t = 1.2. \ldots . n$$

The third group of random time series models contains both autoregressive and moving average parts, where the system at time

t depends on disturbances $\varepsilon_{t-1}.\varepsilon_{t-2}.\ldots$ and on the history of the phenomenon $y_{t-1}.y_{t-2}.\ldots$, the simplest example of those models is the autoregressive-moving average model of order 1, denoted shortly as ARMA(1,1) :

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t + \beta_2 \varepsilon_{t-1} \; ; t = 1.2.\ldots.n$$

3.3 Linear Stochastic Processes

Dynamic models assume presence of a particular form of autocorrelation between the observations of the time series that belong to the processes that follow the behavior of such

models. This might cause some difficulties when dealing with these time series, especially if the autocorrelation coefficients are large. This led the scientists to explore the possibility of studying such processes through a simpler process.

Wold (1938) has published his theory indicating that:

"Every stationary process can be expressed as a linear combination of uncorrelated random variables with mean zero and constant variance σ^2"

3.3.1 Definition of the general linear process

The random process $\{y_t\}$ is called **general linear process** if it is possible to express it in the form:

$$y_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots . \; t = 0. \pm 1. \pm 2. \ldots$$

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_{\mathrm{j}} \varepsilon_{t-\mathrm{j}}$$

Where $\{\varepsilon_t\}$ is the white noise process, μ is a constant, and $\{\psi_t\}$ is a sequence of fixed vales. The process $\{y_t\}$is stationary if one of the following conditions is satisfied:

1- The constants $\psi_1.\psi_2.\ldots$ are finite.

2- The constants $\psi_1.\psi_2.\ldots$ are not finite, but they are asymptotic and fulfill the condition $\sum_{i=0}^{\infty}\psi_i^2 < \infty$, this ensures the variance to be finite. If the process $\{y_t\}$ is stationary then μ is considered the mean of the process, otherwise it is just considered a reference point. In most of the course we will assume $\mu = 0$, this will not affect our discussions of the

different models we will consider), and in case $\mu \neq 0$ we will assume that y_t represent the original series after subtracting the constant μ.

3.4 Invertibility formula

Under certain conditions , we can express the general linear process as a weighted sum of the history of the process $y_{t-1}. y_{t-2}. \dots$ and the current disturbance value ε_t. This formula is known as the $\pi - \text{weights}$ formula, it takes the following form:

$$y_t = \varepsilon_t + \pi_1 y_{t-1} + \pi_2 y_{t-2} + \cdots.$$

And in short as:

$$(1 - \pi_1 \mathrm{B} - \pi_2 \mathrm{B}^2 - \pi_2 \mathrm{B}^2 - \cdots) y_t = \varepsilon_t$$

Or,

$$\boxed{\pi(B) y_t = \varepsilon_t \qquad (1)}$$

Where,

$$\pi(B) = (1 - \pi_1 \mathrm{B} - \pi_2 \mathrm{B}^2 - \pi_2 \mathrm{B}^2 - \cdots)$$

$$\pi(B) = 1 - \sum_{i=1}^{\infty} \pi_i \mathrm{B}^{\mathrm{i}}$$

The constants $\pi_1 . \pi_2 . \ldots$ represent the weights or importance of the variables representing history of the process $y_{t-1} . y_{t-2} . \ldots$. If the

number of the weights that is not equal to zero is limited then we get what we call the autoregressive models of certain order, such as AR (1), and AR (2), these models can be stationary or not, we will discuss this later.

3.5 White noise formula

In the same manner, we can express the general linear process as a weighted sum of the current and past values of the disturbances

$\varepsilon_t . \varepsilon_{t-1} . \varepsilon_{t-2} . \ldots$. This formula is known as ψ-weights formula, it takes the following form:

$$y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots.$$

and in short as:

$$y_t = (1 + \psi_1 \mathrm{B} + \psi_2 \mathrm{B}^2 + \psi_2 \mathrm{B}^2 + \cdots)\varepsilon_t$$

Or,

$$\boxed{y_t = \psi(B)\varepsilon_t \qquad (2)}$$

Where,

$$\psi(B) = \sum_{i=0}^{\infty} \psi_i \mathrm{B}^{\mathrm{i}} \; ; \quad \psi_0 = 1$$

The constants $\psi_1 . \psi_2 . \ldots$ represent the weights or importance of the variables representing the past disturbances $\varepsilon_{t-1} . \varepsilon_{t-2} . \ldots$. If the number of the weights that is not equal to zero is limited then we get what we call the moving average models of certain order, such as MA (1), and MA (2). The polynomial $\psi(B)$ is called the transfer function, or the linear filter that associates the random process $\{y_t\}$ with the white noise process $\{\varepsilon_t\}$. The function $\psi(B)$ is considered

as a generating function for the constants ψ_i, because the coefficient of B^i in the expansion of $\psi(B)$ represent the weights ψ_i.

Also, the relation between the two polynomials $\psi(B)$ and $\pi(B)$ can be found by substituting ε_t from (1) into (2):

$$y_t = \psi(B)\,\pi(B) y_t$$

Which means,

$$1 = \psi(B)\,\pi(B)$$

And thus,

$$\pi(B) = \psi^{-1}(B)$$

Example:

For the model $y_t = \varepsilon_t + 0.5\, y_{t-1}$, find the first three $\pi - weights$

and first three $\psi -$ weights.

solution:

Using the formula $y_t = \varepsilon_t + \pi_1 y_{t-1} + \pi_2 y_{t-2} + \cdots,$ we find:

$$\pi_1 = 0.5 \; ; \pi_2 = \; \pi_3 = 0$$

And for $\psi - \text{weights}$, we find:

$$y_t = \varepsilon_t + 0.5\, y_{t-1} \qquad \text{(i)}$$

Also,

$$y_{t-1} = \varepsilon_{t-1} + 0.5\, y_{t-2} \qquad \text{(ii)}$$

And,

$$y_{t-2} = \varepsilon_{t-2} + 0.5\, y_{t-3} \qquad \text{(iii)}$$

$$y_{t-3} = \varepsilon_{t-3} + 0.5\, y_{t-4} \qquad \text{(iv)}$$

Substitute from (ii) into (i), we find:

$$y_t = \varepsilon_t + 0.5\,[\varepsilon_{t-1} + 0.5\,y_{t-2}\,]$$

$$y_t = \varepsilon_t + 0.5\,\varepsilon_{t-1} + (0.5)^2\,y_{t-2} \qquad \text{(v)}$$

In the same manner, substitute from (iii) into (v), we get:

$$y_t = \varepsilon_t + 0.5\,\varepsilon_{t-1} + (0.5)^2\,\varepsilon_{t-2} + (0.5)^3\,y_{t-3} \qquad \text{(vi)}$$

Also,

$$y_t = \varepsilon_t + 0.5\,\varepsilon_{t-1} + (0.5)^2\,\varepsilon_{t-2} + (0.5)^3\,\varepsilon_{t-3} + (0.5)^4\,y_{t-4}$$

And comparing the last equation with the $\psi - weights$ formula,

we find:

$$\psi_1 = 0.5.\ \psi_2 = (0.5)^2 = 0.25.\ \psi_3 = (0.5)^3 = 0.125$$

Example:

For the model $y_t = \varepsilon_t - 0.3\ \varepsilon_{t-1}$, find the first three $\pi - weights$

and first three $\psi - weights$.

solution:

Comparing with the $\psi - weights$ formula:

$$y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots.$$

We find that:

$$\psi_1 = -0.3 \; ; \psi_2 = \; \psi_3 = 0$$

To find the $\pi - weights$, we rewrite the model as:

$$\varepsilon_t \; = y_t + 0.3 \; \varepsilon_{t-1} \qquad \text{(i)}$$

So that,

$$\varepsilon_{t-1} = y_{t-1} + 0.3 \; \varepsilon_{t-2} \qquad \text{(ii)}$$

$$\varepsilon_{t-2} = y_{t-2} + 0.3 \; \varepsilon_{t-3} \qquad \text{(iii)}$$

$$\varepsilon_{t-3} = y_{t-3} + 0.3 \; \varepsilon_{t-4} \qquad \text{(iv)}$$

Substituting from (ii) into (i), we get:

$$\varepsilon_t = y_t + 0.3 \; [y_{t-1} + 0.3 \; \varepsilon_{t-2} \;]$$

$$\varepsilon_t = y_t + 0.3\, y_{t-1} + (0.3)^2\, \varepsilon_{t-2} \qquad \text{(v)}$$

In the same manner, substituting from (iii) into (v), we find:

$$\varepsilon_t = y_t + 0.3\, y_{t-1} + (0.3)^2\, y_{t-2} \; + (0.3)^3\, \varepsilon_{t-3} \qquad \text{(vi)}$$

And,

$$\varepsilon_t = y_t + 0.3\, y_{t-1} + (0.3)^2\, y_{t-2} \; + (0.3)^3\, y_{t-3} \; + \; (0.3)^4\, \varepsilon_{t-4}$$

Thus,

$$y_t = \varepsilon_t - 0.3\, y_{t-1} - (0.3)^2\, y_{t-2} \; - (0.3)^3\, y_{t-3} \; - \; (0.3)^4\, \varepsilon_{t-4}$$

And comparing the last equation with the π weights formula, we find:

$$\pi_1 = -0.3, \qquad \pi_2 = -(0.3)^2 = -0.09, \quad \pi_3 = -(0.3)^3 = -0.027$$

3.6 Autoregressive Processes

We mentioned earlier that any invertible linear process can be expressed as:

$$y_t = \varepsilon_t + \pi_1 y_{t-1} + \pi_2 y_{t-2} + \cdots.$$

In fact, many of the demographic, economic, environmental, engineering and other applications can be represented in this form using a limited number of constants π as follows :

$$y_t = \varepsilon_t + \pi_1 y_{t-1} + \pi_2 y_{t-2} + \cdots \pi_p y_{t-p} \quad ; \quad t = 0. \pm 1. \pm 2. \ldots$$

We call any process that can be represented in this form as the **Auto-regressive process of order p**, and in the literature it is written in the following format:

$$y_t = \varepsilon_t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} \;\; ; \;\; t = 0. \pm 1. \pm 2. \ldots$$

It is denoted as AR(p), the constants $\phi_1, \phi_2, \ldots \phi_p$ are the main parameters of the model, and they fulfill the invertibility conditions, that is because the number of non-zero π_i weights are limited.

These models might be stationary or not stationary, depending on the values of the parameters $\phi_1. \phi_2. \ldots \phi_p$. The order of the models

in most of the applications does not exceed 2, however in some applications we might need to have larger orders, especially in those where we use the AR models as approximations of other models such as MA models. Thus in this course we will concentrate on autoregressive models of order one and two (AR(1), AR(2)), and just mention some general remarks on the model AR(P).

3.6.1 Auto-regressive model of order one AR (1)

This model takes the form of regressing the value of the series at time t (i.e y_t), on both the value of the series at time $t-1$ (i.e y_{t-1}) and the current value of the disturbance ε_t, the AR(1) model takes the form:

$$y_t = \varepsilon_t + \phi_1 y_{t-1}\,;\;\; t = 0.\pm 1.\pm 2.\ldots$$

Where ε_t is the white noise process, ϕ_1 is a constant value representing the main parameter of the model, and usually we

assume $\{\varepsilon_t\}$ to follow normal distribution with mean zero, and constant variance, that is $\varepsilon_t \sim iid\ N(0.\sigma^2)$. The AR(1) process always fulfill the invertibility condition no matter what the value of ϕ_1, this is because:

$$\pi_1 = \phi_1.\ \pi_i = 0.i > 1$$

i.e. the number of non-zero π_i terms is limited. The AR (1) model can be written in the form:

$$\phi(B)y_t = \varepsilon_t$$

Where $\phi(B) = 1 - \phi_1 B$ is called the autoregressive operator, or the model characteristic function.

3.6.1.1 stationarity condition

We mean by the stationarity conditions, the conditions that the model must satisfy in order to be able to write the model in the

white noise formula. We will denote the past values of the series as:

$$y_{t-1} = \phi_1 \, y_{t-2} + \varepsilon_{t-1}$$

$$y_{t-2} = \phi_1 \, y_{t-3} + \varepsilon_{t-2}$$

$$\vdots$$

$$y_{t-k} = \phi_1 \, y_{t-k-1} + \varepsilon_{t-k}$$

and substituting y_{t-1} in the AR(1) formula, we get:

$$y_t = \varepsilon_t + \phi_1[\phi_1 y_{t-2} + \varepsilon_{t-1}] = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 y_{t-2}$$

substituting y_{t-2} in this form, we get:

$$y_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 y_{t-3}$$

and continue this process k times we get:

$$y_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \cdots + \phi_1^{k-1} \varepsilon_{t-k+1} + \phi_1^k y_{t-k}$$

or,

$$y_t = \sum_{j=0}^{k-1} \phi_1^j \varepsilon_{t-j} + \phi_1^k y_{t-k}$$

Form the previous formula, we notice that if $|\phi_1| < 1$, and $k \longrightarrow \infty$, then the term $\phi_1^k y_{t-k}$ will tend to zero, thus it will be possible to write the AR(1) model in the white noise formula:

$$y_t = \sum_{j=0}^{k-1} \phi_1^j \, \varepsilon_{t-j}$$

And comparing this formula with the white noise formula, we notice that the coefficients in this formula takes the form $\psi_j = \phi_1^j$, with the condition that $|\phi_1| < 1$. Notice that if

$|\phi_1| > 1$, then it is not possible to write the AR(1) model in the white noise formula, so we conclude that the stationarity condition for the AR(1) model is that $|\phi_1| < 1$.

- **Equivalent stationarity condition of AR (1) model**

Stationarity condition for AR(1) model can be checked in a more general way by inspecting the characteristic equation of the model $\phi(B) = 1 - \phi_1 B$, and if $|\phi_1| < 1$ then the root of the characteristic

equation $\phi(B) = 0$ must lie outside the unit circle, i.e. root must satisfy $|B| > 1$.

3.6.1.3 Autocorrelation function of AR (1) model

We will assume that the model satisfies the stationarity condition, $|\phi_1| < 1$, the model has the form:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t$$

Where $\varepsilon_t \sim iid\ N(0.\sigma^2)$, the white noise process.

taking expectations for both sides:

$$E(y_t) = \phi_1 E(y_{t-1}) + 0$$

since the process is stationary, then $E(y_t) = E(y_{t-1})$, thus:

$$E(y_t)(1 - \phi_1) = 0 \Rightarrow E(y_t) = 0$$

also, the variance of y_t is:

$$var(y_t) = \phi_1^2 var(y_{t-1}) + var(\varepsilon_t)$$

and since the process is stationary, then $var(y_t) = var(y_{t-1}) = \gamma(0)$, so:

$$\gamma(0)(1 - \phi_1^2) = \sigma^2$$

or,

$$\gamma(0) = \frac{\sigma^2}{(1 - \phi_1^2)} . \quad |\phi| < 1$$

the auto-covariance at lag one is:

$$\gamma(1) = cov\,(y_t . y_{t-1}) = cov(\phi_1 y_{t-1} + \varepsilon_t \,.\, y_{t-1})$$

$$= \phi_1 cov(y_{t-1} \,.\, y_{t-1}) + cov(\,\varepsilon_t . y_{t-1})$$

$$= \phi_1 \gamma(0) + 0$$

and the auto-covariance at lag 2 is:

$$\gamma(2) = cov\,(y_t . y_{t-2}) = cov(\phi_1 y_{t-1} + \varepsilon_t \,.\, y_{t-2})$$

$$= \phi_1 cov(y_{t-1} \,.\, y_{t-2}) + 0 = \phi_1 \gamma(1)$$

in general, at lag k, it has the form:

$$\gamma(k) = \phi_1 \gamma(k-1). \quad k = 1.2. \ldots$$

Dividing both sides by $\gamma(0)$, we get the auto-correlation function of the AR(1) model:

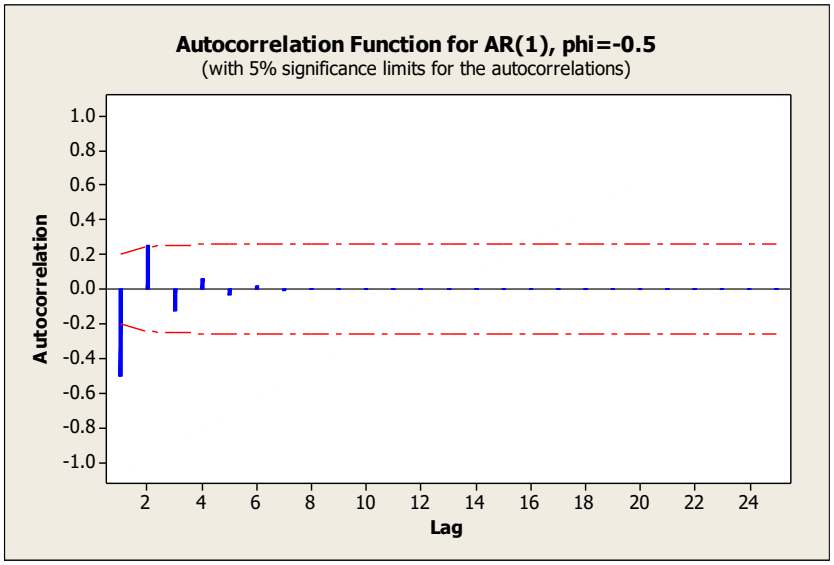
$$\rho(k) = \phi_1 \rho(k-1). \quad k = 1.2. \ldots$$

And by continually substituting we get:

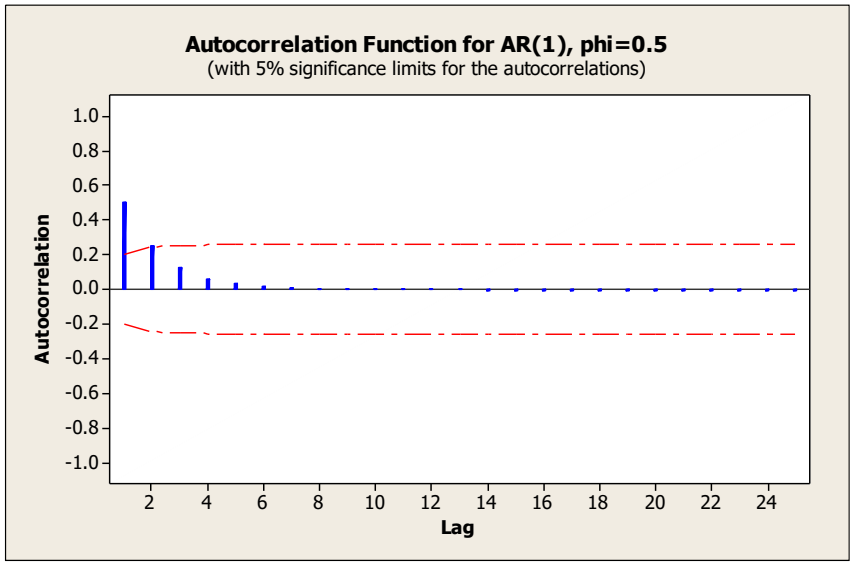
$$\rho(k) = \phi_1^2 \rho(k-2) = \phi_1^3 \rho(k-3) = \cdots = \phi_1^k \rho(0) = \phi_1^k$$

Which indicate that this model remembers everything happened in the past, or we say that it has an infinite memory, however we notice that this memory decrease in an exponential manner as the time lag between current observation y_t and observation y_{t-k} increases.

To show the behavior of the auto-correlation function for the AR(1) model, we plot this function for some values of ϕ_1.

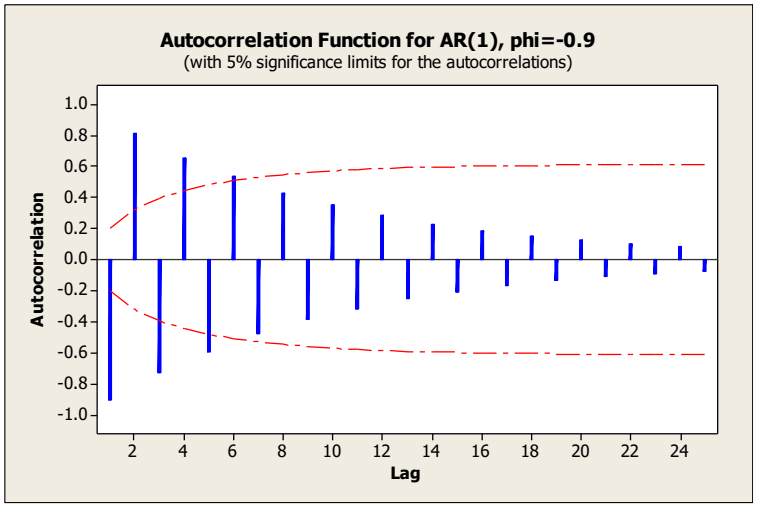


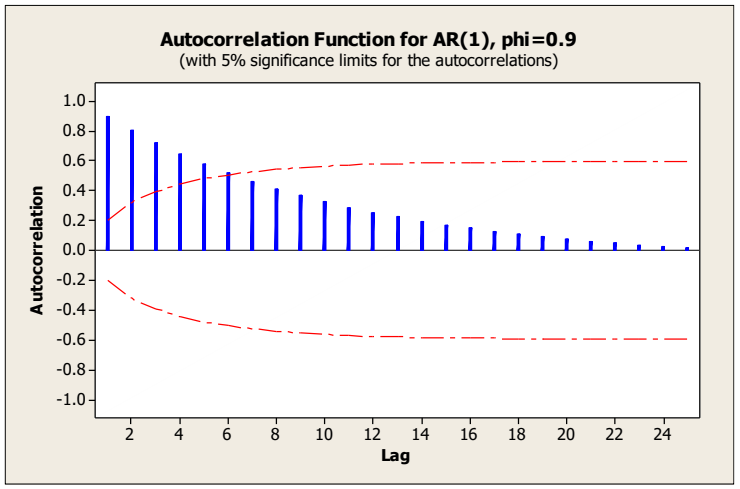
(a) ACF for AR(1) model with $\phi_1 = -0.5$



(b) ACF for AR(1) model with $\phi_1 = 0.5$

We notice from figure (a) that the auto-correlation takes the form of a declining sine-wave form because the parameter value is negative, and from figure (b) the auto-correlation takes the form of a declining exponential form because the parameter value is positive. Also, we note that this decline will be slow at the non-stationarity boundaries $\phi_1 = \pm 1$, for example at $\phi_1 = \pm 0.9$, the ACF will take the form:





(a) ACF for AR(1) model with $\phi_1 = -0.9$

(b) ACF for AR(1) model with $\phi_1 = 0.9$

Note: The general form of the AR(1) model when the model mean is not equal to zero , i.e. when $E(y_t) = \mu$ is:

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \varepsilon_t$$

or,

$$y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t$$

where $\delta = \mu(1 - \phi_1)$.

The AR(1) model could be interpreted for example, if we assume y_t represent the number of population of a certain country at a certain year, then this number is a fraction ϕ_1 (fraction of those who are

still alive) multiplied by the population number in the previous year y_{t-1}, added to them a random component ε_t (representing the new citizens of the country). Another example, y_t might represent number of unemployed people at a certain month, January for example, then this number is a fraction ϕ_1 (fraction of those who are still unemployed) multiplied by the number of unemployed in the previous month y_{t-1}, added to them a random component ε_t (representing the new unemployed looking for job).

3.6.1.4 Partial autocorrelation function for AR (1) model

To find the partial autocorrelation function for the AR(1) model,

$$\phi_{00} = 1 .$$

$$\phi_{11} = \rho_1 = \phi^1 = \phi.$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \phi \\ \phi & \phi^2 \end{vmatrix}}{\begin{vmatrix} 1 & \phi \\ \phi & 1 \end{vmatrix}} = \frac{\phi^2 - \phi^2}{1 - \phi^2} = 0$$

Thus for any time lag k , we can find the partial autocorrelation function (PACF) as:

$$\phi_{\mathrm{kk}} = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_2 \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{vmatrix}}$$

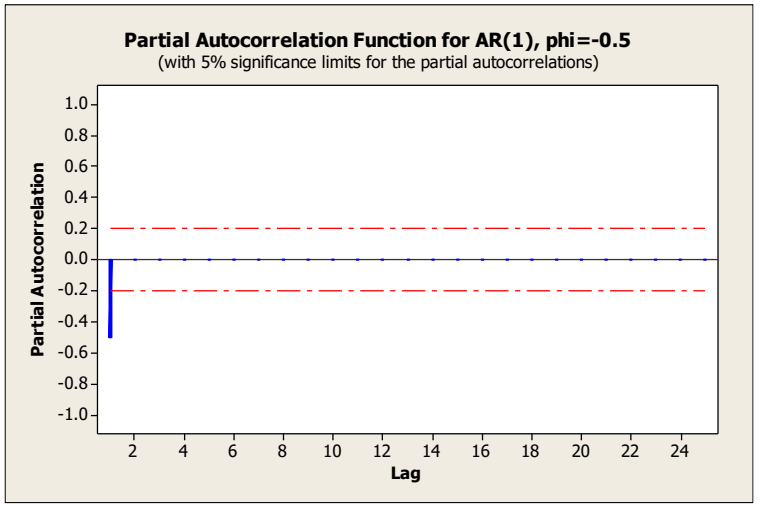
$$= \frac{\begin{vmatrix} 1 & \phi & \dots & \phi \\ \phi & 1 & \dots & \phi^2 \\ \vdots & \vdots & \dots & \vdots \\ \phi^{k-1} & \phi^{k-2} & \dots & \phi^k \end{vmatrix}}{\begin{vmatrix} 1 & \phi & \dots & \phi^k \\ \phi & 1 & \dots & \phi^{k-1} \\ \vdots & \vdots & \dots & \vdots \\ \phi^{k-1} & \phi^{k-2} & \dots & 1 \end{vmatrix}} = \frac{0}{\begin{vmatrix} 1 & \phi & \dots & \phi^k \\ \phi & 1 & \dots & \phi^{k-1} \\ \vdots & \vdots & \dots & \vdots \\ \phi^{k-1} & \phi^{k-2} & \dots & 1 \end{vmatrix}} = 0$$

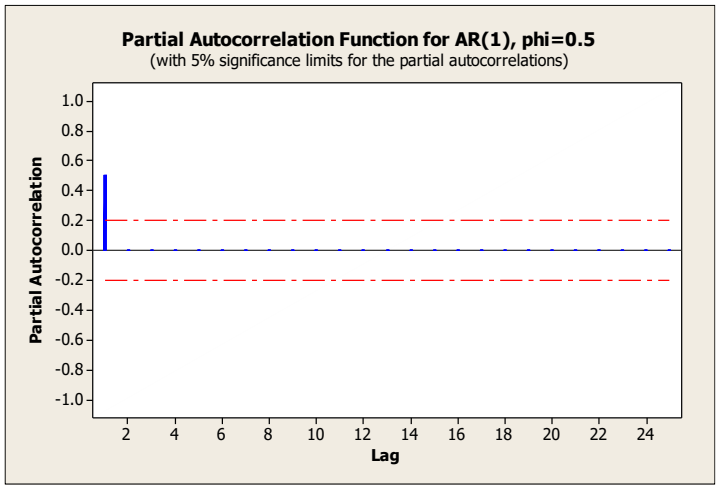
The determinant of the numerator equals zero because the columns are not independent, where we notice that the last

column equals ϕ multiplied by the first column. So, the PACF for the AR(1) model have the form:

$$\phi_{kk} = \begin{cases} 1\,. & k = 0 \\ \phi\,. & k = 1 \\ 0\,. & k \geq 2 \end{cases}$$

The behavior of the PACF for the AR(1) model is as follow:





PACF for AR(1) model when

$\phi < 0$

PACF for AR(1) model when

$\phi > 0$

3.6.2 AR (2) Model

This model takes the form:

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

Where ε_t is the white noise process, i.e. $\varepsilon_t \sim WN(0.\sigma^2)$, and ϕ_1 , ϕ_2 are constant values representing the model parameters.

Now applying the backshift operator, we can rewrite the model in form:

$$(1 - \phi_1 B - \phi_2 B^2) y_t = \delta + \varepsilon_t$$

$$y_t = (1 - \phi_1 B - \phi_2 B^2)^{-1} \delta + (1 - \phi_1 B - \phi_2 B^2)^{-1} \varepsilon_t \qquad (1)$$

Returning to the general linear process:

$$y_t = \mu_Y + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

$$y_t = \mu_Y + (1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \cdots)\varepsilon_t$$

$$y_t = \mu_Y + \psi(B)\varepsilon_t \qquad (2)$$

So, from (1) and (2) the $\psi(B)$ function for AR(2) model is:

$$\psi(\mathrm{B}) = (1 - \phi_1 B - \phi_2 B^2)^{-1}$$

Multiplying both sides by $(1 - \phi_1 B - \phi_2 B^2)$, we get:

$$\psi(\mathrm{B})(1 - \phi_1 B - \phi_2 B^2) = 1$$

That is,

$$(1 + \psi_1 B + \psi_2 B^2 + \cdots)\,(1 - \phi_1 B - \phi_2 B^2) = 1$$

And for this equality to hold, the B^j coefficients for $j \geq 0$, must be equal, as follow:

$$B^1: \psi_1 - \phi_1 = 0 \qquad \Longrightarrow \psi_1 = \phi_1$$

$$B^2: \psi_2 - \phi_1\psi_1 - \phi_2 = 0 \qquad \Longrightarrow \psi_2 = \phi_1\psi_1 + \phi_2 = {\phi_1}^2 + \phi_2$$

$$B^3: \psi_3 - \phi_1\psi_2 - \phi_2\psi_1 = 0 \Longrightarrow \psi_3 = \phi_1\psi_2 + \phi_2\psi_1$$

$$B^4: \psi_4 - \phi_1\psi_3 - \phi_2\psi_2 = 0 \Longrightarrow \psi_4 = \phi_1\psi_3 + \phi_2\psi_2$$

Thus, in general the general form of the ψ_j weights for the AR(2) model has the form:

$$\boxed{\psi_j = \phi_1 \psi_{j-1} + \phi_2 \psi_{j-2} \ . j \geq 2}$$

And for the AR(2) to be stationary, the ψ_j weights must converge, thus we must put some conditions on ϕ_1 and ϕ_2 to satisfy this:

As we remember, the stationarity condition for AR(1) was that $|\phi| < 1$ or equivalently, the solution of the characteristic equation

$(1 - \phi B) = 0$, which is $B = \left|\frac{1}{\phi}\right|$ should be greater than one.

However, for AR(2), we have a quadratic equation:

$$(1 - \phi_1 B - \phi_2 B^2) = 0$$

So we must look at two solutions G_1^{-1} and G_2^{-1}, and are usually called the solutions of the characteristic equation, now:

$$(1 - \phi_1 B - \phi_2 B^2) = (1 - G_1 B)(1 - G_2 B) = 0$$

G_1^{-1} and G_2^{-1} can be real or complex numbers. So, the stationarity conditions for the AR(2) process is that $\left|G_1^{-1}\right| > 1$ and $\left|G_2^{-1}\right| > 1$.

Note: note that $|x|$ means the absolute value for x if it is a real number, but it means $\sqrt{a^2 + b^2}$ if it is a complex number, i.e. one that can be written in the form $x = a + ib$.

Examples:

Suppose that we have an AR(2) model with parameters $\phi_1 = 0.8$ and $\phi_2 = -0.15$, so the characteristic equation has the form:

$$(1 - 0.8B + 0.15B^2) = (1 - 0.5B)(1 - 0.3B) = 0$$

So, the solution is $G_1^{-1} = \frac{1}{0.5} = 2$ and $G_2^{-1} = \frac{1}{0.3} = 3.33$, and both are greater than one in absolute value, so this model is stationary.

Suppose that we have an AR(2) model with parameters $\phi_1 = 1.5$ and $\phi_2 = -0.5$, so the characteristic equation has the form:

$$(1 - 1.5B + 0.5B^2) = (1 - B)(1 - 0.5B) = 0$$

So, one root is $G_1^{-1} = 1$, which is not greater than one, so this model is not stationary.

Suppose that we have an AR(2) model with parameters $\phi_1 = 1$ and $\phi_2 = -0.5$, so the characteristic equation has the form:

$$(1 - B + 0.5B^2) = 0$$

Which means that $a = 1$ and $b = 1$, in this case the solutions are:

$$\left|G_1^{-1}\right| = \left|G_2^{-1}\right| = \sqrt{1^2 + 1^2} = \sqrt{2}$$

and both are greater than in one, so this model is stationary.

An equivalent method of checking stationarity of AR(2) model is by looking directly to the parameters ϕ_1 and ϕ_2:

We say that the AR(2) process is stationary if the following conditions are satisfied:

$$-1 < \phi_2 < 1$$

$$\phi_1 + \phi_2 < 1$$

$$\phi_2 - \phi_1 < 1$$

And if any of them is not satisfied, then the process is not stationary.

3.6.2.1 Autocorrelation function of AR (2) model

For simplicity, we will assume that $\mu = 0$, so the general form of the model is:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

Multiply both sides by Y_{t-k} and taking expectations:

$$\gamma_k = E[y_t y_{t-k}] = \phi_1 E[y_{t-1} y_{t-k}] + \phi_2 E[y_{t-2} y_{t-k}] + E[\varepsilon_t y_{t-k}]$$

And since Y_{t-k} depends only on $\varepsilon_{t-k}. \varepsilon_{t-k-1}. \ldots$, then we have:

$$E[\varepsilon_t y_{t-k}] = \begin{cases} \sigma_\varepsilon^2 & , k = 0 \\ 0 & , k = 1,2,3 \end{cases}$$

So that,

$$\gamma_0 = \phi_1\gamma_{-1} + \phi_2\gamma_{-2} + \sigma_\varepsilon^2$$

$$= \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma_\varepsilon^2,$$

and,

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2}.\ k > 0$$

From which we can get the auto-correlation function ρ_k:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \quad .k = 1.2. \ldots$$

Which is the Yule-Walker equations for this model.

For example, for $k = 1$:

$$\rho_1 = \phi_1 \rho_0 + \phi_2 \rho_1$$

$$\rho_1(1 - \phi_2) = \phi_1 \implies \rho_1 = \frac{\phi_1}{(1 - \phi_2)}$$

for $k = 2$:

$$\rho_2 = \phi_1\rho_1 + \phi_2\rho_0 \implies \rho_2 = \frac{{\phi_1}^2}{(1-\phi_2)} + \phi_2$$

i.e.

$$\rho_2 = \frac{{\phi_1}^2 + \phi_2 - {\phi_2}^2}{(1-\phi_2)}$$

In the same manner, we gat get the form of ρ_k for any value k.

3.6.2.2 Partial autocorrelation function of AR (2) model

$$\phi_{00} = 1 .$$

$$\phi_{11} = \rho_1 = \frac{\phi_1}{(1-\phi_2)} .$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - {\rho_1}^2}{1 - {\rho_1}^2}$$

Where, $\rho_2 = \frac{{\phi_1}^2+\phi_2-{\phi_2}^2}{(1-\phi_2)}$.

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1=\phi_1\rho_0+\phi_2\rho_1 \\ \rho_1 & 1 & \rho_2=\phi_1\rho_1+\phi_2\rho_0 \\ \rho_2 & \rho_1 & \rho_3=\phi_1\rho_2+\phi_2\rho_1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = \frac{0}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = 0$$

Because the last column is a linear combination of the first column (they are not independent). So, it is possible to prove that $\phi_{kk} = 0$ for $k \geq 3$.

Thus the PACF for AR(2) can be written as follow:

$$\phi_{kk} = \begin{cases} 1. & k = 0 \\ \rho_1. & k = 1 \\ \dfrac{\rho_2 - {\rho_1}^2}{1 - {\rho_1}^2}. & k = 2 \\ 0. & k \geq 3 \end{cases}$$

Thus, we can summarize the properties of the AR(2) model as following:

- If the stationarity conditions are satisfied, i.e.:

$-1 < \phi_2 < 1$ and $\phi_2 - \phi_1 < 1$ and $\phi_1 + \phi_2 < 1$

or equivalently, the roots of the characteristic equation:

$(1 - \phi_1 B - \phi_2 B^2) = (1 - G_1 B)(1 - G_2 B) = 0$ satisfy $\left|G_1^{-1}\right| > 1$ and $\left|G_2^{-1}\right| > 1$, then:

$$E(y_t) = \frac{\delta}{(1 - \phi_1 - \phi_2)}$$

which is a constant value for all t.

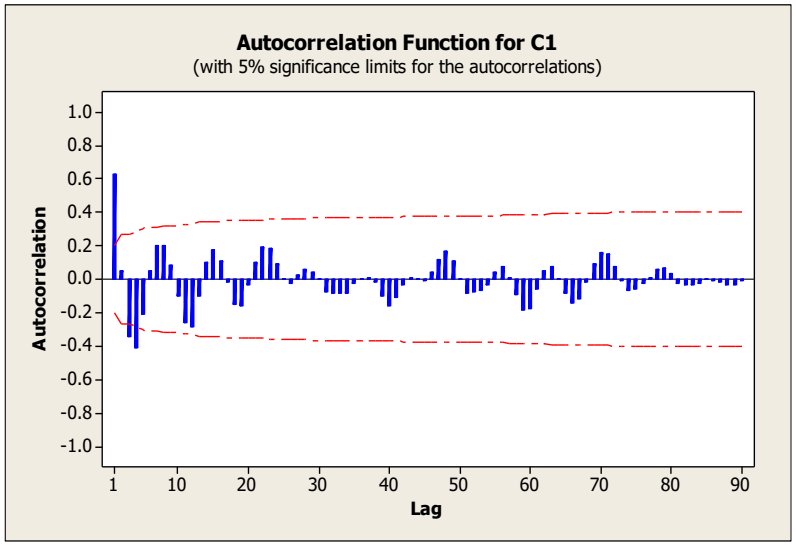
- The ACF depends on the values of the roots of the characteristic equation:

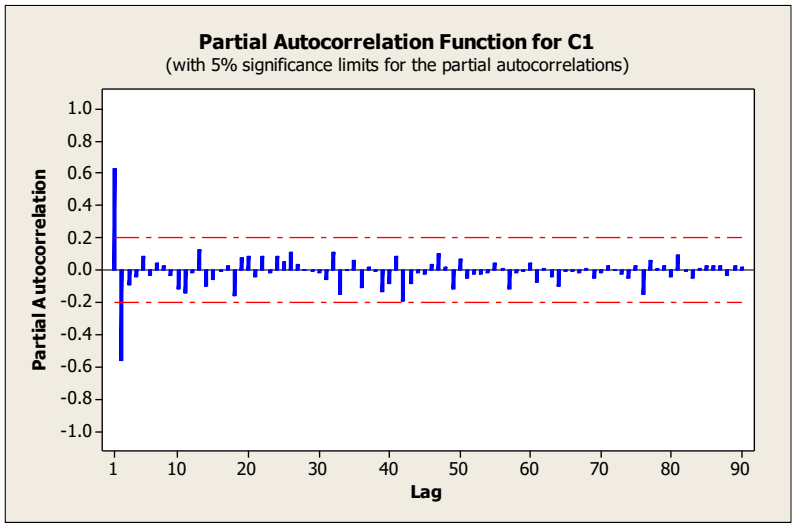
- If the roots are real, then the ACF decline in an exponential fashion.
- If the roots are complex, then the ACF decline in a sin-wave fashion.

- The PACF has only two values not equal to zero (ϕ_{11} and ϕ_{22}), whereas the rest of the values equal zero.

To show the behavior of the PACF for the AR(2) model we take the following examples.

The following figure shows the ACF and PACF for $\phi_1 = 1, \phi_2 = -0.5$:





ACF for AR(2) model when

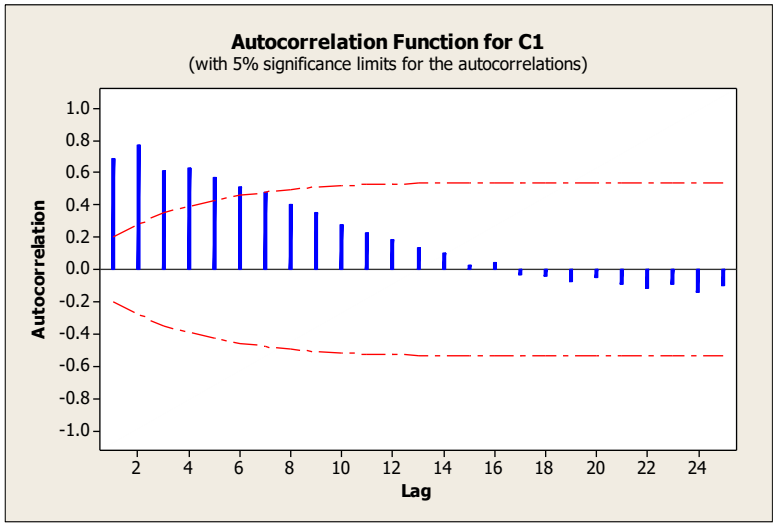
$\phi_1 = 1. \phi_2 = -0.5$

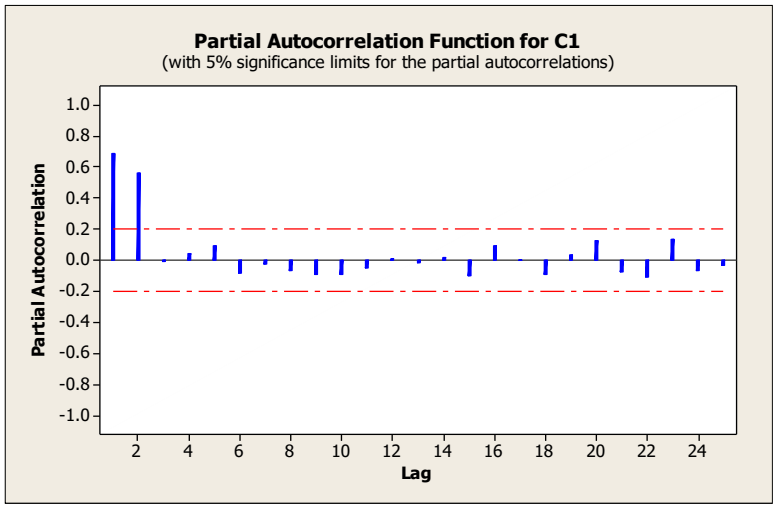
PACF for AR(2) model when

$\phi_1 = 1. \phi_2 = -0.5$

We note from figure (a) that the ACF takes the form of a decaying sine-wave, and from figure (b) that the PACF has only two coefficients differ from zero, and the function cut-off after two time lags.

The following figure shows the ACF and PACF for $\phi_1 = 0.4. \phi_2 = 0.5$:





ACF for AR(2) model when

$\phi_1 = 0.4. \phi_2 = 0.5$

PACF for AR(2) model when

$\phi_1 = 0.4 \,. \phi_2 = 0.5$

The ACF decline in an exponential format, and again the PACF cuts-off after two time lags.

3.6.3 Autoregressive Model of order p

This has the following form:

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

Where $\varepsilon_t \sim WN(0.\sigma^2)$.

using the backshift operator:

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi^p B^p)\, y_t = \delta + \varepsilon_t$$

or:

$$\phi(B)\, y_t = \delta + \varepsilon_t$$

These models are always invertible regardless of the values of parameters ϕ_i, this is because the number of non-zero π_i

weights are limited. AR(p) models might be stationary or not depending the values of the coefficients ϕ_i, however, it can be shown that if the roots of the characteristic function $\phi(B) = 0$ fall outside the unit circle, then the model is stationary.

The autocorrelation function of the AR (p) model can be shown to satisfy the following difference equation:

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p} \quad .k \geq 1$$

we will not derive mathematical solution of this function, however, we will just mention the forms this function can take (it is very similar to the forms of the AR(1), and AR(2) cases):

The ACF extend infinitely and consist of a mixture of decaying exponential or sine-wave functions. So, always the ACF function is a good indicator whether a series in practical applications can be modeled by autoregressive models. However, it is not enough in

determining the order p of the autoregression, so we have to examine the PACF, which cuts-off after the order p.

3.7 Moving Average Processes

We mentioned earlier that any stationary linear process can be written in the form:

$$y_t = \varepsilon_t + \sum_{j=1}^{\infty} \psi_j \, \varepsilon_{t-j} \; ; \; where \sum_j {\psi_j}^2 < \infty$$

In fact many phenomena in economics or social sciences can be represented (may be after first or second difference) in the same manner, however with limited number of constants ψ_j, as follow:

$$y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots + \psi_q \varepsilon_{t-q}$$

The processes that can be represented in this form is called the Moving Average of order q, or MA(q) in short. In literature, it is written in a special format, so that they can be distinguished from other operations:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

The constants θ_i are the main parameters of the model.

The MA models are always stationary no matter what the parameters values are, because the number of non-zero ψ_1 values in the linear process representation are limited:

$$\psi_1 = -\theta_1.\ \psi_2 = -\theta_2. \ldots . \psi_q = -\theta_q\ ;\ \ \psi_j = 0.\ j > q$$

Note: Sometimes it may be necessary to express these models using the past values of the series y_{t-1} , y_{t-2}, … , this means that we use the invertibility formula, in which case we must put some conditions on the parameters θ_i, these conditions are called the invertibility conditions, we will see this conditions when discussing the models MA(1) and MA(2).

In most applications that arise in economics and management, engineering, environmental studies, the value of q is usually less than or equal 2, so we will confine ourselves to discussing these two models, and just mention some properties of the general MA(q) model.

3.7.1 Moving Average of first order MA (1)

We say that the process $\{y_t\}$ follow a moving average model of order one if it can be represented as:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} \; ; t = 0. \pm 1. \pm 2. \ldots$$

Where θ_1 represent the main parameter of the model, and $\{\varepsilon_t\}$ is the white nose process $\varepsilon_t \sim iid \;\; N(0.\sigma_\varepsilon^2)$.

The MA(1) model is considered as one of the important time series analysis models that is used in modeling inventory, quality control, temperature, pollution percentages, and general economic indicators after being affected by sudden disturbances either from within the system such as worker strikes, or from outside the system such as wars or disasters, etc.

As mentioned earlier that the MA(1) model is always stationary, no matter what the value of the parameter θ_1 is, because:

$$\psi_1 = -\theta_1 \; ; \;\; \psi_j = 0. \;\; j > 1$$

The model can be written in short as:

$$y_t = \theta(B)\varepsilon_t$$

Where $\theta(B) = 1 - \theta_1 B$ is a polynomial.

The linear filter $\theta(B)$ is called the moving average operator, it link the process $\{y_t\}$ as an output with the process $\{\varepsilon_t\}$ as input.

3.7.1.1 The autocorrelation function of MA (1)

The model is:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Taking the expectation of both sides:

$$E(y_t) = \varepsilon_t - \theta_1 \varepsilon_{t-1} = 0$$

and taking the variance of both sides:

$$var(y_t) = \gamma(0) = var(\varepsilon_t - \theta_1 \varepsilon_{t-1})$$

$$= var(\varepsilon_t) + \theta_1{}^2 \, var(\varepsilon_{t-1})$$

$$= \sigma_\varepsilon^2 (1 + \theta_1{}^2)$$

and the auto-covariance at lag one is:

$$\gamma(1) = cov\,(y_t . y_{t-1}) = cov(\varepsilon_t - \theta_1 \varepsilon_{t-1} \,.\, \varepsilon_{t-1} - \theta_1 \varepsilon_{t-2})$$

$$= -\theta_1 cov(\varepsilon_{t-1} . \varepsilon_{t-1}) = -\theta_1 \sigma_\varepsilon^2$$

and at lag two:

$$\gamma(2) = cov\,(y_t . y_{t-2}) = cov(\varepsilon_t - \theta_1 \varepsilon_{t-1} \,.\, \varepsilon_{t-2} - \theta_1 \varepsilon_{t-3}) = 0$$

Similarly, one can show that: $\gamma(3) = \gamma(4) = \cdots = 0$

So the auto-covariance function for the MA(1) model can be written as:

$$\gamma(k) = \begin{cases} \sigma_\varepsilon^2\left(1 + {\theta_1}^2\right). & k = 0 \\ -\theta_1\sigma_\varepsilon^2\,. & k = 1 \\ 0\quad . & k = 2.3. \ldots \end{cases}$$

Note that the expectation, variance, and auto-covariance functions of this model do not depend on time t, (which is expected to be, since moving average processes are always stationary). Now dividing by variance $\gamma(0)$, we get the autocorrelation function for the MA(1) model:

$$\rho(k) = \begin{cases} 1\,. & k = 0 \\ \dfrac{-\theta_1}{\left(1 + {\theta_1}^2\right)}\,. & k = 1 \\ 0 & k = 2.3. \ldots \end{cases}$$

Which means that autocorrelation function of MA(1) processes cuts-off after the first time lag, which means that observations one time lag apart are correlated, while at larger lags they are not correlated. Also, note that if the sign of θ_1 is negative, then $\rho(1)$ is positive, which means that large values of the series y_t tend to be followed by large values, and small values are followed by small values, in this case the process $\{y_t\}$ is more smooth than the white

noise process , and this smoothness increases as θ_1 approaches -1, and the reverse situation occurs when θ_1 is positive.

3.7.1.2 The Partial autocorrelation function of MA (1)

From the definition of partial auto-correlation we have,

$$\phi_{00} = 1 .$$

$$\phi_{11} = \rho_1 = -\left(\frac{\theta_1}{1+\theta_1^2}\right).$$

Using the determinants to find the partial autocorrelation functions, we find:

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{0 - {\rho_1}^2}{1 - {\rho_1}^2} = \frac{-\theta_1^2}{1 + \theta_1^2 + \theta_1^4}$$
$$= \frac{-\theta_1^2(1 - \theta_1^2)}{1 - \theta_1^6}$$

(where me multiplied numerator and denominator by $(1 - \theta_1^2)$).

For k=3 we get:

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & 0 \\ 0 & \rho_1 & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = \frac{{\rho_1}^3}{1 - 2{\rho_1}^3} = \frac{-\theta_1^3(1 - \theta_1^2)}{1 - \theta_1^8}$$

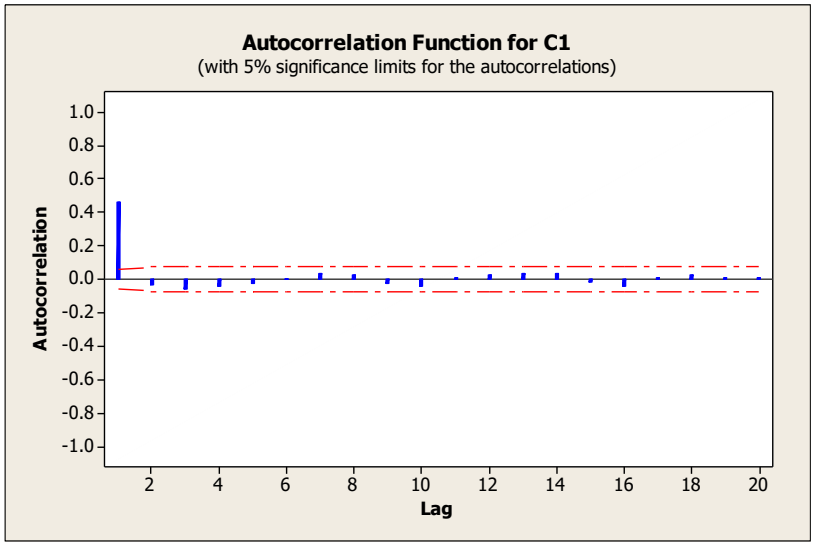
In general we can prove that $\phi_{kk} = \frac{-\theta_1^k(1-\theta_1^2)}{1-\theta_1^{2(k+1)}}$ for all $k > 0$. Thus the PACF for the MA(1) model takes the same form as the ACF for the AR models:

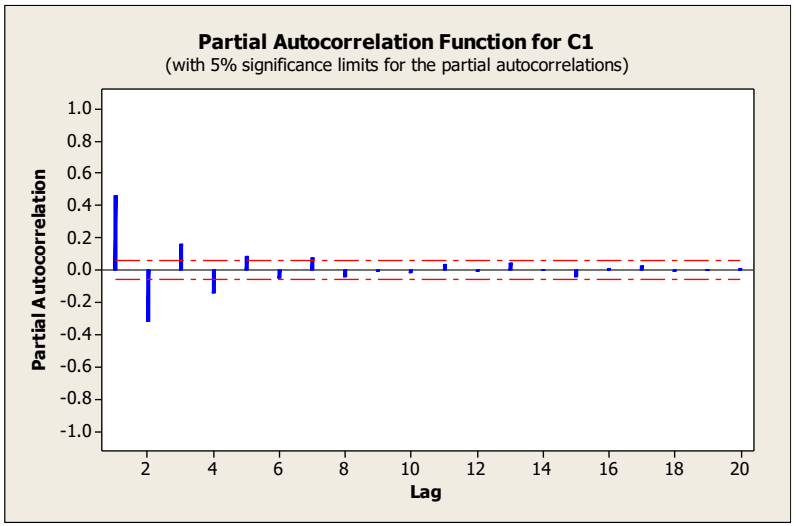
i) If $0 < \theta < 1$; then the PACF follow a damped exponential function.

ii) If $-1 < \theta < 0$; then the PACF follow a damped sine-wave function.

To show the behavior of the ACF and PACF for the MA(1) model we take the following examples.

1- The following figure shows the ACF and PACF for $\theta_1 = -0.7$:



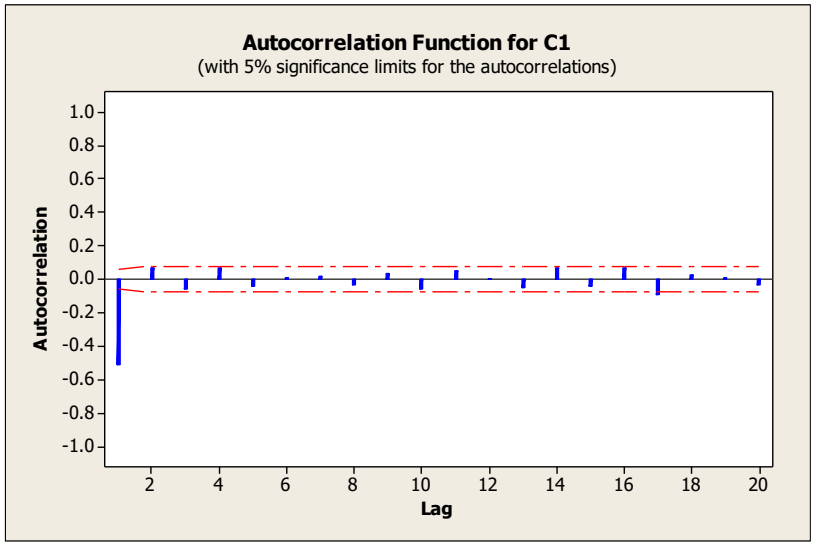


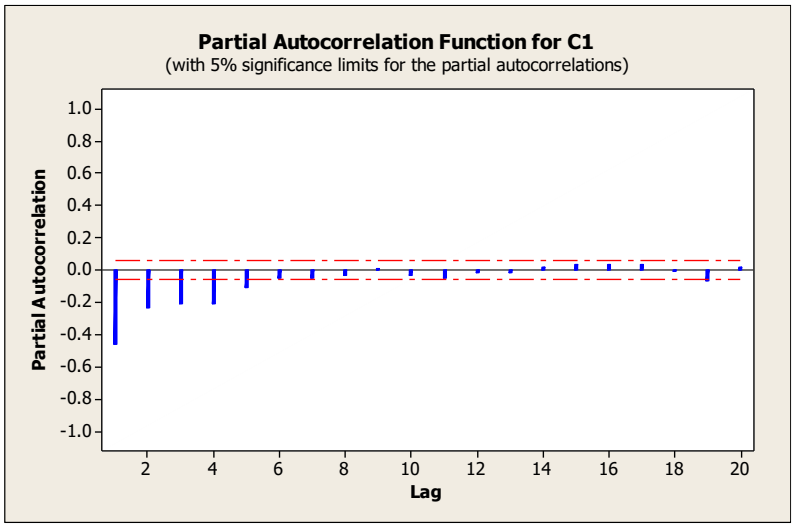
a) ACF for MA (1) model when $\theta_1 = -0.7$

b) PACF for MA (1) model when $\theta_1 = -0.7$

We note from figure (a) that the ACF cuts-off after lag 1, and from figure (b) that the PACF takes the form of a decaying sine-wave. Also, note that $\hat{\rho}_1 = 0.4698$ and its sign is opposite to the sign of θ_1.

2- The following figure shows the ACF and PACF for $\theta_1 = 0.7$:





a) ACF for MA(1) model when $\theta_1 = 0.7$

b) PACF for MA(1) model when $\theta_1 = 0.7$

We note from figure (a) that the ACF cuts-off after lag 1, and from figure (b) that the PACF takes the form of a decaying exponential

function. Also, note that $\hat{\rho}_1 = -0.4698$ and its sign is opposite to the sign of θ_1.

3.7.1.3 Invertibility

We have already mentioned the invertibility, and explained the importance of writing the model in terms of the past values of the series $y_{t-1}. y_{t-2}. \ldots$, also, we have mentioned that to be able to achieve this goal we have to put some conditions on the weights π_i, the definition of MA(1) model is:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

rewriting it as:

$$\varepsilon_t = y_t + \theta_1 \varepsilon_{t-1}$$

From which we can get:

$$\varepsilon_{t-1} = y_{t-1} + \theta_1 \varepsilon_{t-2}$$

$$\varepsilon_{t-2} = y_{t-2} + \theta_1 \varepsilon_{t-3}$$

$$\vdots$$

$$\varepsilon_{t-k} = y_{t-k} + \theta_1 \varepsilon_{t-k-1}$$

And by continue substituting in $\varepsilon_t = y_t + \theta_1 \varepsilon_{t-1}$, we get:

$$\varepsilon_t = y_t + \theta_1 y_{t-1} + \theta_1^2 y_{t-2} + \theta_1^3 y_{t-3} + \cdots + \theta_1^k y_{t-k} + \theta_1^{k+1} \varepsilon_{t-k-1}$$

If we continue substitute for large number of times, i.e. letting $k \to \infty$, then last term $(\theta_1^{k+1} \varepsilon_{t-k-1})$ will not diminish to zero unless we put the condition that $|\theta| < 1$, whereas, if $|\theta| > 1$ then it will not

diminish to zero, and as a consequence the observations in the MA(1) model will be affected by all observations in the history of the series.

3.7. 1.4 importance of Invertibility

Invertibility is a special characteristic concerned with the models and is completely independent in terms of concept and importance from stationarity. Some points about its importance are:

1. Invertibility ensures that the value y_t is affected after a specific period of time by the nearby observations more than

being affected by observations very distant apart, in fact we see this effect decreases in an exponential manner.

2. Invertibility ensures the existence of a single model corresponding to a specific auto-correlation function. We have found for MA (1) model that:

$$\rho_1 = \frac{-\theta_1}{\left(1 + {\theta_1}^2\right)}$$

Cross Multiplication and rearranging terms, we get:

$${\theta_1}^2 \rho_1 + \theta_1 + \rho_1 = 0$$

or,

$$\theta_1{}^2 + \frac{\theta_1}{\rho_1} + 1 = 0$$

it is a quadratic function in θ_1 , which has two roots their multiplication equal 1, and thus if $\theta_1{}^*$is one root, then the second will be $\frac{1}{\theta_1{}^*}$, this means that there are <u>two</u> MA(1) models having <u>two different</u> values for θ_1 but <u>have the same</u> auto-correlation function!

3. Invertibility makes it possible sometimes to use MA(q) with a small order as an alternative for a model that uses a large number of previous observation:

$$y_t = \varepsilon_t + \theta_1 y_{t-1} + \theta_1^2 y_{t-2} + \theta_1^3 y_{t-3} + \cdots$$

Example:

If $\{y_t\}$ is a MA(1) process with $\theta_1 = 0.5,$ what is the auto-correlation function for this process, then show that there exist another value for θ_1 satisfy this auto-correlation function. Which value satisfy the invertibility condition?

Solution:

$$\rho_1 = -\left(\frac{\theta_1}{1+\theta_1^2}\right) \; ; \; \rho_k = 0 \, . k > 1$$

So if $\theta_1 = 0.5$, then:

$$\rho_1 = -\left(\frac{0.5}{1+(0.5)^2}\right) = -0.4$$

Now, if we used the other root that satisfy this equation, which is $\frac{1}{\theta_1} = \frac{1}{0.5} = 2$, then:

$$\rho_1 = -\left(\frac{\frac{1}{0.5}}{1+\left(\frac{1}{0.5}\right)^2}\right) = -0.4$$

This means that ${\theta_1}^* = \frac{1}{0.5}$ gives the same value for ρ_1 as the value $\theta_1 = 0.5$, so we have two MA(1) models having the same auto-correlation function:

$$\rho_k = \begin{cases} -0.4 & . \quad k = 1 \\ 0 & . \quad k = 2.3. \ldots \end{cases}$$

The first model $MA(1)$ with parameter 0.5, the other with parameter 2, of course the first one satisfies the invertibility condition ($|\theta| < 1$).

3.7.2 Moving Average of second order MA (2)

We say that the process $\{y_t\}$ follow a moving average model of order two if it can be represented as:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \; ; t = 0. \pm 1. \pm 2. \ldots$$

Where θ_1, θ_2 represent the main parameters of the model, and $\{\varepsilon_t\}$ is the white nose process $\varepsilon_t \sim iid\ \ N(0.\sigma_\varepsilon^2)$.

The MA(2) model is similar to the MA(1) model, but it has more ability in modeling a more complicated situations, as it is used in modeling important economic indicators after being affected by sudden disturbances when effects of such disturbances extend to two time lags.

Also, the MA(2) model is always stationary, no matter what the value of the parameter θ_1, θ_2 are, since:

$$\psi_1 = -\theta_1 \,; \psi_2 = -\theta_2 \ ; \ \psi_j = 0 \,. \ \ j > 2$$

The model can be written in short as:

$$y_t = \theta(B)\varepsilon_t$$

Where $\theta(B) = 1 - \theta_1 B - \theta_2 B^2$ is a polynomial in the operator B.

3.7.2.1 The autocorrelation function of MA (2)

The model is:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

Taking the expectation of both sides:

$$E(y_t) = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} = 0$$

and taking the variance of both sides:

$$var(y_t) = \gamma(0) = var(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2})$$

$$= var(\varepsilon_t) + {\theta_1}^2 var(\varepsilon_{t-1}) + {\theta_2}^2 var(\varepsilon_{t-2})$$

$$= \sigma_\varepsilon^2\left(1 + {\theta_1}^2 + {\theta_2}^2\right)$$

and the auto-covariance at lag one is:

$$\gamma(1) = cov\,(y_t . y_{t-1})$$

$$= cov(\varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} . \varepsilon_{t-1} - \theta_1\varepsilon_{t-2} - \theta_2\varepsilon_{t-3})$$

$$= -\theta_1\sigma_\varepsilon^2 + \theta_1\theta_2\sigma_\varepsilon^2$$

$$= -\sigma_\varepsilon^2\theta_1(1 - \theta_2)$$

and at lag two:

$$\gamma(2) = cov\,(y_t . y_{t-2})$$

$$= cov(\varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} . \varepsilon_{t-2} - \theta_1\varepsilon_{t-3} - \theta_2\varepsilon_{t-4})$$

$$= -\theta_2\sigma_\varepsilon^2$$

Similarly, one can show that: $\gamma(3) = \gamma(4) = \cdots = 0$

So the auto-covariance function for the MA(2) model can be written as:

$$\gamma(k) = \begin{cases} \sigma_\varepsilon^2\left(1 + {\theta_1}^2 + {\theta_2}^2\right). & k = 0 \\ -\sigma_\varepsilon^2\theta_1(1 - \theta_2). & k = 1 \\ -\theta_2\sigma_\varepsilon^2. & k = 2 \\ 0. & k = 3.4. \ldots \end{cases}$$

Note that the expectation, variance, and auto-covariance functions of this model do not depend on time t, (which is expected to be, since moving average processes are always stationary). Now

dividing by variance $\gamma(0)$, we get the autocorrelation function for the MA(2) model:

$$\rho(k) = \begin{cases} \dfrac{-\theta_1(1-\theta_2)}{(1+{\theta_1}^2+{\theta_2}^2)}. & k=1 \\ \dfrac{-\theta_2}{(1+{\theta_1}^2+{\theta_2}^2)}. & k=2 \\ 0. & k=3.4.\ldots \end{cases}$$

Which means that autocorrelation function of MA(2) processes cuts-off after the two time lags, thus we say that MA(2) models have a memory size of 2.

3.7.2.2 The Partial autocorrelation function of MA (2)

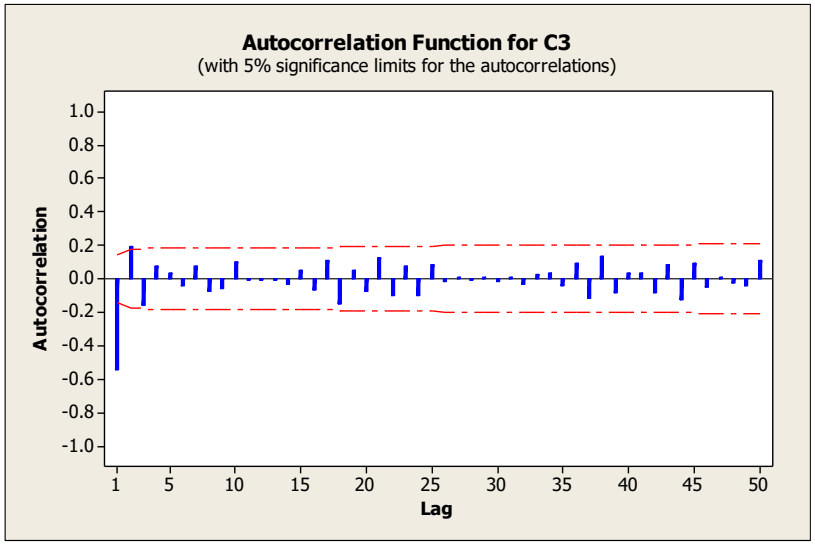
We will not derive the mathematical form of this function due to the mathematical complications, however, we will just mention the properties and form of this function:

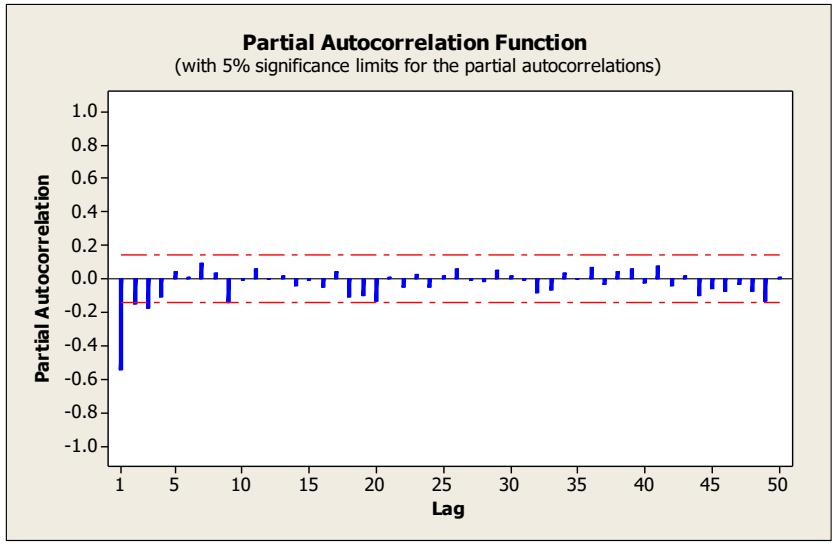
1- If the roots of the quadratic function $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 = 0$ are real, then the PACF will be in form of a decaying exponential function.

2- If the roots of the quadratic function $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 = 0$ are complex, then the PACF will be in form of a decaying sine-wave function.

To show the behavior of the ACF and PACF for the $MA(2)$ model we take the following examples for some values of θ_1 and θ_2:

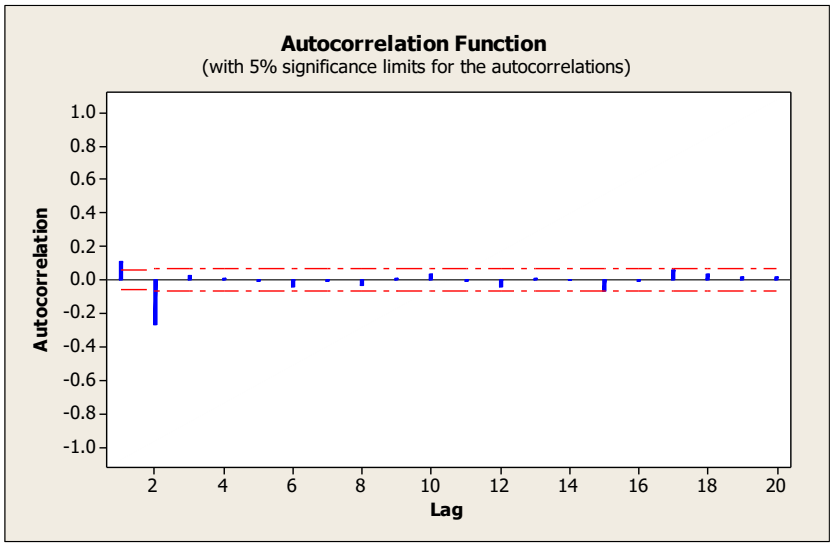
1. The following figure shows the ACF and PACF for $\theta_1 = 0.7\ , \theta_2 = -0.1$:

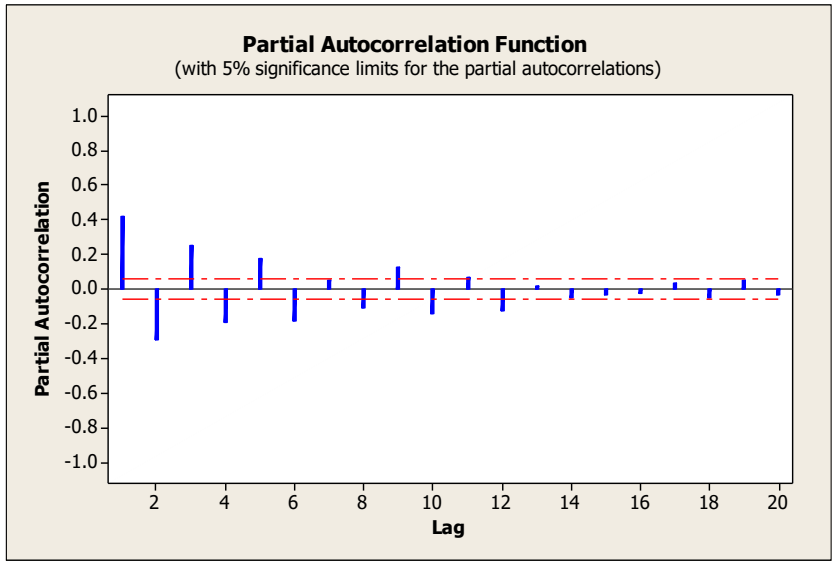




We note from figure (a) that the ACF cuts-off after lag 2, and from figure (b) that the PACF takes the form of a decaying exponential form.

2) The following figure shows the ACF and PACF for $\theta_1 = 1 \, . \theta_2 = -0.7$:





a) ACF for MA(2) model when $\theta_1 = 1 \, . \theta_2 = -0.7$

b) PACF for MA(2) model when $\theta_1 = 1 \, . \theta_2 = -0.7$

We note from figure (a) that the ACF cuts-off after lag 2, and from figure (b) that the PACF takes the form of a decaying sine-wave function.

3.7.2.3 Invertibility

We will not derive invertibility conditions for the $\mathrm{MA}(2)$ model, however we will just mention these conditions:

- $-1 < \theta_2 < 1$
- $\theta_1 + \theta_2 < 1$
- $\theta_2 - \theta_1 < 1$

Which as we can see are very similar to the stationarity conditions of the AR(2) model.

Example: If the model that best fits the process $\{y_t\}$ is

$y_t = \varepsilon_t + 0.8\varepsilon_{t-1} - 0.15\varepsilon_{t-2}$, where $\{\varepsilon_t\}$ is the white noise process, does this model satisfy the invertibility conditions?

Solution:

From the model equation, we see that the model parameters are $\theta_1 = -0.8$. $\theta_2 = 0.15$. Now applying the invertibility conditions:

(i) $|\theta_2| = |0.15| < 1$

(ii) $\theta_1 + \theta_2 = -0.8 + 0.15 = -0.65 < 1$

(iii) $\theta_2 - \theta_1 = 0.15 - (-0.65) = 0.95 < 1$

Therefore, all invertibility conditions are satisfied, and the process is invertible.

3.7.3 Moving Average of order q

This model can be written on the form:

$$y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q} \;\; ; t = 0. \pm 1. \pm 2. \ldots$$

Where $\varepsilon_t \sim WN(0\,.\sigma^2)$, and the constants $\theta_1.\theta_2.\ldots\theta_q$ are the model parameters. These models are always stationary. The models MA (q) can be invertible or non-invertible depending on the constants θ_i , but generally it can be shown that this process is invertible if the roots of the equation:

$$\theta(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right) = 0$$

all lie outside the unit circle.

And for autocorrelation function for the MA (q) models, it can be shown to have the following form:

$$\rho_k = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k+1} + \cdots + +\theta_{q-k}\theta_q}{(1 + {\theta_1}^2 + \cdots + {\theta_q}^2)} & ; k = 1.2. \ldots . q \\ 0 & ; k > q \end{cases}$$

We will not derive this mathematical equation, however we will show the pattern it can take, which is very similar to the MA(1), and MA(2) case. The ACF cuts-off after q time lags, this indicate that these processes have a memory of size q, also we can prove that

there are 2^q models with different parameters that have the same ACF , however, only one of them satisfy the invertibility condition.

As for the partial auto-correlation function it has the same pattern as MA(2) model, i.e. :

1- If the roots of the quadratic function $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q = 0$ are real, then the PACF will be in form of a decaying exponential function.

2- If the roots of the quadratic function $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q = 0$ are complex, then the PACF will be in form of a decaying sine-wave function.

3.8 Autoregressive- Moving Average Processes

We say that $\{y_t\}$ follow an Autoregressive-Moving average process of order (p, q), in short ARMA(p, q) model, if it has the following form:

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

Where $\varepsilon_t \sim WN(0\,.\sigma^2)$, and the constants $\phi_1.\phi_2.\ldots\phi_p$ and $\theta_1.\theta_2.\ldots\theta_q$ are the model parameters. We can express this process in the form:

$$\phi(B)y_t = \theta(B)\varepsilon_t \qquad (1)$$

Or,

$$\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\right)y_t = \left(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q\right)\varepsilon_t$$

Where:

$\phi(B)$: the auto-regressive operator, a polynomial in powers of B.

$\theta(B)$: the moving average operator, a polynomial in powers of B.

Now, notice from (1) that:

$$y_t = \frac{\theta(B)}{\phi(B)}\varepsilon_t$$

Which is in the form of the general linear process:

$$y_t = \psi(B)\varepsilon_t \qquad (2)$$

That is y_t can be written as an infinite moving average process, and in this case we require the roots of the characteristic equation $\phi(B) = 0$, to lie outside the unit circle as a stationarity condition for this model.

Note also, that (1) can be put alternatively in the form:

$$\varepsilon_t = \frac{\phi(B)}{\theta(B)}\, y_t$$

Which is in the form of the invertibility formula:

$$\varepsilon_t = \Pi(B)\, y_t \qquad (3)$$

That is ε_t can be written as an infinite auto-regressive process, and in this case we require the roots of the characteristic equation $\theta(B)=0$, to lie outside the unit circle as an invertibility condition for this model.

By noting both (2) and (3) we see that:

$$\Pi(B) = \psi^{-1}(B)$$

or,

$$\Pi(B)\psi(B) = 1$$

The weights ψ_j and π_j can be found by equating the coefficients of B^j in both sides of equations (2) and (3), we will see this for the model ARMA(1,1).

3.8.1 ARMA (1,1) model

We say that $\{y_t\}$ is an ARMA(1,1) process if it can be represented as:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Where $\varepsilon_t \sim WN(0.\sigma^2)$, and the constants $\phi_1.\theta_1$ are the model parameters.

This model can be put in the form:

$$\phi(B)y_t = \theta(B)\varepsilon_t$$

or,

$$(1-\phi_1 B)y_t = (1-\theta_1 B)\varepsilon_t$$

The ARMA(1,1) process is stationary if $|\phi_1| < 1$, in this case it can be expressed as an infinite moving average process, as follow:

$$y_t = \psi(B)\varepsilon_t$$

where,

$$\psi(B) = \frac{(1-\theta_1 B)}{(1-\phi_1 B)}$$

$$\Longrightarrow (1-\phi_1 B)\psi(B) = (1-\theta_1 B)$$

and thus,

$$(1-\phi_1 B)(1+\psi_1 B+\psi_2 B^2+\cdots)=(1-\theta_1 B)$$

equating the coefficients of B^j in both sides, we have;

B^1: $\psi_1-\phi_1=-\theta_1 \quad \Rightarrow \psi_1=\phi_1-\theta_1$

B^2: $\psi_2-\phi_1\psi_1=0 \quad \Rightarrow \psi_2=\phi_1\psi_1=\phi_1(\phi_1-\theta_1)$

B^3: $\psi_3-\phi_1\psi_2=0 \quad \Rightarrow \psi_3=\phi_1\psi_2=\phi_1^2\psi_1=\phi_1^2(\phi_1-\theta_1)$

Thus it is possible to get the general expression for the ψ_j weights for the ARMA(1,1) process as:

$$\psi_j=\phi_1\psi_{j-1}=\phi_1{}^{j-1}(\phi_1-\theta_1).\ \ j>0$$

The ARMA(1,1) process is invertible if $|\theta_1|<1$, in this case it can be expressed as an infinite auto-regressive process, as follow:

$$\varepsilon_t = \Pi(B) y_t$$

where,

$$\Pi(B) = \frac{(1-\phi_1 B)}{(1-\theta_1 B)}$$

$$\Rightarrow (1-\theta_1 B)\Pi(B) = (1-\phi_1 B)$$

and thus,

$$(1-\theta_1 B)(1 + \pi_1 B + \pi_2 B^2 + \cdots) = (1-\phi_1 B)$$

equating the coefficients of B^j in both sides, we get:

$\pi_1 = \phi_1 - \theta_1$

$\pi_2 = \theta_1 \pi_1 = (\phi_1 - \theta_1)$

$\pi_3 = \theta_1^2 \pi_2 = \theta_1^2 (\phi_1 - \theta_1)$

$\vdots$

Thus it is possible to get the general expression for the π_j weights for the ARMA(1,1) process as:

$$\pi_j = \phi_1 \pi_{j-1} = {\theta_1}^{j-1} (\phi_1 - \theta_1). \quad j > 0$$

It is clear from the expression of ψ_j and π_j weights that ARMA(1,1) models can be used as an appropriate approximations for either MA(∞) or AR(∞), but with merit of having a limited number of parameters (just 2!) (parsimonious law), thus mixed models are generally used instead of the moving average or the autoregressive models with large orders.

3.8.1.1 autocorrelation function for ARMA (1,1) model

The model function is:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Taking expectation on both sides, we find:

$$E(y_t) = \phi_1 E(y_{t-1}) + 0$$

therefore:

$$E(y_t) = 0$$

Taking variance of both sides:

$$var(y_t) = \gamma(0) = \phi_1^2\gamma(0) + \sigma_\varepsilon^2 + \theta_1^2\,\sigma_\varepsilon^2 - 2\phi_1\theta_1\sigma_\varepsilon^2$$

hence:

$$\gamma(0) = \frac{\sigma_\varepsilon^2(1 + \theta_1^2 - 2\phi_1\theta_1)}{1 - \phi_1^2}$$

and the auto-covariance at lag one is:

$$\begin{aligned}\gamma(1) &= cov\,(y_t . y_{t-1}) \\ &= cov(\phi_1 y_{t-1} + \varepsilon_t - \theta_1\varepsilon_{t-1} \cdot\ y_{t-1}\ \ ) \\ &= \phi_1\gamma(0) - \theta_1\sigma_\varepsilon^2\end{aligned}$$

Substituting the value of $\gamma(0)$, we get:

$$\gamma(1) = \frac{\sigma_\varepsilon^2(\phi_1 - \theta_1)(1 - \phi_1\theta_1)}{1 - \phi_1^2}$$

and at lag two:

$$\gamma(2) = cov\,(y_t . y_{t-2})$$
$$= cov(\phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \cdot y_{t-2})$$
$$= \phi_1 \gamma(1)$$

Generally, we can show that:

$$\gamma(k) = \phi_1 \gamma(k-1) \quad ; k = 2.3. \ldots$$

So the auto-correlation coefficient at lag one is:

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{(\phi_1 - \theta_1)(1 - \phi_1\theta_1)}{(1 + \theta_1^2 - 2\phi_1\theta_1)}$$

and the auto-correlation coefficient at lag k is:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi_1 \rho(k-1) \quad ; k = 2.3. \ldots$$

or,

$$\rho(k) = \phi_1^{k-1}\rho(1) \quad ; k = 2.3. \ldots$$

$$\rho(k) = \begin{cases} \dfrac{(\phi_1 - \theta_1)(1 - \phi_1\theta_1)}{(1 + \theta_1^2 - 2\phi_1\theta_1)}. & k = 1 \\ \phi_1^{k-1}\rho(1). & k = 2.3\ldots \end{cases}$$

So, it is clearly noted that for the ARMA(1,1) process, the ACF exhibits an exponential decay starting from ρ_1 not from ρ_0, as is the case in the AR(1) process. Also, note that value of ρ_1 depends on both parameters ϕ_1, θ_1, and its sign depends on the quantity $(\phi_1 - \theta_1)$, if $\phi_1 > \theta_1$ then $\rho_1 > 0$, and vice versa. After lag 1, the function

will start to decay in exponential manner if $\phi_1 > 0$, or in a decaying sine-wave format if $\phi_1 < 0$.

Thus, we notice the resemblance of the ACF shape of ARMA(1,1) model to that of the AR(1) model, the only difference is that decay starts after ρ_1 not after ρ_0.

Example:

If $y_t = 0.5y_{t-1} + \varepsilon_t + 0.9\varepsilon_{t-1}$, find the autocorrelation function and plot it, show the difference between this function and the AR(1) with same parameter.

Solution:

We have $\phi_1 = 0.5,\ \theta_1 = -0.9$, so using the formula of the ACF of the ARMA(1,1) model,

$$\rho(k) = \begin{cases} \dfrac{(\phi_1 - \theta_1)(1 - \phi_1\theta_1)}{(1 + \theta_1^2 - 2\phi_1\theta_1)}, & k = 1 \\ \phi_1^{k-1}\rho(1), & k = 2.3\ldots \end{cases}$$

we get:

$$\rho(1) = \frac{(0.5 + 0.9)(1 + 0.45)}{1 + 0.9^2 - 2(0.5)(0.91)} = 0.75$$

$$\rho(2) = \phi_1^{2-1}\rho(1) = (0.5)(0.75) = 0.375$$

$$\rho(3) = \phi_1^{3-1}\rho(1) = (0.5^2)(0.75) = 0.1875$$

$$\rho(4) = \phi_1^{4-1}\rho(1) = (0.5^3)(0.75) = 0.09375$$

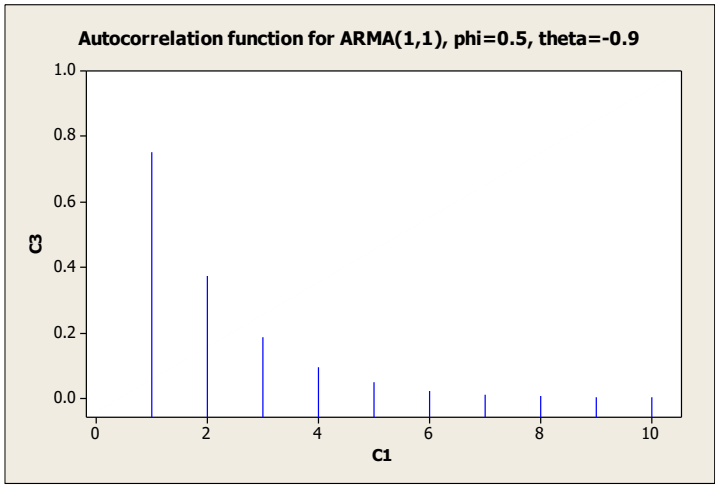
$$\rho(5) = \phi_1^{5-1}\rho(1) = (0.5^4)(0.75) = 0.046875$$

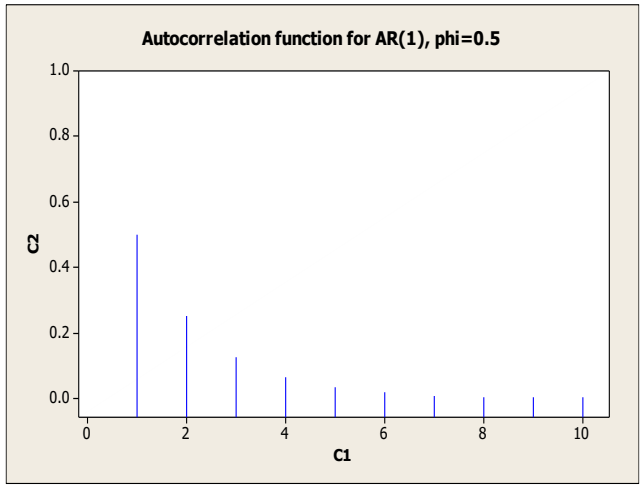
Whereas, for the AR(1) model with parameter $\phi_1 = 0.5$, and using the ACF:

$$\rho(k) = \phi_1^k \rho(0) = \phi_1^k \quad ; k = 1.2.3. \ldots$$

We get:

$$\rho(1) = 0.5 \ . \quad \rho(2) = 0.25 \ . \ \rho(3) = 0.125 \ . \ \rho(4) = 0.0625$$





(a) ACF for ARMA(1,1) model when $\phi_1 = 0.5. \theta_1 = -0.9$

(b) ACF for AR(1) model when $\phi_1 = 0.5$

So we notice the resemblance of both function, but in ARMA(1,1), the exponential decay starts from $\rho(2)$, whereas in AR(1) the decay starts from $\rho(1)$.

3.8.1.2 partial autocorrelation function model ARMA (1,1)

We can deduce the PACF for the ARMA(1,1) model by applying the definition of partial autocorrelation that we have previously addressed,

$$\phi_{00} = 1 .$$

$$\phi_{11} = \rho_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1-2\phi_1\theta_1+{\theta_1}^2} .$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - {\rho_1}^2}{1-\rho_1}$$

$$\phi_{33} = \frac{\rho_3 - \phi_{21}\rho_2 - \phi_{22}\rho_1}{1 - \phi_{21}\rho_1 - \phi_{22}\rho_2}. \;\; \phi_{21} = \phi_{11} - \phi_{22}\phi_{11}$$

The PACF for ARMA(1,1) either decay in an exponential manner, or in a sine-wave manner, exactly as the case of MA(1), except that it starts after the initial value $\phi_{11} = \rho_1$.

3.9 Integrated Autoregressive-Moving averages processes

Most of the actual time series that arise in practical applications in many areas of knowledge are **not stationary** in the mean, and thus,

we must use the difference transformation to make it stationary. Let us assume that d is the minimum order of the differences that must be taken to render the series stationary. Models that describe these processes are symbolized as $\mathrm{ARIMA}(p, d, q)$, so that to distinguish them from the stationary $\mathrm{ARMA}(p, q)$ models.

Thus, we say that a process $\{y_t\}$ is an $\mathrm{ARIMA}(p, d, q)$ process if it is possible to express it in the form:

$$\phi(B)\nabla^d y_t = \theta(B)\varepsilon_t$$

Where,

$$\phi(B) = \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right).$$

$$\theta(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right).$$

$$\nabla^d = (1 - B)^d$$

i.e.

$$y_t \sim ARIMA(p.d.q)$$

Usually the transformed series $\nabla^d y_t$ is denoted as z_t, i.e. is expressed as:

$$\phi(B)\, z_t = \theta(B)\varepsilon_t$$

Where $z_t \sim ARMA(p, q)$ which is a stationary process.

Example:

Express the ARIMA(1,1,1) in its final form.

Solution:

The ARIMA(1.1.1) model has the form:

$$(1-\phi_1 B)(1-B)y_t = (1-\theta_1 B)\varepsilon_t$$

Now putting $z_t = (1-B)y_t$, we get the model:

$$(1-\phi_1 B)z_t = (1-\theta_1 B)\varepsilon_t$$

i.e.

$$z_t = \phi_1 z_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Substituting for $z_t = y_t - y_{t-1}$, we get:

$$y_t - y_{t-1} = \phi_1(y_{t-1} - y_{t-2}) + \varepsilon_t - \theta_1\varepsilon_{t-1}$$

or,

$$y_t = (1 + \phi_1)y_{t-1} - \phi_1 y_{t-2} + \varepsilon_t - \theta_1\varepsilon_{t-1}$$

Which is the final form for the ARIMA(1.1.1) model.

Note that y_t in the previous format looks like an ARMA(2,1), which is true, however with these parameter values it is not stationary, and that after differencing $\{z_t\}$ we have turned it into a stationary ARMA(1,1) process.

Chapter 4: Parameter Estimation

We will assume that the order of the model ARMA(p,q) have been determined, i.e., we have determined the values of p and q.

Hence, we need to estimate the values of the parameters σ_a^2 , $\boldsymbol{\phi}$, and $\boldsymbol{\theta}$. In what follow, we will discuss some methods for doing so.

4.1 Method of Moments

This method is considered the simplest among estimation methods, where, as we know, the sample moments are equated to the

corresponding theoretical moments, and solving the resulting equations, one can get the required estimates.

4.1.1 Autoregressive Models

AR (1) model:

As we have shown before, $\rho_1 = \phi_1$, and we estimate ρ_1 by the sample autocorrelation coefficient r_1, thus the method of moments estimate for ϕ_1 is:

$$\hat{\phi}_1 = r_1$$

AR (2) model :

Since there are two parameters to be estimated, namely, ϕ_1 and ϕ_2, thus we need two equations for the estimation process, in this regard we can use the Yule-Walker (recall that the Yul-Walker equations have the form:

$$\rho_K = \phi_1\rho_{K-1} + \phi_2\rho_{K-2} + \cdots + \phi_P\rho_{K-P}$$

so if there are two parameters, we need the following two equations:

$$\rho_1 = \phi_1 + \phi_2\rho_1$$

$$\rho_2 = \phi_1\rho_1 + \phi_2$$

Now replace ρ_1 by r_1, and ρ_2 by r_2 , and solving these equations we get:

$$\hat{\phi}_1 = \frac{r_1(1 - r_2)}{1 - r_1^2}$$

$$\hat{\phi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2}$$

AR (p) model:

In this case we need to solve the following system of Yule-Walker equations:

$$r_1 = \phi_1 + r_1\phi_2 + \cdots + r_{p-1}\phi_p$$

$$r_2 = r_1\phi_1 + \phi_2 + \cdots + r_{p-2}\phi_p$$

$$\vdots$$

$$r_p = r_{p-1}\phi_1 + r_{p-2}\phi_2 + \cdots + \phi_p$$

which will require some more effort, but maybe mathematical software can be used for this purpose.

4.1.2 The Moving Average models

Method of moments for these models is not as easy as we have seen for AR models, it might be even impossible for models with large orders, let us consider the MA (1) model:

As we have shown earlier that:

$$\rho_1 = -\frac{\theta}{1+\theta^2}$$

so replacing ρ_1 by r_1, we get:

$$r_1 = -\frac{\theta}{1+\theta^2}$$

From which, we will get a quadratic equation in θ:

$$r_1\theta^2 + \theta + r_1 = 0$$

In case $|r_1| < 0.5$, then the real roots of the equation are:

$$\hat{\theta} = \frac{-1 \pm \sqrt{1 - 4r_1^2}}{2r_1}$$

One solution satisfy the invertibility condition $|\theta| < 1$, it is possible to check that this solution is:

$$\hat{\theta} = \frac{-1 + \sqrt{1 - 4r_1^2}}{2r_1}$$

For higher order MA models, the solutions will be more complicated.

4.1.3 Estimating the white noise variance σ_a^2

For any stationary ARMA (p, q) model, $\gamma_0 = Var(y_t)$ is estimated using the sample variance of the time series y_t:

$$S^2 = \frac{\sum_{i=1}^{n}(y_t - \bar{y})^2}{n-1}$$

and then we use the relationship between σ_ε^2 and the parameters θ or ϕ to get $\hat{\sigma}_\varepsilon^2$ for any model we want, for example:

4.1.3.1 AR (p) model

We use the following relationship that we have already obtained when discussing the AR(p) model:

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma_\varepsilon^2 \qquad (*)$$

The relationship $\rho_k = \frac{\gamma_k}{\gamma_0} \Rightarrow \gamma_k = \gamma_0\rho_k$, enable us to write (*) in

the form:

$$\gamma_0 = \phi_1\gamma_0\rho_1 + \phi_2\gamma_0\rho_2 + \cdots + \phi_p\gamma_0\rho_p + \sigma_\varepsilon^2$$

form which we have:

$$\gamma_0 - \phi_1\gamma_0\rho_1 - \phi_2\gamma_0\rho_2 - \cdots - \phi_p\gamma_0\rho_p = \sigma_\varepsilon^2$$

or:

$$\sigma_\varepsilon^2 = \left(1 - \phi_1\rho_1 - \phi_2\rho_2 - \cdots - \phi_p\rho_p\right)\gamma_0$$

and thereby, estimate of the white noise variance is:

$$\hat{\sigma}_{\varepsilon}^{2} = (1 - \hat{\phi}_1 r_1 - \hat{\phi}_2 r_2 - \cdots - \hat{\phi}_p r_p) S^2$$

For example, for AR (1) model, we have one parameter estimated as $\hat{\phi}_1 = r_1$, thus the estimate of σ_{ε}^{2} is:

$$\hat{\sigma}_{\varepsilon}^{2} = (1 - r_1^2) S^2$$

And one can estimate σ_{ε}^{2} for any AR model, for example , for AR(2) model , the equation is

$$\hat{\sigma}_{\varepsilon}^{2} = (1 - \hat{\phi}_1 r_1 - \hat{\phi}_2 r_2) S^2$$

and replace the parameter estimates $\hat{\phi}_1$ and $\hat{\phi}_2$ in terms of r_1 and r_2.

4.1.3.2 MA (q) models

We use the following relationship that we have already obtained when discussing the MA(q) model, which connects between σ_ε^2 and the parameters $\theta_1, \theta_2, \ldots, \theta_q$:

$$\gamma_0 = \sigma_\varepsilon^2 \left(1 + \theta_1^2 + \cdots + \theta_q^2\right) \Rightarrow \sigma_\varepsilon^2 = \frac{\gamma_0}{\left(1 + \theta_1^2 + \cdots + \theta_q^2\right)}$$

Estimating γ_0 by the sample variance S^2, we get the following estimate of σ_ε^2 :

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{S^{2}}{(1 + \hat{\theta}_{1}^{2} + \hat{\theta}_{2}^{2} + \cdots + \hat{\theta}_{q}^{2})}$$

For example, for MA (1) model:

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{S^{2}}{(1+\hat{\theta}_{1}^{2})} \quad \text{where} \quad \hat{\theta} = \frac{-1+\sqrt{1-4r_{1}^{2}}}{2r_{1}}$$

For the mixed model ARMA (1,1), it can be shown that the equation for estimating the white noise variance is given the following relationship:

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{\left(1-\hat{\phi}_{1}^{2}\right)}{\left(1+\hat{\theta}_{1}^{2}-2\hat{\theta}_{1}\hat{\phi}_{1}\right)} S^{2}$$

Example: Suppose that we have observed a time series Y_t of size $n = 121,$ and we have decided that AR (2) model is suitable for modelling Y_t , also, we have estimated the sample autocorrelation coefficients $r_1 = 0.936$ and $r_2 = 0.802$. The series mean $\mu = 5.1069$, and $\gamma_0 = Var(y_t) = S^2 = 1.99487$. Hence, using the following relations:

$\hat{\phi}_1 = \frac{r_1(1-r_2)}{1-r_1^2}$, and $\hat{\phi}_2 = \frac{r_2-r_1^2}{1-r_1^2}$, we can get the parameter estimates as follows:

$$\hat{\phi}_1 = \frac{0.936(1-0.802)}{1-0.936^2} = 1.50$$

and,

$$\hat{\phi}_2 = \frac{0.802-0.936^2}{1-0.936^2} = -0.598$$

and the estimate of the white noise variance is:

$$\hat{\sigma}_\varepsilon^2 = \left[1-\hat{\phi}_1 r_1 - \hat{\phi}_2 r_2\right] S^2$$

$$\hat{\sigma}_{\varepsilon}^{2} = [1 - 1.50 \times 0.936 - (-0.598)0.802]1.99487 = 0.388$$

So we may write the estimated model of this time series in the form:

$$y_t - 5.1069 = 1.5(y_{t-1} - 5.1069) + 0.598(y_{t-2} - 5.1069) + \varepsilon_t$$

where $\varepsilon_t \sim WN(0\,.0.388)$.

Note that the model can be written in an equivalent form as follow:

$$y_t = -5.6074 + 1.5y_{t-1} + 0.598y_{t-2} + \varepsilon_t$$

4.2 Least Squares method

4.2.1 AR (1) model:

The model takes the form:

$$y_t - \mu = \phi(y_{t-1} - \mu) + \varepsilon_t$$

The idea of least squares is to minimize the sum of squared errors:

$$\varepsilon_t = (y_t - \mu) - \phi(y_{t-1} - \mu)$$

That is, to minimize the term:

$$S(\phi.\mu) = \sum_{t=2}^{n} \varepsilon_t^2 = \sum_{t=2}^{n} [(y_t - \mu) - \phi(y_{t-1} - \mu)]^2$$

Thus, we find estimates of the parameters ϕ and μ by finding the corresponding values that minimize the term $S(\phi.\mu)$, so:

$$\frac{\partial S}{\partial \mu} = \sum_{t=2}^{n} 2[(y_t - \mu) - \phi(y_{t-1} - \mu)](-1 + \phi) = 0$$

And solving for μ, we find:

$$\hat{\mu} = \frac{\sum_{t=2}^{n} y_t - \phi \sum_{t=2}^{n} y_{t-1}}{(n-1)(1-\phi)}$$

Note that for large n,

$$\sum_{t=2}^{n} \frac{y_t}{n-1} \approx \sum_{t=2}^{n} \frac{y_{t-1}}{n-1} \approx \bar{y}$$

Thus whatever the value of ϕ,then:

$$\hat{\mu} \approx \frac{\bar{y} - \phi\bar{y}}{1-\phi} = \frac{\bar{y}(1-\phi)}{1-\phi} = \bar{y}$$

So we notice that the least squares method estimate μ approximately as $\bar{y}$, in case of large sample sizes.

Now, to estimate ϕ, we differentiate $S(\phi.\bar{y})$ with respect to ϕ and equate it to zero,

$$S(\phi,\mu)=\sum_{t=2}^{n}[(y_t-\mu)-\phi(y_{t-1}-\mu)]^2$$

$$\frac{\partial S(\phi.\bar{y})}{\partial\phi}=-\sum_{t=2}^{n}2[(y_t-\bar{y})-\phi(y_{t-1}-\bar{y})](y_{t-1}-\bar{y})=0$$

From which we get:

$$\hat{\phi}=\frac{\sum_{t=2}^{n}(y_t-\bar{y})(y_{t-1}-\bar{y})}{\sum_{t=2}^{n}(y_{t-1}-\bar{y})^2}$$

Note that in the denominator, we have one missing term, namely $(y_n - \bar{y})^2$, which will make $\hat{\phi}$ exactly equal r_1, but for large sample sizes the effect of this missing term will be negligible, and hence the method of moments and least squares method produce approximately equal estimates for $\hat{\phi}$ for large sample sizes.

4.2.2 MA(1) model

This model takes the form:

$$y_t = \varepsilon_t - \theta \varepsilon_{t-1}$$

We can rewrite the model in the form:

$$\varepsilon_t = y_t + \theta \varepsilon_{t-1}$$

and conditioning that $\varepsilon_0 = 0$, we find:

$$\varepsilon_1 = y_1$$

$$\varepsilon_2 = y_2 + \theta\varepsilon_1 = y_2 + \theta y_1$$

$$\varepsilon_3 = y_3 + \theta\varepsilon_2 = y_3 + \theta y_2 + \theta^2 y_1$$

$$\vdots$$

$$\varepsilon_n = y_n + \theta y_{n-1} + \cdots + \theta^{n-1} y_1$$

Now the value of θ is estimated by minimizing the sum of squares:

$$S(\theta) = \sum_{t=1}^{n} \varepsilon_t^2 = \sum_{t=1}^{n} (y_n + \theta y_{n-1} + \cdots + \theta^{n-1} y_1)^2$$

Which is a non-linear equation in θ, thus cannot be solved immediately, but we can use any numerical optimization method to solve it (for example , by the Gauss- Newton method). The same method is used in the case of higher order moving average models.

Chapter 5: Forecasting

5 .1 Introduction

The problem of forecasting is summarized in how to employ the model that passes all diagnostic tests together with the observed time series at hand to predict future values that did not occur yet, i.e. the values $y_{t+1}. y_{t+2}. \ldots$. In other words, we want to use the current and previous observations to predict the observation that will occur after l periods of time, i.e. y_{t+l} . $l = 1.2. \ldots$. We usually denote l as *forecast horizon* or, *lead time*.

Complete statistical inference for the variable y_{t+l} requires knowledge of its conditional density function, that is, its density function given that history of the time series is known up to time t. This is called the *predictive distribution*. Usually, we look for one suitable value to represent the center of this distribution in order to use it as a point estimate of the variable y_{t+l}, in addition we construct a predictive interval around this point.

The best value representing the center of the predictive distribution is the (average) or the expected value of the conditional distribution of the variable y_{t+l} given that the history of the series $y_1 . y_2 . \ldots . y_t$ is known. This conditional expectation is considered the best point estimate of this variable, because it fulfils an important characteristic which is the minimum

mean square errors, meaning that if the model for y_t is correct, then there is no other forecast produce a smaller mean squared errors.

A quick review of some of the properties of the conditional expectation:

If X and Y are random variables having joint density function $f(x.y)$, and marginal functions $f(x)$ and $f(y)$ respectively, then the conditional density function for Y given $X = x$ is:

$$f(Y|X = x) = \frac{f(x.y)}{f(x)}$$

The conditional expectation for Y given $X = x$ is:

$$E(Y|X=x) = \int_{-\infty}^{\infty} y\, f(Y|X=x)\ dy$$

Note that this is the mean of the conditional distribution, therefore all the characteristics of the mean function applies, for example:

a) $E(aY + bZ|X=x) = aE(Y|X=x) + bE(Z|X=x)$

b) $E(h(Y)|X=x) = \int_{-\infty}^{\infty} h(y) f(Y|X=x)\ dy$

Also, the mean of the conditional distribution, has the following properties:

1) $E(h(X)|X=x) = h(x)$

Which means that knowing that $X = x$, i.e. it takes a fixed value, then $h(x)$ is considered a constant function.

2) $E(E(Y|X)) = E(Y)$, and if Y and X are independent then, $E(Y|X) = E(Y)$.

5.2 Forecasting functions for ARMA models

As we have already mentioned, one of the objectives of time-series analysis is to build mathematical models and use them in forecasting future values of the time series.

Let us consider the series $\{Y_t\}$ and suppose that we can write it in the form of ARMA (p, q)

model or the general linear model form:

$$\phi(B)y_t = \theta(B)\varepsilon_t$$

Which can be written as:

$$y_t = \frac{\theta(B)}{\phi(B)}\varepsilon_t$$

$$= \psi(B)\varepsilon_t$$

Also, suppose that we have observed the series up to time t, i.e. that we have the observations $y_t. y_{t-1}. \ldots.$, let's denote it as $\tilde{y} = (y_t. y_{t-1}. \ldots)$, we will discuss how to use the available observations up to time t to predict the future value of the series at time $t+1$, i.e. y_{t+1}. We denote this predictor at time t and for one step in the future as $y_t(1)$, and in general for l steps in the future as $y_t(l)$, where t is called the time origin, and l is called the lead time.

5.3 Minimum Mean Square Error Forecast

We will denote this as $\hat{y}_t(l)$, and it is given by:

$$\hat{y}_t(l) = E(y_{t+l}|y_t . y_{t-1} . \dots) \qquad (*)$$

In other words, it is the conditional expectation of the studied phenomenon at time $t + l$, provided that the values of the phenomenon until the time t are known. We will discuss below how to get the forecasts for some ARMA models.

5.4 AR (1) Model

As we know the general form of the

AR(1) model is:

$$y_t - \mu = \phi(y_{t-1} - \mu) + \varepsilon_t$$

If we want to predict one step in the future, we replace t with $t + 1$:

$$y_{t+1} - \mu = \phi(y_t - \mu) + \varepsilon_{t+1}$$

Applying the definition of minimum mean square error forecast by taking the conditional expectation of both sides:

$$\boxed{\hat{y}_t(1) - \mu = E[(y_{t+1} - \mu)|y_t . y_{t-1} . \dots . y_1]}$$

we get:

$$\hat{y}_t(1) - \mu = \phi[E(y_t|y_t . y_{t-1} . \dots . y_1) - \mu] + E(\varepsilon_{t+1}|y_t . y_{t-1} . \dots . y_1)$$

using property number (1) of the conditional expectation we get:

$$E(y_t|y_t.y_{t-1}.\dots.y_1) = y_t$$

and since ε_{t+1} is independent from $y_t.y_{t-1}.\dots.y_1$, we get from property (2):

$$E(\varepsilon_{t+1}|y_t.y_{t-1}.\dots.y_1) = E(\varepsilon_{t+1}) = 0$$

thus,

$$\boxed{\hat{y}_t(1) = \mu + \phi(y_t - \mu)}$$

In the same way, we can find the forecast for any value l, where we replace t with $t + l$ as follows:

$$y_{t+l} - \mu = \phi(y_{t+l-1} - \mu) + \varepsilon_{t+l}$$

Thus $\hat{y}_t(l)$ is given by the conditional expectation:

$$\hat{y}_t(l) = \mu + \phi[E(y_{t+l-1}|y_t.\, y_{t-1}.\, \dots.\, y_1) - \mu] + E(\varepsilon_{t+l}|y_t.\, y_{t-1}.\, \dots.\, y_1)$$

$$\boxed{\hat{y}_t(l) = \mu + \phi[\hat{y}_t(l-1) - \mu].\quad l \geq 1}$$

Note that the previous equation provide forecasts for lead time l in terms of previous forecasts $\hat{y}_t(l-1)$. Also we can use this equation to find a prediction of any value l, in terms of the original values :

$l = 1$: $\hat{y}_t(1) = \mu + \phi(y_t - \mu)$

$$l = 2: \quad \hat{y}_t(2) = \mu + \phi[\hat{y}_t(1) - \mu]$$

$$= \mu + \phi^2(y_t - \mu)$$

$$l = 3: \quad \hat{y}_t(3) = \mu + \phi[\hat{y}_t(2) - \mu] = \mu + \phi[\mu + \phi^2(y_t - \mu) - \mu]$$

$$= \mu + \phi^3(y_t - \mu)$$

In general,

$$\boxed{\hat{y}_t(l) = \mu + \phi^l(y_t - \mu), \quad l \geq 1}$$

Example: Suppose that we have the following

AR (1) model:

$$y_t = 10 + 0.7(y_{t-1} - 10) + \varepsilon_t$$

and that the current value of the series is equal to 10.6, then one-time period ahead forecast is given as:

$$\hat{y}_t(1) = 10 + \phi^1(y_t - 10),$$

$$= 10 + 0.7 \times (10.6 - 10) = 10.42$$

and for two-time periods ahead, the forecast is:

$$\hat{y}_t(2) = 10 + \phi^2(y_t - 10),$$

$$= 10 + 0.7^2(10.6 - 10) = 10.294$$

of course, it was possible to get the forecasts in terms of previous forecasts $\hat{y}_t(.)$:

$$\hat{y}_t(1) = 10.42$$

$$\hat{y}_t(2) = 10 + 0.7[\hat{y}_t(1) - 10]$$

$$= 10 + 0.7[10.42 - 10] = 10.294$$

Remark: We can evaluate the error of one-step ahead forecast for the AR(1) model, as follow:

$$e_t(1) = y_{t+1} - \hat{y}_t(1)$$

$$= \mu + \phi(y_t - \mu) + \varepsilon_{t+1} - [\mu + \phi(y_t - \mu)] = \varepsilon_{t+1}$$

The white noise process $\{\varepsilon_t\}$ can now be reinterpreted as a sequence of one-step ahead forecast errors. We shall see that this is true for all ARMA models.

Also, the equation implies that $e_t(1)$ is independent of the process history $y_t, y_{t-1}, \ldots$ up to time t. If this were not so, the dependence could be exploited to improve our forecast.

5.5 MA(1) Model

As we know the general form of the model is:

$$y_t = \mu + \varepsilon_t - \theta\varepsilon_{t-1}$$

If we want to predict one step in the future, we replace t with $t + 1$:

$$y_{t+1} = \mu + \varepsilon_{t+1} - \theta\varepsilon_t$$

Applying the definition of minimum mean square error forecast by taking the conditional expectation of both sides:

$$\hat{y}_t(1) = \mu + E(\varepsilon_{t+1}|y_t . y_{t-1} . \cdots . y_1) - \theta E(\varepsilon_t|y_t . y_{t-1} . \cdots . y_1)$$

But:

$$E(\varepsilon_{t+1}|y_t.y_{t-1}.\cdots.y_1) = 0$$

$$E(\varepsilon_t|y_t.y_{t-1}.\cdots.y_1) = \varepsilon_t$$

so the one-step ahead forecast is:

$$\boxed{\hat{y}_t(1) = \mu - \theta\varepsilon_t}$$

and the forecast error is:

$$e_t(1) = y_{t+1} - \hat{y}_t(1)$$

$$= (\mu + \varepsilon_{t+1} - \theta\varepsilon_t) - (\mu - \theta\varepsilon_t) = \quad \varepsilon_{t+1}$$

which is the same result we obtained for the process AR (1).

To forecast future values in the process MA (1) for values $l > 1$:

$$\hat{y}_t(l) = \mu + E(\varepsilon_{t+l}|y_t.y_{t-1}.\dots.y_1) - \theta E(\varepsilon_{t+l-1}|y_t.y_{t-1}.\dots.y_1)$$

$$= \mu + E(\varepsilon_{t+l}) - \theta E(\varepsilon_{t+l-1})$$

$$= \mu + 0 - (\theta)0 = \mu \,.\, l > 1$$

In other words, in the process MA (1) if we want to predict for a period greater than one , the best prediction this process provide us is the mean of the series.

5.6 Some results for the general ARMA (p, q) process

The relationship that gives the forecasts of this model are as follows:

$$\hat{y}_t(l) = \mu + \phi_1[\hat{y}_t(l-1) - \mu] + \phi_2[\hat{y}_t(l-2) - \mu] + \cdots$$
$$+ \phi_p[\hat{y}_t(l-p) - \mu] - \quad \theta_1 E(\varepsilon_{t+l-1} | y_t . y_{t-1} . \cdots . y_1)$$
$$- \cdots - \theta_q E(\varepsilon_{t+l-q} | y_t . y_{t-1} . \cdots . y_1)$$

Where:

$$E(\varepsilon_{t+j} | y_t . y_{t-1} . \cdots . y_1) = \begin{cases} 0 & . j \geq 1 \\ \varepsilon_{t+j} & . j \leq 0 \end{cases}$$

For example, for ARMA (1,1):

The model has the form,

$$y_t = \mu + \phi(y_{t-1} - \mu) + \varepsilon_t - \theta\varepsilon_{t-1}$$

and forecasts are given by the relation:

$$\hat{y}_t(1) = \mu + \phi(y_t - \mu) - \theta\varepsilon_t \,. \qquad (1)$$

$$\hat{y}_t(2) = \mu + \phi[\hat{y}_t(1) - \mu]. \qquad (2)$$

and in general:

$$\hat{y}_t(l) = \mu + \phi[\hat{y}_t(l-1) - \mu]. \quad l \geq 2 \qquad (3)$$

also we can use the relations (1) to (3) to find forecasts in terms of the original values of the series as follow:

$$\hat{y}_t(l) = \mu + \phi^l(y_t - \mu) - \phi^{l-1}\theta\varepsilon_t. \quad l \geq 1$$

In the same way that has been used previously, we can find the forecasting error for the one-step ahead forecast $l = 1$ of the ARMA(1,1) model as follows:

$$e_t(1) = y_{t+1} - \hat{y}_t(1)$$

$$= \phi(y_t - \mu) + \varepsilon_{t+1} - \theta\varepsilon_t - [\phi(y_t - \mu) - \theta\varepsilon_t] = \varepsilon_{t+1}$$

which is the same result that have already been obtained for the other models.

The forecast error for any lead time could be written as (we will not prove this):

$$e_t(l) = \sum_{j=0}^{l-1} \psi_j \varepsilon_{t+l-j}$$

And therefore any ARMA model we have:

$$E[e_t(l)] = \sum_{j=0}^{l-1} \psi_j \, E(\varepsilon_{t+l-j}) = 0, \quad l \geq 1$$

This means that the average forecast error is equal to zero, i.e. they are unbiased. The forecast error variance is:

$$Var[e_t(l)] = \sigma_\varepsilon^2 \sum_{j=0}^{l-1} {\psi_j}^2, \quad \geq 1$$

From which we note that the forecast error variance increases as lead time increase.

5.7 Confidence intervals for forecasts

If we assume that the terms of the white noise process follow the normal distribution, then it is also possible to show that the forecast error $e_t(l)$ will also follow the normal distribution, then a $(1-\alpha)100\%$ for the future value y_{t+l} is given as,

$$\hat{y}_t(l) \pm z_{1-\frac{\alpha}{2}}\sqrt{var(e_t(l))}$$

5.8 Forecast update for ARMA (p, q) models

Suppose for instance that we study a monthly time series, and that we have observed the series until month number 6, and we have

forecasted the values of the series for months: 7,8, and 9, that is we have lead time $l = 3$. Assume that later we got the actual value of the series for the month 7. Then we can use this new value to modify our forecast for the months 8 and 9, this procedure is called *forecast update*.

In general, we have the observations $y_1, y_2, \dots, y_t$, let the time origin is t, and lead time l, our forecast for $(l + 1)$ steps ahead is denoted $\hat{y}_t(l + 1)$, and when the observation at time $t + 1$ become available, i.e. observation y_{t+1}, then we want to update our original value to be $\hat{y}_{t+1}(l)$. The equation for getting this update is:

$$\hat{y}_{t+1}(l) = \hat{y}_t(l+1) + \psi_l[y_{t+1} - \hat{y}_t(1)]$$

Example:

Suppose that the model which was applied to a time series is the AR (2), and the time origin was $t = 121$, that is we have the observed time series $y_1.y_2.\ldots.y_{121}$, and that we have the following ψ values, $\psi_1 = 1.563$ and $\psi_2 = 1.46$ and that we got the following forecasts from the model:

$$\hat{y}_{121}(1) = 5.81027, \hat{y}_{121}(2) = 5.48419,$$

$$\hat{y}_{121}(3) = 5.3215$$

Suppose now that we have obtained the actual value for time $t = 122$, which is $y_{122} = 5.9$, then our update for the forecast at time $t = 123$ (i.e. $l = 1$) becomes:

$$updated\ value = value\ before\ update + \psi_1[forecast\ error]$$

$$\hat{y}_{122}(1) = \hat{y}_{121}(2) + \psi_1[y_{122} - \hat{y}_{121}(1)]$$

$$= 5.48419 + 1.563[5.9 - 5.81027]$$

$$= 5.62444$$

Also, our update for the forecast at time $t = 124$ (i.e. $l = 2$) becomes:

$$updated\ value\ = value\ before\ update + \psi_2[forecast\ error\ ]$$

$$\hat{y}_{122}(2) = \hat{y}_{121}(3) + \psi_2[y_{122} - \hat{y}_{121}(1)]$$

$$= 5.3215 + 1.46[5.9 - 5.81027]$$

$$= 5.4525$$

Chapter 6: Box-Jenkins Methodology

The methodology developed by the scientists Box and Jenkins in their important book:

" Time Series Analysis, Forecasting and Control (1976) ", consist of several steps:

1- identification

2- estimation

3- diagnosis

4- forecasting

We have already discussed briefly the estimation step (chapter 4), and forecasting step (chapter 5), in the following sections we will look at identification and diagnosis steps with application to some data sets to be able to understand this methodology well.

6.1 identification

The first stage of the analysis of time series is to identify the initial model appropriate to the observed time series data. The meaning of identification is to choose the rank of the three parameters $(p.d.q)$, where d represent the order of differencing needed to make the series stationary, p represent number of past observations that should be

included in the initial model, i.e. the autoregressive order. Whereas, q represent number of white noise terms to be included in the initial model, i.e. the moving average order.

Application of Box and Jenkins methodology requires in addition to the theoretical foundations, skill and experience and some amount of personal judgment of the researcher. Here are some important points regarding the application of this methodology:

a) In this stage selection of the initial adequate $ARIMA(p.d.q)$ model for the time series is governed by theoretical and scientific foundations, and the skill of the researcher and his ability to judge

how the data characteristics is compatible with the characteristics of the random process that may have produced this data set.

b) The selected model in the initial stage is not final and may be modified or improved, or even to reach to a completely different model in the advanced stages of study and analysis.

c) In this stage, the researcher might arrive to different appropriate models, he has to carry these models with him for further stages of analysis hoping that at the end he will keep the best model capable of representing the characteristics of the time series data set he is analyzing.

6.1.1 Determine the rank of differences (*d*)

We mentioned earlier that most of the time series data that arise in the various application fields might show signs of non-stationarity either in the mean, the variance or in both.

In fact, non-stationarity may occur in several ways. We have earlier mentioned that judging the stationarity of certain time series by examining the roots of the characteristic equation $\phi(B) = 0$. If the roots of this equation lie outside the unit circle, it means that the series is stationarity, in which case the autocorrelation function decrease rapidly with increasing time lags. However, if a root is located on the unit circle, it means that the process or the series is not stationarity

but homogeneous. This kind of non-stationarity is the characteristic of most of the actual time series that arise in practical applications. It can be converted to a stationary series using the mathematical transformations we have seen before.

Now, how to determine appropriate value of d in order to convert non-stationarity series in the mean to a stationary one? In fact, the first thing to check before determining the value of d is to check the stationarity of the series variance, by checking the time scatter plot of the original series y_t. If the variance is not stable, it must be made stationary by taking logarithms of the original series. Usually logarithms succeed in stabilizing the variance, but in some cases we may need to use

another transformation such as square root or cubic root or any other transformation. After that, to determine the value d we follow the following steps:

- Plotting time curve of the original series y_t, and the sample autocorrelation function (SACF) r_k (the correlogram). If the time curve does not show obvious signs of existence of trend, and r_k decrease rapidly to zero as time lag increase, then the series is considered stationary and we do not need to take any differences, i.e. let $d = 0$, and move on to deciding the values of p and q.

- If the time curve shows lack of stationary in the mean and the SACF decay slowly with increasing time lag, then we must take the first differences of the time series, and then again plot the time curve, and the correlogram for the series of first differences, z_t. If both shows no sign of non-stationarity, then we let $d = 1,$ and move on to deciding the values of p and q.

- If the time curve of the series z_t still shows lack of stationary in the mean, and the SACF decay slowly with increasing time lag, then we must take the second differences of the time series, and study the transformed series in the same manner as above.

- Usually small values for d, like $d = 0,1,2$ are enough to make the time series stationary in most practical applications. Also, you should pay attention to the seriousness of taking unnecessary differences, although taking differences of a stationary series also produces stationary series, however, this process of over differencing leads to:

(1) a model that contains unnecessary additional parameters,

(2) a more complicated auto-correlation pattern,

(3) increases the variance of the series.

Example:

consider the following series:

$$y_t = \varepsilon_t$$

Where $\{\varepsilon_t\}$ is the white noise process. Discuss the stationarity of the series, and then take the first differences of the series, and again discuss the stationarity and the variance of the differenced series.

solution:

As we note the original series y_t is exactly the white noise process, which, as we know, is stationary, and have no parameters, and has auto-correlation function equal to zero for all time lags $k > 0$.

Now let's take the first differences transformation of the process y_t:

$$z_t = \nabla y_t = y_t - y_{t-1} = \varepsilon_t - \varepsilon_{t-1}$$

Thus we see that the resulting model is the moving average model of order one, with parameter $\theta = 1$, which , of course does not fulfil the invertibility condition , but it is stationary, because all the moving average models are stationary. Thus by this transformation we have complicated the model (from the simple white noise model to non invertable MA(1) model). The variance of the original model is:

$$var(y_t) = var(\varepsilon_t) = \sigma_\varepsilon^2$$

and the autocorrelation function is:

$$\rho_k = 0. \quad k \geq 1$$

Now the variance of the transformed model is:

$$var(z_t) = var(\varepsilon_t) + var(\varepsilon_{t-1}) = 2\sigma_\varepsilon^2$$

This means that the transformation has made the variance increase to double the original variance.

The autocorrelation function is:

$$\rho_k = \begin{cases} \dfrac{-\theta}{1+\theta^2}. & k = 1 \\ 0\,. & k > 1 \end{cases}$$

so we note that the degree of complexity of the correlation function has increased after transformation.

6.1.2 determine the order of the moving average and autoregressive models

After determining the necessary differences to render the series stationarity (and before that determining the need to take a logarithmic, or a square-root or other transformations to stabilize the variance), one must determine the order of the autoregressive and the moving average parts of the

model. The autocorrelation function and the partial autocorrelation function are the most effective tools in distinguishing between $AR(p)$, $MA(q)$ or $ARMA(p,q)$ models and determining the order of each of them. We here recall the theoretical forms of these functions for the $AR(p)$, $MA(q)$ or $ARMA(p,q)$ models:

Model	ρ_k	ϕ_{kk}
$AR(1)$	Approach zero exponentially or in a sinusoidal manner	Cut off completely after the first time lag

$AR(2)$	Approach zero exponentially or in a sinusoidal manner	Cut off completely after the second time lag
$AR(p)$	Approach zero exponentially or in a sinusoidal manner	Cut off completely after time lag p
$MA(1)$	Cut off completely after the first time lag	Approach zero exponentially or in a sinusoidal manner
$MA(2)$	Cut off completely after the second time lag	Approach zero exponentially or in a sinusoidal manner
$MA(q)$	Cut off completely after a time lag q	Approach zero exponentially or in

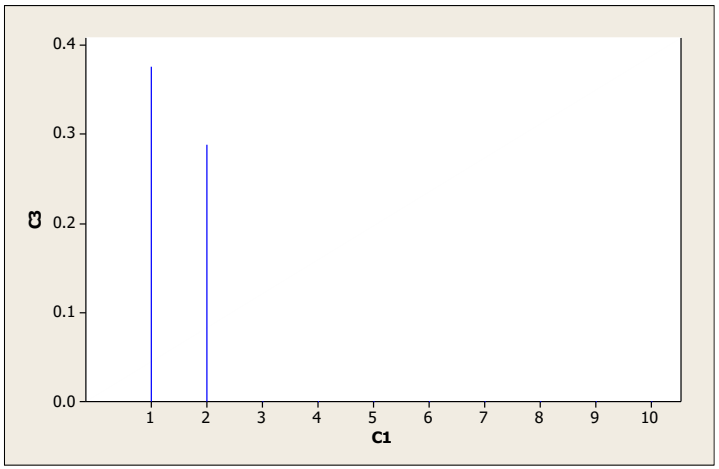
		a sinusoidal manner
$ARMA(p.q)$	Gradually approaching zero after $(q-p)$ lags exponentially or in a sinusoidal manner	Gradually approaching zero after $(p-q)$ lags exponentially or in a sinusoidal manner

The characteristics of the autocorrelation and the partial autocorrelation functions mentioned in the table above are the theoretical characteristics of the stochastic process, but, as we know, there exist differences between the theoretical characteristics of the stochastic process that generated the observed time series (what is called in the field of statistics as

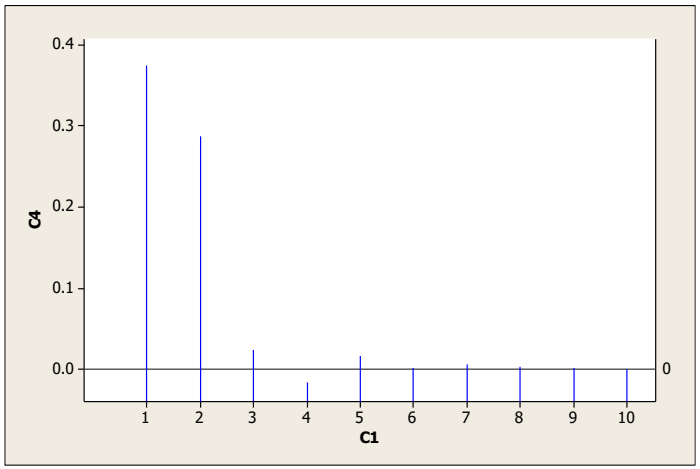
population), and the properties of the observed time series (what is called the sample) because of the sampling errors. Anyway, if the length of the series (sample size) is large, then we expect that the sample autocorrelation function r_k will reflect approximately the characteristic of the theoretical autocorrelation function ρ_k, the same is true for r_{kk} and ϕ_{kk}.

To explain this, let's consider that sampling is from a MA(2) process, where in this process the autocorrelation function ρ_k is characterized by complete cut off after time lag 2, as in figure (a) below, however, the sample that might result from such processes

might not produce an estimated autocorrelation function r_k that cut off exactly after time lag 2, see figure (b),



b) theoretical autocorrelation function ρ_k for MA(2) model



a) sample autocorrelation function r_k for MA(2) model

This means that the sample generated from MA(2) process might produce an estimated autocorrelation function with two large

values at time lags one and two, together with small autocorrelations (but do not exactly equal zero) at other time lags, but, we might consider them equal to zero. So, how do we formally test that these coefficients do not significantly differ from zero? To answer this question, we recall the results deduced by Bartlet (1940) where it was shown that one can use the test statistic

$z = \frac{r_k}{SE(r_k)}$ to test that the function ρ_k cuts off after a certain number of lags, q, for instance. We can infer this statistically by testing the significance of the coefficients of ρ_k after lag q. The initial impression we got from figure (b) is that the theoretical

autocorrelation function might take the form in figure (a), in this case, to ascertain this first impression is to test the hypothesis:

$$H_0: \rho_3 = 0 \quad \text{vs} \quad H_1: \rho_3 \neq 0$$

If H_0 is accepted, then we have to test,

$$H_0: \rho_4 = 0 \quad vs \quad H_1: \rho_4 \neq 0$$

If H_0 is accepted, then we have to test ρ_5, and so forth.

Usually, the results of the tests are clear by simply comparing r_k with double the standard error without the need to calculate the test statistic z, where we reject $H_0: \rho_k = 0$ if:

$$|r_k| > 2SE(r_k). \quad k = q + 1. q + 2. \ldots$$

With respect to the partial autocorrelation function, how can we test the significance of its coefficients, i.e., how to decide on the order of the AR(p) model? To answer this question we refer to the results deduced by Anderson and Quenelle, where one can use the statistic $z = \frac{r_{kk}}{SE(r_{kk})} = \frac{r_{kk}}{1/\sqrt{n}} = r_{kk}\sqrt{n}$ which follow approximately the standard normal distribution to test the cut off point for the function ϕ_{kk} after any time lag. So, to infer statistically about the

significance of the coefficients of ϕ_{kk} after time lag $p+1$, we have the following hypothesis:

$$H_0: \phi_{kk} = 0 \quad vs \quad H_1: \phi_{kk} \neq 0 \quad ; k = p+1. p+2. \ldots$$

If these coefficients do not differ significantly from zero, then we can accept the hypothesis that the theoretical function ϕ_{kk} cut off after time lag p, and hence we choose the right order of the AR(p) model.

With regard to mixed ARMA(p,q) models, in fact the situation is more complicated to identify their order than the pure AR (p) or pure MA (q) models, but we just mention here that both the

autocorrelation and partial autocorrelation functions decay exponentially or in the form of sine wave functions.

Example:

The following data illustrate the autocorrelation and partial autocorrelation functions for a time series with length 100 observations. Specify initial model suitable for the series:

k	1	2	3	4	5	6	7	8	9	10
r_k	0.405	-0.073	0.08	0.11	0.092	-0.09	0.1	0.1	-0.09	0.052

r_{kk}	0.405	0.32	0.24	-0.11	0.09	-0.02	0.01	0.03	-0.05	0.03

Solution:

The autocorrelation function r_k seems to cut off after the first time lag, thus, we first conduct a test for the significance of ρ_1 assuming that the stochastic process generated the data is purely random , that is a white noise process, i.e. $q = 0$, thus for all time lags k, we have,

$$SE(r_k) \cong \sqrt{\frac{1}{n}} = \sqrt{\frac{1}{100}} = 0.1.\ k > 0$$

So, to test the hypothesis:

$$H_0: \rho_1 = 0 \quad vs \quad H_1: \rho_1 \neq 0$$

We use the test statistic:

$$z = \frac{r_1}{SE(r_1)} \approx \frac{0.405}{0.1} \cong 4.05 > 2$$

Hence, we reject H_0, and deduce that ρ_1 is significantly different from zero, and that the stochastic process generated the series cannot be a pure random process. The question now arises; can we assume that all other autocorrelation coefficients do not differ significantly from zero? To answer this question, we have to

calculate the standard error of the process assuming the process is MA (1), i.e.

$$SE(r_k) \cong \sqrt{\frac{1}{n}\,(1 + 2\,r_1^2)}. \quad k > 1$$

$$\cong \sqrt{\frac{1}{100}\,[1 + 2\,(0.405)^2]} = 0.115$$

$$hence, \qquad 2SE(r_k) \cong 2(0.115) = 0.23 \;\; ; k > 1$$

By inspecting all estimated autocorrelation coefficients in the table, we see that $|r_k| < 0.23$ for all values k=2,3,… , we see that autocorrelation function cuts off after the first time lag which

indicate that the MA(1) model is a tentative possible model for the series.

Example:

The following data illustrate the autocorrelation and partial autocorrelation functions for a time series with length 92 observations. Specify initial model suitable for the series:

k	1	2	3	4	5	6	7	8	9
r_k	0.66	0.42	0.29	0.19	0.09	-0.01	0.01	0.02	0.01
r_{kk}	0.66	0.39	0.01	0.02	-0.01	-0.03	0.02	0.01	0.01

Solution:

The partial autocorrelation function r_{kk} seems to cut off after the second time lag, thus, we first conduct a test for the significance of ϕ_{22} assuming that the stochastic process generated the data is AR(1), thus we have,

$$SE(r_{kk}) \cong \sqrt{\frac{1}{n}} = \sqrt{\frac{1}{92}} = 0.104$$

So, to test the hypothesis:

$$H_0: \phi_{22} = 0 \quad vs \quad H_1: \phi_{22} \neq 0$$

We use the test statistic:

$$|z| = \frac{r_{22}}{SE(r_{22})} \approx \frac{0.39}{0.104} \cong 3.7 > 2$$

Hence, we reject H_0, and deduce that ϕ_{22} is significantly different from zero, and that the stochastic process generated the series cannot be AR(1) process, hence we assume it is AR(2), thus the standard error for all time lags $k > 2$ is:

$$SE(r_{kk}) \cong \sqrt{\frac{1}{n}} = \sqrt{\frac{1}{92}} = 0.104 \quad ; k > 2$$

$$2SE(r_{kk}) \cong 0.208 \quad ; k > 2$$

By inspecting all estimated partial autocorrelation coefficients in the table, we see that $|r_{kk}| < 0.208$ for all values k=2, 3..., thus there is evidence that it cuts off after the second time lag. Hence, the AR (2) model seems a tentative possible model for the series, especially that the autocorrelation function seems to decay exponentially.

Example:

The following data illustrate the autocorrelation and partial autocorrelation functions for a monthly time series of length was 400 months representing the number of car accidents occurred in a city:

k	1	2	3	4	5	6	7	8	9
r_k	0.85	0.45	0.28	0.15	0.10	0.06	0.03	0.02	0.01
r_{kk}	0.85	0.61	0.45	0.40	0.30	0.20	0.11	0.10	0.05

Specify initial model suitable for the series, if the series length was 400 months.

Solution:

Obviously, both autocorrelation and partial autocorrelation functions do not seem to cut off after short time lags, which might indicate that mixed model is suitable for modelling the data. Note also that r_k start decay from r_1 not from r_0 which might indicate that ARMA(1,1) model might be suitable to model the data, what

support this choice is that behavior of r_{kk} seems similar to MA(1) behavior.

6.3 Diagnostics

Time Series model identified in the first stage depends on an important theoretical hypothesis of the stochastic process that generated the data set, and on the general form of the model and the random shocks ε_t. This means that parameter estimates and its statistical properties and inferences have no meaning unless these assumptions are fulfilled, or at least cannot be rejected for the available data set. Thus, investigating the appropriateness of these

assumptions is a corner stone of studying and analyzing time series. Such investigation is called model diagnostics, which can be seen as a balance between theoretical assumptions the model is based on and the practical output of the estimation stage. Diagnostics is the third stage of Box-Jenkins methodology, after initial identification of the tentative model and estimation of its parameters, then comes the third stage of making sure that estimated model comply with theoretical assumptions, or that at least do not show a clear deviation from these assumptions. This stage is the most serious and important stage of the analysis, as it can assure us that the model is adequate and thus can be used for

forecasting, or it might show that the model has to be modified according to these diagnostics. In general, model diagnostics depend on conducting several checks and tests, the most important are:

1- stationarity analysis
2- invertibility analysis
3- residual analysis
4- fitting a lower model
5- fitting a higher model

6.3.1 Stationarity Analysis

We have mentioned before that the conditions for stationarity requires that the roots of the characteristic equation $\phi(B) = 0$ must all be outside unity circle. Therefore, in the estimation stage, if the absolute value of each root is outside the unit circle then this indicates that the process generated the observed series is stationary, but if the absolute value of one root is close to 1, this indicate the need to take additional differences, adjusting the initial model Consequently.

Example:

Assume that the identified and estimated model for an observed time series is ARIMA (1,0,1), that is, it has the form:

$$(1-\phi_1 B)y_t = (1-\theta_1 B)\varepsilon_t$$

If the parameter ϕ_1 does not differ significantly from 1 , then the model can be re - written in the form:

$$(1 - B)y_t = (1-\theta_1 B)\varepsilon_t$$

or,

$$z_t = (1-\theta_1 B)\varepsilon_t$$

Where,

$$z_t = (1 - B)y_t = y_t - y_{t-1}$$

This process is stationary, which means that the model

ARIMA (0,1,1) or IMA (1,1) may be better than ARIMA(1,0,1) to model the time series.

Example:

After initial estimation of the model ARIMA (2,0,1) for time series data y_t, it was found that one of the roots of the characteristic

equation $\phi(B) = 0$ is near to 1. Suggest a better model for the data than the initial model.

solution:

The original model is $(1 - \phi_1 B - \phi_2 B^2) y_t = (1 - \theta_1 B)\varepsilon_t$, and since one of the roots of $1 - \phi_1 B - \phi_2 B^2 = 0$ is near to 1, then the original model could be written as:

$$(1 - B)(1 - \phi_1 B) y_t = (1 - \theta_1 B)\varepsilon_t$$

Which means that the series is not stationary, thus;

$$(1 - \phi_1 B) z_t = (1 - \theta_1 B)\varepsilon_t$$

is a stationary process, this means that the model ARIMA(1,1,1) might be a better model than the original ARIMA (2,0,1) model.

6.3.2 invertibility analysis

We have mentioned the importance of invertibility condition for time series models, and thus it is very important to examine the estimates of the moving average parameters to check that the invertibility conditions are satisfied. These conditions are that the roots of the equation $\theta(B) = 0$ should all be outside the unit circle. However, if one root was near to one, then this might indicate we have taken extra unnecessary differences.

Example:

Assume that the identified and estimated model is ARIMA(1,1,1), i.e. has the form:

$$(1-\phi_1 B)z_t = (1-\theta_1 B)\varepsilon_t$$

where,

$$z_t = y_t - y_{t-1} = (1-B)y_t \qquad (1)$$

assuming that the value of the parameter θ_1 does not differ significantly from 1, this means:

$$(1-\phi_1 B)z_t = (1-B)\varepsilon_t$$

or,

$$(1-\phi_1 B)(1-B)^{-1} z_t = \varepsilon_t \qquad (2)$$

Substituting from (1) into (2):

$$(1-\phi_1 B) y_t = \varepsilon_t$$

Which means that the model ARIMA(1,0,0) may be better than the original model ARIMA(1,1,1) in modeling the time series.

Example:

After initial estimate of the model ARIMA (1,1,2) for the time series data, it was found that one of the roots of the equation $\phi(B) = 0$ is near to 1. Propose an alternative model that might be a better fit to the data than the original model.

Solution:

The original model is $(1-\phi_1 B)(1-B)y_t = (1-\theta_1 B-\theta_2 B^2)\ \varepsilon_t$

As one of the roots of the equation $1-\theta_1 B-\theta_2 B^2=0$ is near to 1, then the model could be written as :

$$(1-\phi_1 B)(1-B)y_t = (1-B)(1-\theta_1 B)\varepsilon_t$$

or,

$$(1-\phi_1 B)y_t = (1-\theta_1 B)\varepsilon_t$$

which means that the model ARIMA(1,0,1) might be better in fitting the data than the original ARIMA(1,1,2) model.

6.3.3 Residual analysis

If the model that was chosen in the first phase truly represents the characteristics of the random process that generated the time series at hand, then the residuals resulting from the estimation phase should fulfill the theoretical assumptions postulated for the random shocks ε_t, or at least, these residuals do not show serious deviations from these assumptions, the most serious one being "ε_t are not correlated".

If we assume that $\hat{\varepsilon}_1 . \hat{\varepsilon}_2 . \ldots . \hat{\varepsilon}_n$ represent the residuals after fitting the initial model to the available time series observations, and this model was a good fit for the data, then the model residuals should not show any patterns or regular movements that can be predicted, in other words, the residuals should reflect the main characteristics of the variables ε_t, which are:

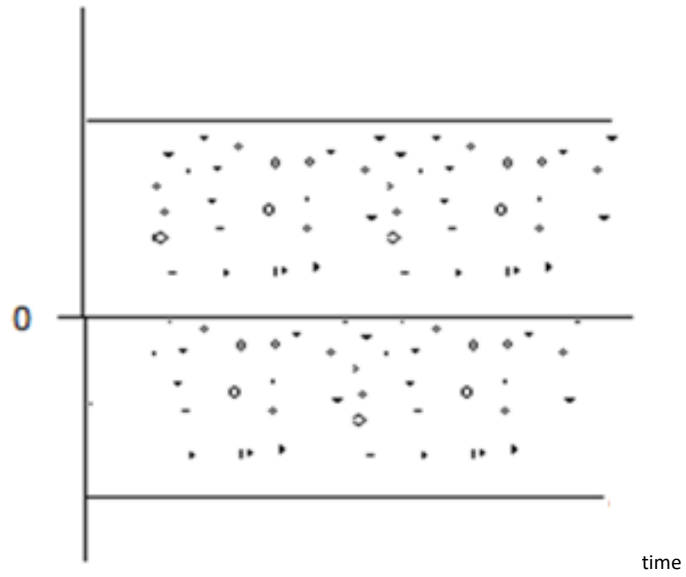
1- random variables
2- with zero mean
3- and a constant variance
4- and follow the normal distribution
5- and are uncorrelated

For checking these assumptions, we have to plot the residuals as a time series, check the autocorrelation function for the residuals $\hat{\varepsilon}_t$, plot the histogram for the residuals, conduct some non-parametric tests for checking the randomness and normality of the residuals and that their mean is not significantly different from zero, use the modified Box-Pierce statistic. We will go through these steps in some detail in the following sections.

6.3.3.1 Plotting the residuals

The first and most important step in the residual analysis is to plot the residuals graphically, where the horizontal axis represents time

and vertical axis represents residuals $\hat{\varepsilon}_t$. This is a vital and irreplaceable step, as it can reveal the principal features of the residuals such as the trend, the variance, and outliers if they exist, in such a way even the statistical tests might not be able to reveal. If the initial model was adequate, then this means that it can accommodate all the patterns and the regular movements in the time series data, leaving residuals free of any pattern, thus the residual plot should show them oscillating with a constant variance around the vertical line passing through zero. Also, this plot should be looking random and free of any information that can be used in forecasting the time series.



6.3.3.2 Randomness of the residuals

The randomness of the residuals is tested by Runs test around zero, which is one of the non-parametric tests, the command to perform the test in MINITAB is:

MTB > RUNS 0 C_k

where the column C_k that contains the estimated residuals.

6.3.3.3 Test that the residuals mean is equal to zero

The hypothesis that we test here is:

$$H_0: E(\varepsilon_t) = 0 \quad vs \quad H_1: E(\varepsilon_t) \neq 0$$

which is a two-tailed test and we use the test statistic $u = \frac{\overline{\hat{\varepsilon}_t}}{se(\overline{\hat{\varepsilon}_t})}$, which has the standard normal distribution. So, at significance level $\alpha = 0.05$, we consider $E(\varepsilon_t) = 0$, if $|u| < 1.96$ (assuming the

sample size is at least 30, which is satisfied in most time series data). The command to perform the test in MINITAB is:

MTB > OneZ C_k ;

SUBC> Test 0

6.3.3.4 Constant variance

As mentioned in previous sections, plotting the residuals reveals important issues, including whether the residual variance is constant or not. If the variance is constant, the plot will

approximately reveal this point. If we observe increasing or decreasing variance in the residual plot, then we must return to the original series and use some transformation to try to stabilize the variance, and analyze the data again.

6.3.3.5 Autocorrelation function of residuals

If the errors ε_t are purely random variables, then the estimated residuals $\hat{\varepsilon}_t$ must reflect this fact, thus the autocorrelation function must be free of any spikes, that is, all the autocorrelation

coefficients ought to be small in order to accept that the corresponding theoretical coefficients are not significantly different from zero. We check every autocorrelation coefficient separately, thus we have to check the sampling distribution of these coefficients. Anderson (1942) have shown that if the model was appropriate, then the autocorrelation coefficients for large and medium sample sizes are uncorrelated and follow normal distribution with standard deviation $n^{-1/2}$. Hence, the autocorrelation coefficient of the residuals at a certain lag that fall

outside the interval $\pm 2/\sqrt{n}$ support that the corresponding theoretical coefficient is significantly different from zero.

In spite of the simplicity of conducting this test, however the approximate variance $\frac{1}{n}$ is greater than the actual variance for the autocorrelation coefficients at small lags. Thus, if the autocorrelation function is free of any spikes, then this is an important indication that ε_t represent purely random variables, however it is not sufficient, as some autocorrelation coefficients at small lags might be inside the interval $\pm 2/\sqrt{n}$ but actually the corresponding theoretical coefficient is significantly different from

zero if compared to the true standard deviation which is less than $1/\sqrt{n}$. This means that it is not sufficient to plot the autocorrelation coefficient with the interval limits $\pm 2/\sqrt{n}$ to conclude that ε_t are random, but we have to conduct further checks and tests to assure that these variables are random.

In fact, the results and outcomes of the estimation stage and calculation of the autocorrelation function for the residuals remains particularly important, even if these results do not support that the model is appropriate because the spikes noted in the autocorrelation function might be used to adjust the initial

model. For example, if the autocorrelation function of the residuals shows a spike at the first time lag, this may be an evidence for the need to add a moving average parameter to the initial model especially if the partial autocorrelation function of the residuals behaves in an exponential function shape. Suppose for example that the initial model we have chosen for the series y_t is an MA (1) which has the form:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} = (1 - \theta_1 B)\varepsilon_t$$

If we assume that examination of the autocorrelation function of the residuals show that the errors are not random, but follows the MA (1) model as well, then,

$$\varepsilon_t = a_t - ca_{t-1} = (1 - cB)a_t$$

where $\{a_t\}$ is a white noise process. Substituting for ε_t we get:

$$y_t = (1-\theta_1 B)(1 - cB)a_t$$

$$= a_t - \theta_1^* a_{t-1} - \theta_2^* a_{t-2}$$

where,

$$\theta_1^* = (\theta_1 + c) \quad ; \theta_2^* = -c\theta_1$$

This means that $\{y_t\}$ follow MA(2) and not MA(1), in which case we have to go back and fit an MA(2) for the time series, estimate its parameters and perform diagnostic checks again to make sure it fits the data well.

On the other hand, if the autocorrelation function of the residuals decreases exponentially, or gradually approaching zero interchanging in sign, then the original initial model

MA(1) may need the inclusion of an autoregressive parameter, especially if the partial autocorrelation function of the residuals completely cut off after the first time lag. In this case,

the initial model is modified to the ARMA (1,1) model, fitting it to the time series, estimate its parameters and perform diagnostic checks again to make sure it fits the data well.

6.3.3.6 Modified Box and Pierce statistic

Checking every coefficient of the autocorrelation function of the residuals is an important indication of the appropriateness of the model assumptions, the most important assumption is the randomness of the ε_t variables. But, it is not sufficient to just perform this diagnostic for two reasons. First – which we have

mentioned above- that there exist some difficulties at small time lags that lead mistakenly to consider a theoretical autocorrelation coefficient at a small time lag not significantly different from zero, when in fact it differs significantly from zero if we used the true variance instead of the approximate variance $\frac{1}{n}$. The second reason is that some spikes might exist especially at large time lags, but the model is still considered appropriate, since the randomness of the variables ε_t does not prevent existence of some large coefficients in the sample (because the estimated residuals $\hat{\varepsilon}_t$ are considered as a

sample from the process $\{\varepsilon_t\}$), upon which we may accept that the corresponding theoretical coefficients are different from zero.

For these reasons it is necessary to examine the appropriateness of the model using a different philosophy. Instead of checking every autocorrelation coefficient $r_{\hat{\varepsilon}_t}(j)$ separately, it is possible to check that a group of coefficients all together are equal to zero. Suppose that we denote the first k terms of the residual autocorrelation coefficients as $r_{\hat{\varepsilon}_t}(1), r_{\hat{\varepsilon}_t}(2), \ldots, r_{\hat{\varepsilon}_t}(k)$ resulting from fitting ARMA(p,q) model to the series y_t, Box and Pierce

(1970) proposed a test such that if the fitted model is appropriate then the statistic:

$$Q = n \sum_{j=1}^{k} r_{\hat{\varepsilon}_t}^2(j)$$

has, for large sample sizes, a χ^2 distribution with $(k - p - q)$ degrees of freedom. Thus if some coefficients are not sufficiently close to zero, then Q tends to be large. In general, we do not reject the randomness of the autocorrelation coefficients –or equivalently- the appropriateness of the model if calculated value of Q is less than the tabulated χ_α^2 where,

$$P\left[\chi^2_{(k-p-q)} > \chi^2_{\alpha}\right] = \alpha$$

α is the significance level. The value of k is subjective and is chosen by the analyst, the power of the test decrease as k increase. The statistic Q works well if the sample size is large or moderately large, however for small sample sizes it's power decrease. For small sample sizes the approximation of Q by the χ^2 distribution is not good, for this reason Ljung-Box introduced a modified statistic in the form:

$$Q^* = n(n+2)\sum_{j=1}^{k}\frac{r^2_{\hat{\varepsilon}_t}(j)}{(n-j)}$$

which has a better approximation to the χ^2 distribution with $(k-p-q)$ degrees of freedom.

Example:

The following table shows the first 12 autocorrelation coefficients for the residuals resulting from the fitting ARMA $(1,1)$ model for a time series of length 100 observations.

k	*1*	2	3	4	5	6	7	8	9	10	11	12
$r_{\hat{\varepsilon}_t}(k)$	*0.03*	0.04	-0.3	-0.1	0.01	-0.03	0.02	-0.05	0.3	0.1	0.08	-0.1

1-Test the significance of each theoretical correlation coefficient (i.e. that it is different from zero at each time lag).

2-Test the appropriateness of the model using Box-Pierce statistic.

3-Test the appropriateness of the model using modified Ljung-Box statistic.

Solution:

1-We first calculate the standard error $\frac{1}{\sqrt{n}} = \frac{1}{\sqrt{100}} = 0.1$, and hence the approximate 95% confidence limits are $\pm\frac{2}{\sqrt{n}} = \pm 0.2$, then comparing each correlation coefficient with this interval, we see

that $\rho_\varepsilon(3)$ and $\rho_\varepsilon(9)$ are both significantly different from zero at significance level 5%.

2- Box-Pierce statistic:

$$Q = n \sum_{j=1}^{k} r_{\hat{\varepsilon}_t}^2(j) = 100[(0.03)^2 + (0.04)^2 + \cdots + (-0.01)^2] = 22.28$$

and since the tabulated value is $\chi^2_{0.05\ .10} = 18.3$, that is $Q > \chi^2_\alpha$, then we say that there is some doubt about the appropriateness of the model.

3- Modified Ljung-Box statistic:

$$Q^* = n(n+2)\sum_{j=1}^{k}\frac{r^2_{\hat{\varepsilon}_t}(j)}{(n-j)}$$

$$= 100(102)\left[\frac{(0.03)^2}{99} + \frac{(0.04)^2}{98} + \cdots + \frac{(-0.01)^2}{88}\right] = 24.33$$

and since the tabulated value is $\chi^2_{0.05\,.10} = 18.3$, that is $Q^* > \chi^2_{\alpha}$, then we say that there is some doubt about the appropriateness of the model.

Example:

The following table shows the first 10 autocorrelation coefficients for the residuals resulting from the fitting ARMA (0,2,1) model for a time series of length 123 observations.

k	1	2	3	4	5	6	7	8	9	10
$r_{\hat{\varepsilon}_t}(k)$	0.01	0.02	-0.01	-0.10	0.10	0.01	0.02	0.04	0.03	0.1

Test the appropriateness of the model using Box-Pierce statistic.

Solution:

Since the model used a difference of order 2, then we lose two observations from the series, hence the effective number of observations is:

$$n^* = 123 - 2 = 121$$

Hence,

$$Q = 121[(0.01)^2 + (0.02)^2 + \cdots + (-0.1)^2] = 4.0656$$

the tabulated value is $\chi^2_{0.05,9} = 16.9$. Because $Q < \chi^2_{\alpha}$, we deduce that there is no non-random pattern in the first 10 autocorrelations of the residuals, and hence the model is appropriate for the observed time series.

6.3.5 Fitting the lower order model

We have mentioned previously that the identification stage depends partially on personal judgment of the researcher, as testing the cut off points of both the autocorrelation and partial autocorrelation functions depend on the used significance level, where large significance levels are used for small time lags, and small significance levels for larger time lags. Sometimes the model might contain a parameter of large order, then simplification of the model, i.e. fitting the next lower model is achieved by dropping the largest order parameter from the model.

Thus, it is necessary to perform some additional checks apart from the residual analysis and the estimation stage outcomes. We must study whether the lager order parameter is significantly different from zero by comparing this parameter estimate with double its standard error. If it is less than double of the standard error, then it is preferred to omit this parameter from the model. But before omitting the parameter one must investigate the correlation of the parameter estimate with all other parameter estimates. If we notice existence of strong correlation, then this is a good indication that the model can be further simplified, and thus fitting a lower order model is justified. It is important to subject the reduced

model for all diagnostic tests and checks to make sure the other parameters could compensate for the effect of dropping the higher order parameter.

6.3.6 Fitting the higher order model

We can also answer the following question: Can the model efficiency be improved by adding an extra parameter? for example if the original model we have found suitable for the data is an MA(1) then, one can add an extra moving average parameter to this model, and hence fit an MA(2) model to the data, and study the

improvement in the diagnostic checks of the model, also study the significance of the added parameter θ_2, and the correlation between $\hat{\theta}_2$ and $\hat{\theta}_1$. If it is found that the added parameter θ_2 is not significant , or that the correlation between $\hat{\theta}_2$ and $\hat{\theta}_1$ is large, then we have to drop the added parameter θ_2, and just keep the original model MA(1), and vice versa.

Of course, one could have added an autoregressive parameter ϕ_1 to the original model, i.e. fit the model ARMA(1,1) to the data and study the efficiency of adding the parameter ϕ_1 in the same manner.

What we want to emphasize here, is that testing of omitting or adding some parameters depend to a high extent on the experience of the researcher and his personal judgment, that's why we say that the identification and diagnostic checks are the most important stages in the modern time series analysis, and are the vital steps in getting trustable forecasts.

6.4 Practical example of time series analysis

The following time series represent a count of the number homicide cases recorded in Australia during 1915-1993, it represents the rate number of yearly homicide for every 100000.

Analyze these data, write a full report about your findings, use the proposed model to forecast rate of homicide cases for the next five years.

6.4.1 General form of a data analysis report

Introduction:

We have a time series data that represent rate number of yearly homicide for every 100000 recorded in Australia during 1915-1993. Since the data are recorded serially over the years, then we would expect them to be correlated over time. Thus we can study the autocorrelation structure of the data to see how it behave, and based on this structure we can develop a mathematical model that can describe how the rate number of homicide develop over time in Australia, we will also use this developed model to forecast the

rate of homicide in the next five years, and construct a 95% confidence limits for these forecast.

Data description:

Figure (1) shows rate number of yearly homicide for every 100000 recorded in Australia during 1915-1993:

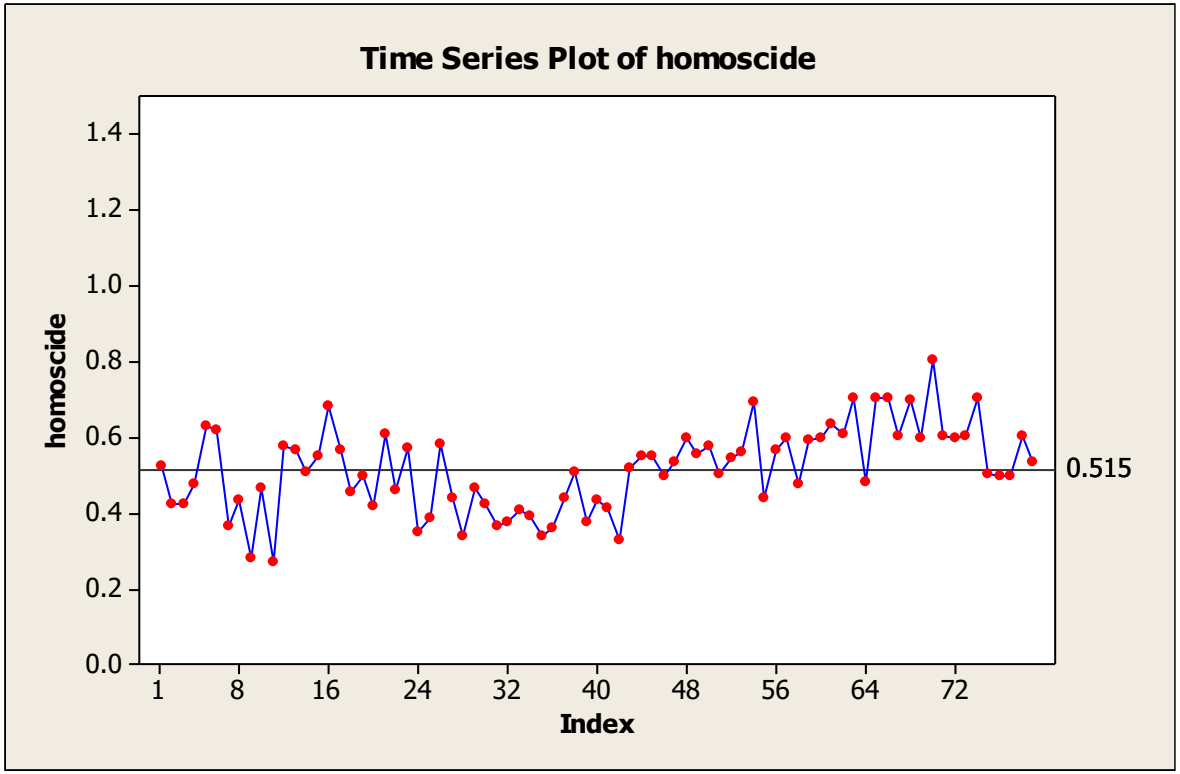


Figure (1): rate number of yearly homicide for every 100000 recorded in Australia during 1915-1993

From figure (1) we notice that the data seems to be stationary in the mean as we do not notice any long term increase or decrease in

the series, the series oscillate around the mean (the value 0.5149). We also, do not notice any seasonal pattern in the data, or any outliers.

a)The autocorrelation function of the data:

Figure (2) shows the autocorrelation function of the data:

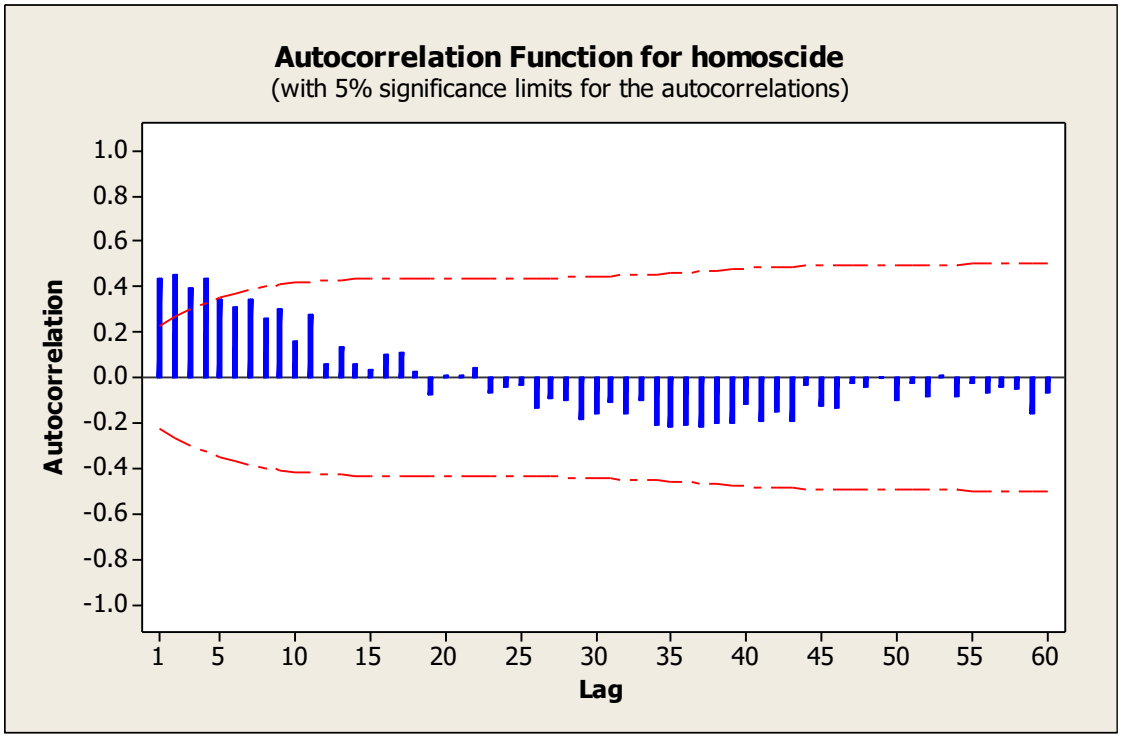


Figure (2): autocorrelation function of rate number of yearly homicide for every 100000 recorded in Australia during 1915-1993

We notice that the autocorrelation function takes the form of an exponential decay function, this is a common feature of the autoregressive models.

b) Figure (3) shows the partial autocorrelation function of the data:

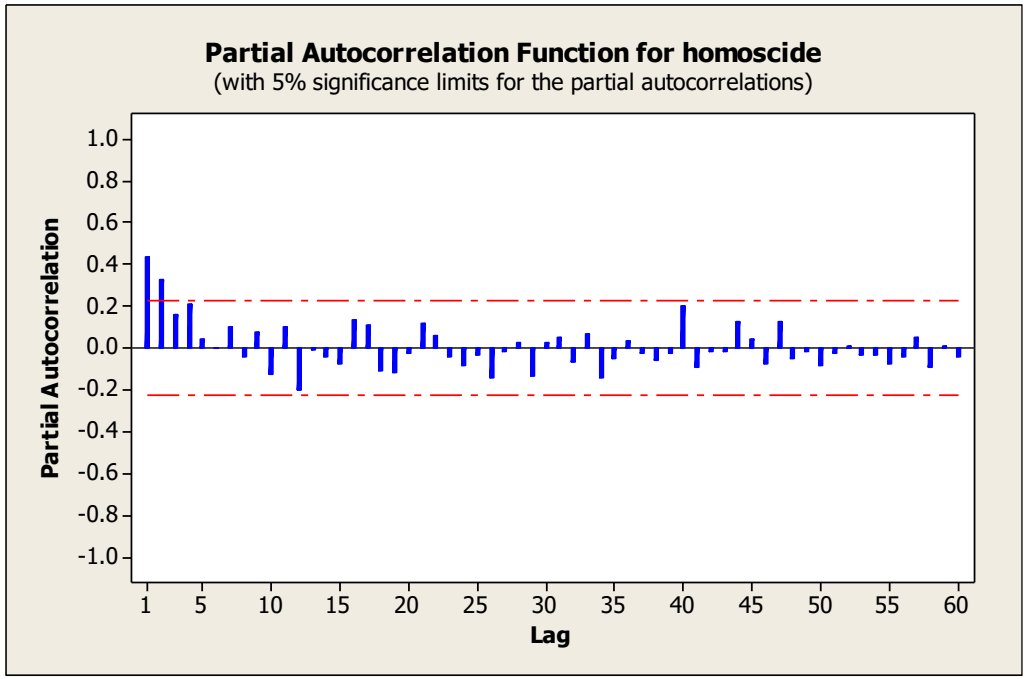


Figure (3): Partial autocorrelation function of rate number of yearly homicide for every 100000 recorded in Australia during 1915-1993

We notice that in the Partial autocorrelation function two values at time lags $k = 1.2$ seems to differ significantly from

zero, also we can imagine that the function takes the form of an exponential decay function. Thus, from the structure of the estimated autocorrelation and partial autocorrelation functions of the data we can propose that the models AR(1), AR(2), or ARMA(1,1) are potential models to describe the evolution of the rate number of yearly homicide for every 100000 recorded in Australia during 1915-1993.

Fitting proposed models:

(i) Autoregressive model of order one AR(1):

We obtained the following results when fitted the AR(1) model:

Type	Coef	SE Coef	T	P
AR 1	**0.4385**	0.1024	4.28	**0.000**
Constant	**0.28923**	0.01126	25.68	**0.000**
Mean	0.51514	0.02006		

Number of observations: 79
Residuals: SS = 0.771706 (backforecasts excluded)
MS = 0.010022 DF = 77

Modified Box-Pierce (**Ljung-Box**) Chi-Square statistic

Lag	**12**	**24**	**36**	**48**
Chi-Square	30.8	37.8	50.1	73.7
DF	10	22	34	46
P-Value	**0.001**	**0.019**	**0.037**	**0.006**

As we note from the table, the parameter estimates are significant (i.e. they differ significantly from zero), thus have to

be kept in the model. Looking at the p_value for the estimated model parameter $\hat{\phi}_1 = 0.4385$, which we use to test the hypothesis $H_0: \phi_1 = 0$ vs $H_1: \phi_1 \neq 0$, since the p_value equal to 0 (less than 5% or 1% whatever we used), then we reject H_0 and conclude that that ϕ_1 should be kept in the model. Now looking at the result of Ljung-Box statistic, which is used to test the hypothesis:

$$H_0: \rho_1 = \cdots = \rho_K = 0 \quad \text{vs} \quad H_1: at\ least\ two \neq 0$$

This hypothesis test that residuals of the fitted model up to time lag k are uncorrelated, hence in case we accept H_0 we will

deduce that the model is suitable to the data. But from the table above, we note that all p_values for any k are less than 5%, hence we reject H_0 and deduce that the model is not appropriate for modelling all the autocorrelation structure in the data. We can also plot the autocorrelation and partial autocorrelation functions of the residuals to check this point;

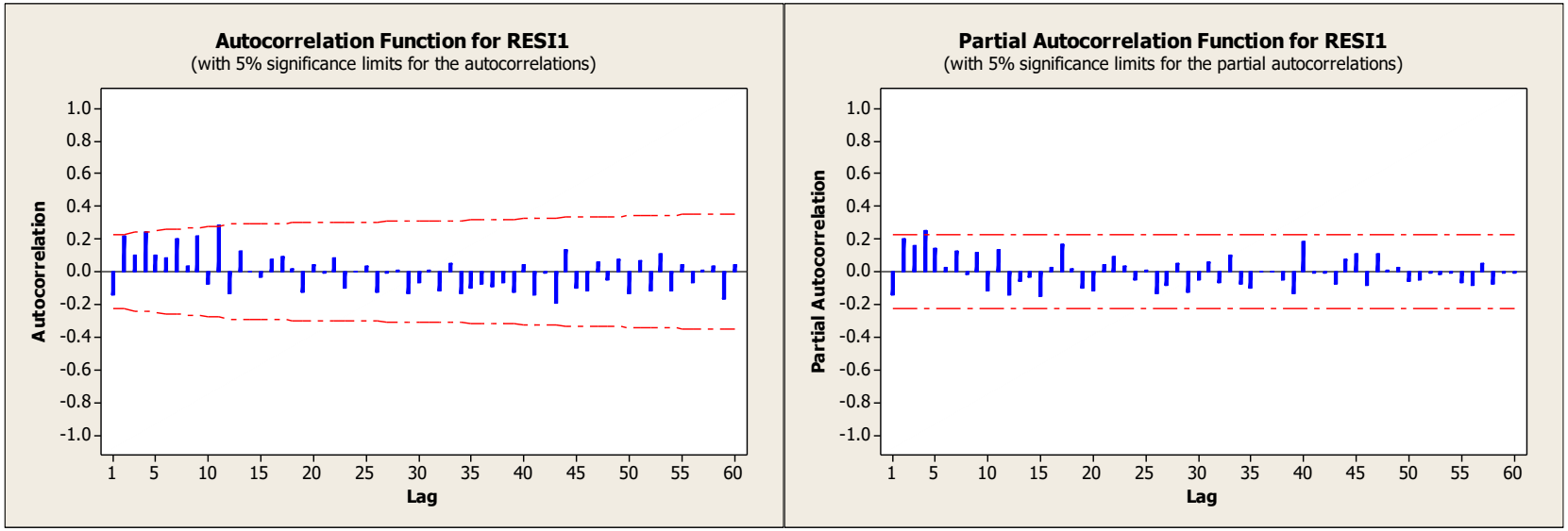


Figure (4): autocorrelation and partial autocorrelation functions of the residuals of AR(1) model

As we note from figure (4), the autocorrelation function shows that some autocorrelation in the residuals at lags k=2,4 still exists, which means that the model couldn't model them properly. The same comment for the partial autocorrelation

function, as it seems that some autocorrelation structure is still not accounted for by the AR(1) model, hence we move to the next proposed model.

(ii) Autoregressive model of order two AR(2):

We obtained the following results when fitted the AR(2) model:

```
Final Estimates of Parameters

Type          Coef  SE Coef      T      P
AR   1      0.2937   0.1083   2.71  0.008
AR   2      0.3312   0.1087   3.05  0.003
Constant   0.19334  0.01071  18.06  0.000
Mean       0.51535  0.02854

Number of observations:  79
Residuals:    SS =  0.688118 (backforecasts excluded)
              MS =  0.009054  DF = 76
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag                12      24      36      48
Chi-Square       11.6    22.5    32.7    51.1
DF                  9      21      33      45
P-Value         0.234   0.371   0.484   0.245
```

As we note from the above table, all parameters included in the model are significantly different from zero and hence have to be

retained in the model. Also, the p_values for testing the hypothesis $H_0: \rho_1 = \rho_2 = \cdots = \rho_K = 0$ are not significant for all values of k, hence we accept H_0 and deduce that the model is tentatively appropriate for the data. Plotting the autocorrelation and partial autocorrelation function for the residuals of the AR(2) model, we get:

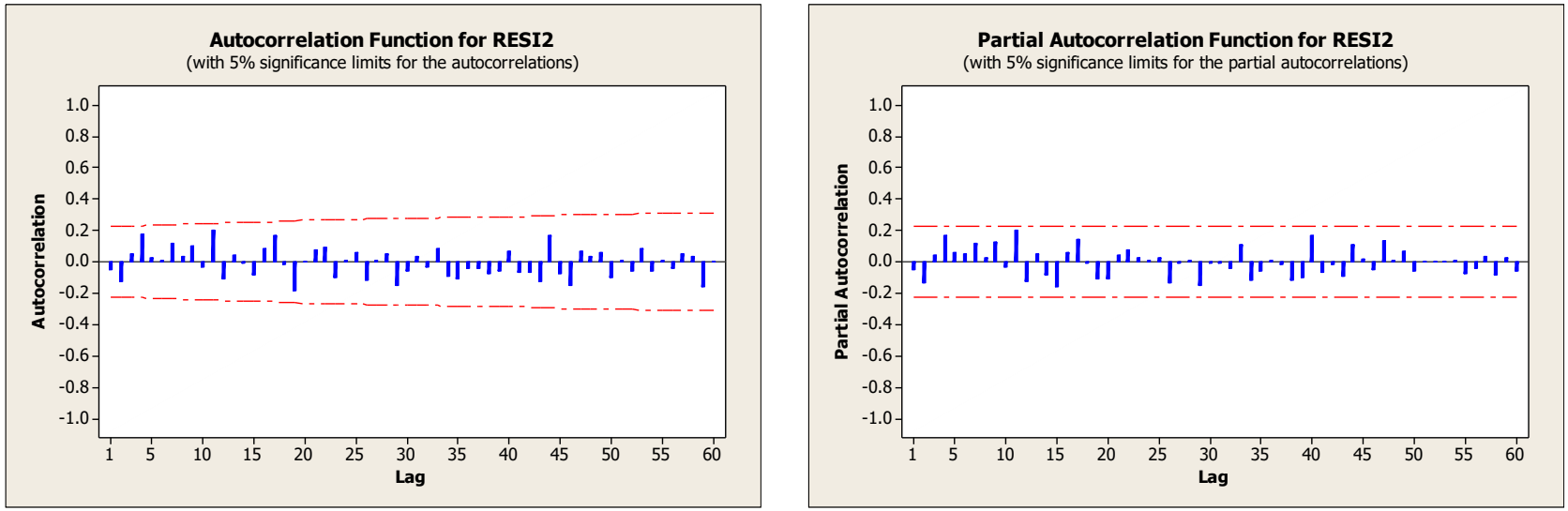


Figure (5): autocorrelation and partial autocorrelation functions of the residuals of AR(2) model

As we note from figure (5), the residuals of the AR(2) model are much better from those of AR(1) as they do not show any unexplained autocorrelation structure in the residuals.

Now, we have to perform the diagnostic checks to verify whether the model residuals fulfill the assumptions of the white noise process ε_t,

where as we know, $\hat{\varepsilon}_t$ are actually estimates for the terms of the white noise process. The following figure shows results of diagnostic checks of the model residuals:

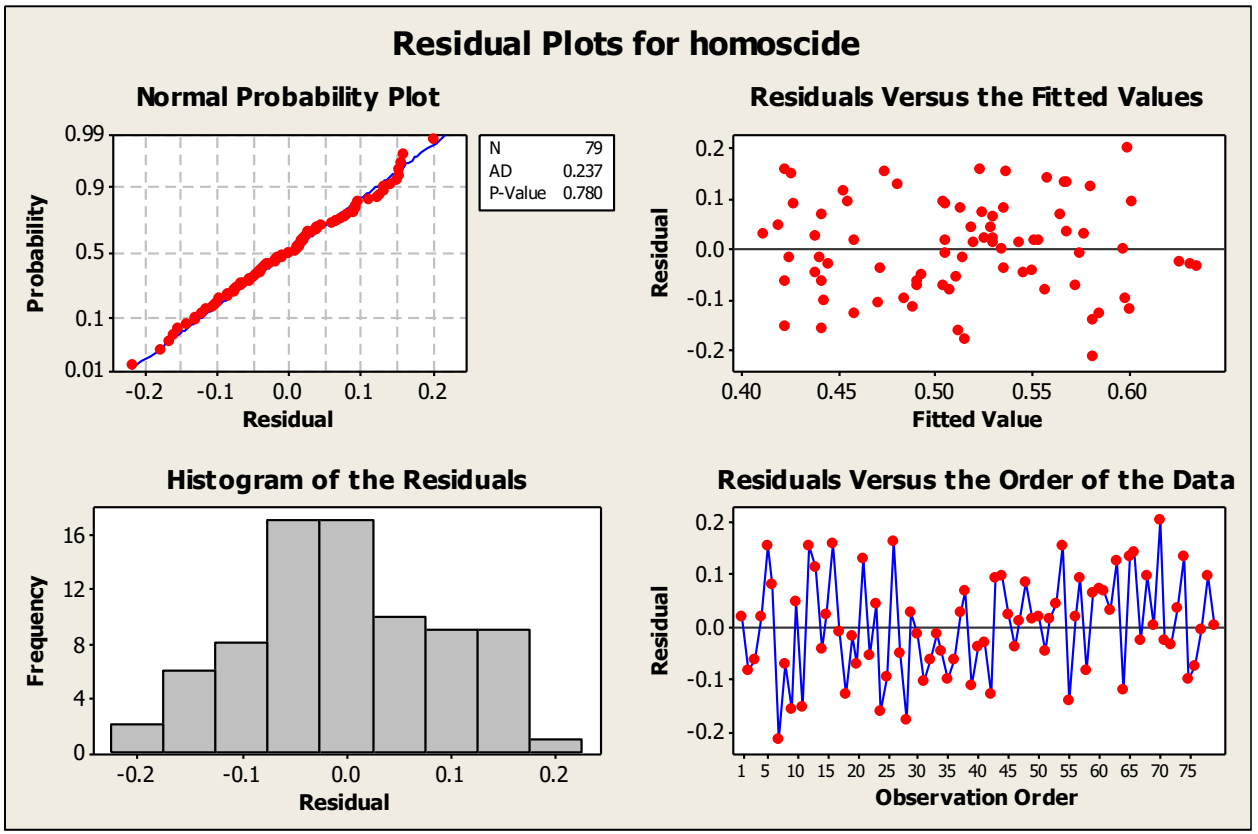


Figure (6): Diagnostic plots for the residuals of AR(2) model

a) Residuals follow the normal distribution:

Checking figure (6), we note that the probability plot shows percentiles of the residuals that agree to a high extent with those of the normal distribution, also the figure shows the result of applying a non-parametric test for goodness of fit, the Anderson-Darling test for the hypothesis:

$$H_0: residulas\ of\ the\ model\ follow\ Normal\ distribution$$

The P_value of the test is 0.780, which means the acceptance of H_0. Also, note that the histogram of the data resembles to a good extent the normal histogram.

b) Variance of the residuals is constant:

The plot at the top right hand side of figure (6), shows residuals against the estimated fitted values, which indicate that the variance is constant and does not change with time.

c) Mean of the residuals is zero:

We can conduct a t-test for testing the hypothesis that residuals mean is zero, the MINITAB output provide us with the following output:

One-Sample T: RESI3

Test of mu = 0 vs not = 0

Var	N	Mean	StDev	SE Mean	95% CI	T	P
RESI3	79	0.0006	0.093924	0.010567	(-0.020460, 0.021615)	0.05	**0.957**

Since the P_value of the test is 0.957, which means the acceptance of the zero mean hypothesis of the residuals.

d) Randomness of the residuals:

Using the Runs test, which is a non-parametric test for testing the hypothesis that the residuals are random versus that they are not random, the MINITAB provide us with the following results:

```
Runs test for RESI3

Runs above and below K = 0

The observed number of runs = 40
The expected number of runs = 40.4937
39 observations above K, 40 below
P-value = 0.911
```

Since the P_value of the test is 0.911, which means that we accept the hypothesis of the residuals randomness.

e) Residuals are uncorrelated:

We have already mentioned the result of the Ljung-Box test, which in fact is a test for the uncorrelation of the residuals, and we have accepted this hypothesis.

- Stationarity analysis:

The estimated values of the model parameters are:

$\hat{\phi}_1 = 0.2937.\ \hat{\phi}_2 = 0.3312$, and applying the stationarity conditions for this model:

(i) $|\boldsymbol{\phi_2}| < \mathbf{1} \Rightarrow |\mathbf{0.3312}| < \mathbf{1}$

(ii) $\boldsymbol{\phi_1} + \boldsymbol{\phi_2} < \mathbf{1} \Rightarrow \mathbf{0.2937} + \mathbf{0.3312} = \mathbf{0.6249} < \mathbf{1}$

(iii) $\boldsymbol{\phi_2 - \phi_1 < 1 \Rightarrow\ 0.3312 - 0.2937 = 0.0375 < 1}$

So the estimated parameters of the model satisfy the stationarity condition.

Hence, we note that the AR(2) model has passed all diagnostic checks, and thus we conclude that it is suitable to model rate number of homicide cases in Australia during 1915-1993, and the form of the model is:

$$Y_t = 0.19334 + 0.2937\, Y_{t-1} + 0.3312\, Y_{t-2} + \varepsilon_t$$

Where, Y_t is rate number of homicide cases at year t, and the variance of the white noise process ε_t is estimated as

MS = 0.009054.

(iii) As previously mentioned, ARMA(1,1) model was a tentative model for our data, thus we are going to fit it and see how good it is for modelling the rate number of homicide cases in Australia.

The results of applying this model to the data in MINITAB is shown in the following table:

```
Final Estimates of Parameters
```

```
Type          Coef    SE Coef       T       P
AR   1       0.9384     0.0603   15.56   0.000
MA   1       0.6959     0.1234    5.64   0.000
Constant     0.0317     0.00319   9.94   0.000
Mean         0.51579    0.05191

Number of observations:  79
Residuals:    SS =  0.649709 (backforecasts excluded)
              MS =  0.008549  DF = 76

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag             12     24     36     48
Chi-Square    10.0   17.9   26.5   42.1
DF               9     21     33     45
P-Value      0.348  0.654  0.782  0.594
```

As we note from the above table, all parameters included in the model are significantly different from zero and hence have to be

retained in the model, as all P_values of the parameters θ_1, ϕ_1 and the constant δ are all equal to zero. Also, the result of the Ljung-Box test indicate that the model is adequate for the data since all the p_values are larger than $\alpha = 0.05$. In addition, the parameter estimates fulfill the stationarity and invertibility conditions.

Model diagnostics:

The following figure shows results of diagnostic checks of the model residuals:

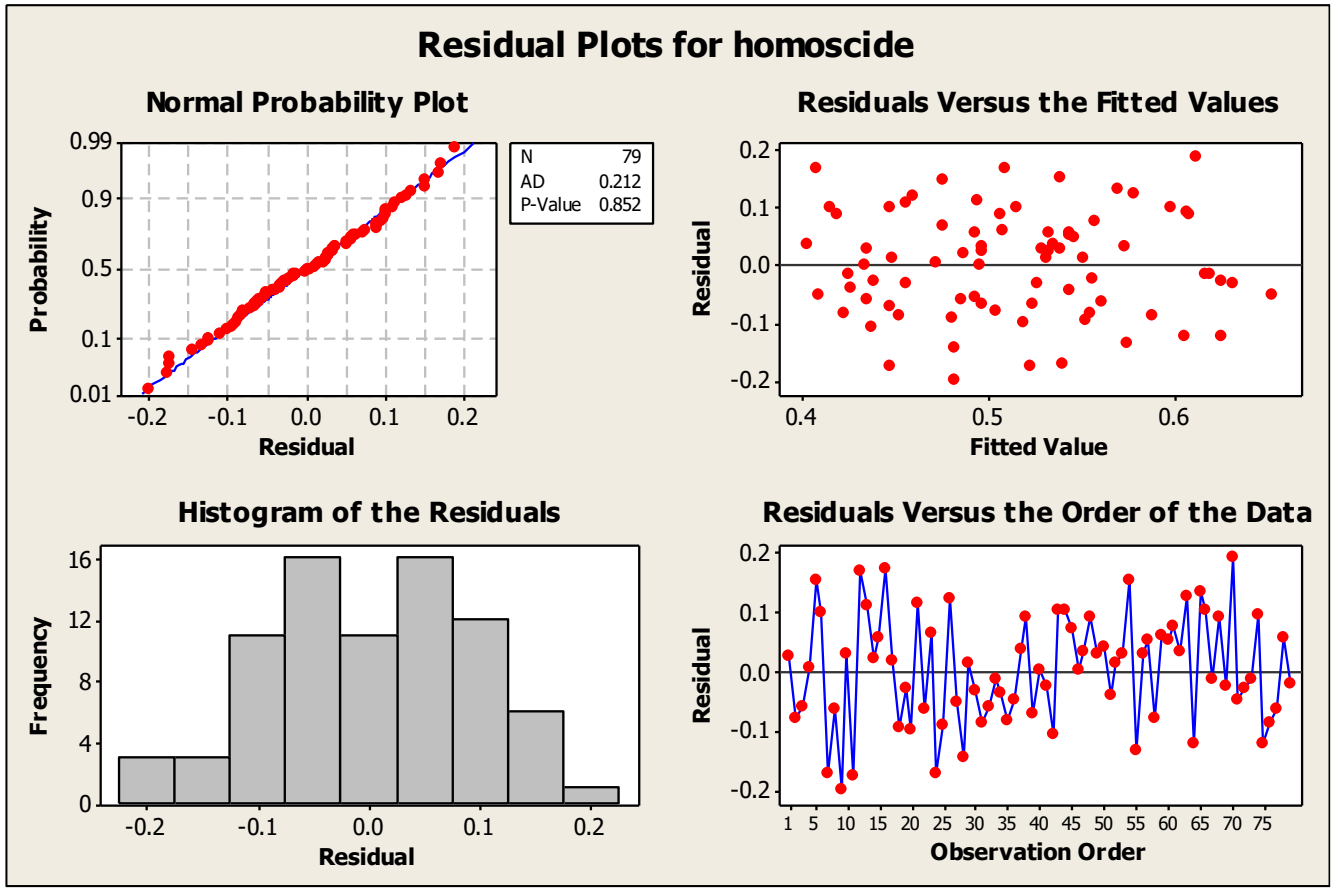


Figure (8): Diagnostic plots for the residuals of ARMA(1,1) model

From figure (8), it is evident that the residuals follow the normal distribution , and that their variance is constant and do not change with time. The rest of the diagnostic checks are as follow:

(a) Mean of the residuals is zero:

We can conduct a t-test for testing the hypothesis that residuals mean is zero, the MINITAB output provide us with the following output:

One-Sample T: RESI1

Test of mu = 0 vs not = 0

Var	N	Mean	StDev	SE Mean	95% CI	T	P
RES	79	0.001949	0.091246	0.010266	(-0.01848, 0.02238)	0.19	**0.850**

Since the P_value of the test is 0.850, which means the acceptance of the zero mean hypothesis of the residuals.

(b) Randomness of the residuals:

Using the Runs test, which is a non-parametric test for testing the hypothesis that the residuals are random versus that they are not random, the MINITAB provide us with the following results:

Runs Test: RESI1

```
Runs test for RESI1

Runs above and below K = 0

The observed number of runs = 38
The expected number of runs = 40.3418
42 observations above K, 37 below
P-value = 0.594
```

Since the P_value of the test is 0.594 which means that we accept the hypothesis of the residuals randomness.

(c) Residuals are uncorrelated:

We have already mentioned the result of the Ljung-Box test, which in fact is a test for the uncorrelation of the residuals, and we have accepted this hypothesis. Plotting the ACF and PACF for the residuals, we get:

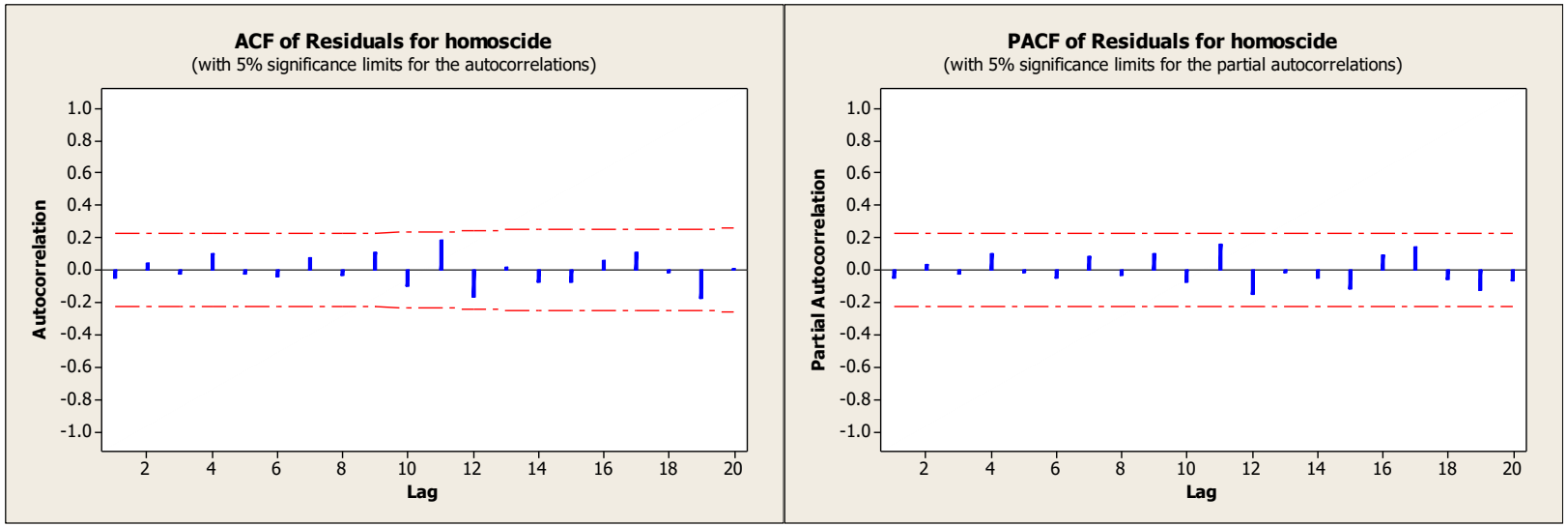


Figure (9): autocorrelation and partial autocorrelation functions of the residuals of ARMA(1,1) model

As we note, the ARMA(1,1) succeeded in modelling all the autocorrelation structure in the data.

Figure (10): autocorrelation and partial autocorrelation functions of the first differences of the residuals of ARMA(1,1) Model

Hence, we note that the ARMA(1,1) model has passed all diagnostic checks, and thus we conclude that it is suitable to model rate number of homicide cases in Australia during 1915-1993, and the form of the model is:

$$Y_t = 0.031771 + 0.9384\, Y_{t-1} + \varepsilon_t - 0.6959\, \varepsilon_{t-1}$$

Where, Y_t is rate number of homicide cases at year t, and the variance of the white noise process ε_t is estimated as MS = 0.008549.

Since we have proposed two models that can successfully model the correlation structure available in the data, hence we have to use some

comparison criteria to choose the best model of the two, from these criteria are:

a) Akaike information criterion (AIC):

This criterion is defined as: $AIC(m) = n \ln(\hat{\sigma}_\varepsilon^2) + 2m$

b) Bayesian information criterion (BIC):

This criterion is defined as: $BIC(m) = n \ln(\hat{\sigma}_\varepsilon^2) + 2m \ln(n)$

Where, m : number of estimated parameters

n: Is the number of available observations (if any differences are taken, then it is the total number of observations after the difference).

$\hat{\sigma}_{\varepsilon}^{2}$: is the estimated variance of the model residuals (or the estimated variance of the white noise process)

Now, we summarize the results in the following table:

Model	n	m	$\hat{\sigma}_{\varepsilon}^{2}$	AIC	BIC
AR(2)	79	2	0.009054	-367.659	-354.1816
ARMA(1,1)	79	2	0.008549	-372.193	-358.7155

Since the model to be selected is the one with the lowest value of the comparison criterion, thus we see from the table that both criterion select the ARMA(1,1) to model the homicide rate in Australia.

Using the model to forecast the homicide rate in Australia for the next five years:

The following figure shows the forecast the homicide rate in Australia for the next five years:

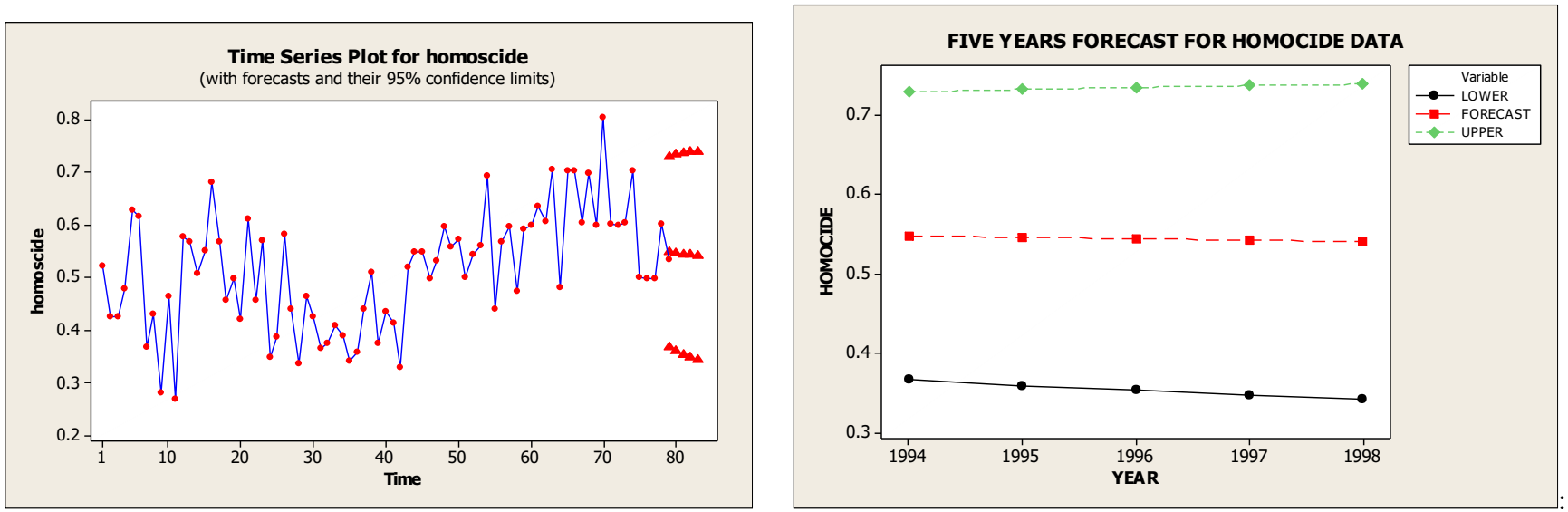


Figure (11): forecast the homicide rate in Australia for the next five years using the ARMA(1,1) Model

The following table shows these forecasts, together with a 95% C.I.:

Table (1): Forecasting homicide rate for every 100000 capita in Australia

for five years using ARMA(1,1) model

Year	***Lower limit***	***Forecast***	***Upper limit***
1994	0.366362	0.547620	0.728877
1995	0.359149	0.545659	0.732170
1996	0.352802	0.543819	0.734836
1997	0.347194	0.542093	0.736991
1998	0.342218	0.540472	0.738726

*Base year (1993) where homicide rate for every 100000 capita is 0.53395

As we note from these results, that we expect the homicide rate to increase in 1994 to 0.547620 for every 100000 capita, then the rate will start to decline in an average yearly rate of 0.30 %.

Chapter seven: Seasonal Models

As we have seen in the previous chapters, the stochastic time series models could successfully model the correlation structure in the data. However, in case the data show a seasonal behavior, then the model should incorporate a component that reflect such seasonality.

7.1 Autoregressive seasonal models

Assume for example that we have a quarterly time series, then we say that it follows a seasonal autoregressive model of order one if we can express the current value of the series y_t as a linear function

of the value of the series at the same season in the previous year y_{t-s} (here we assume $S = 4$) plus a random variable term ε_t, that is:

$$y_t = \Phi_1 y_{t-s} + \varepsilon_t$$

Where Φ_1 represent the seasonal autoregressive parameter, this model is denoted as SAR(1).

In the same manner, we can add seasonal autoregressive parameters to this model to get SAR(P), which can be expressed as:

$$\left(1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_p B^{Ps}\right) y_t = \varepsilon_t$$

or,

$$y_t = \Phi_1 y_{t-s} + \Phi_2 y_{t-2s} + \cdots + \Phi_P y_{t-Ps} + \varepsilon_t$$

It can be proven that the autocorrelation function for the seasonal autoregressive model is very much similar to the ACF of the usual autoregressive model, except that the autocorrelation coefficients appear at multiples of S, i.e. at the multiples of the seasonal period. For example, for the SAR(1) model, with a positive parameter Φ_1, and seasonal period length $s = 4$, then the autocorrelation coefficients will appear at multiples of the number 4, and will gradually decline to zero, see figure (7.1).

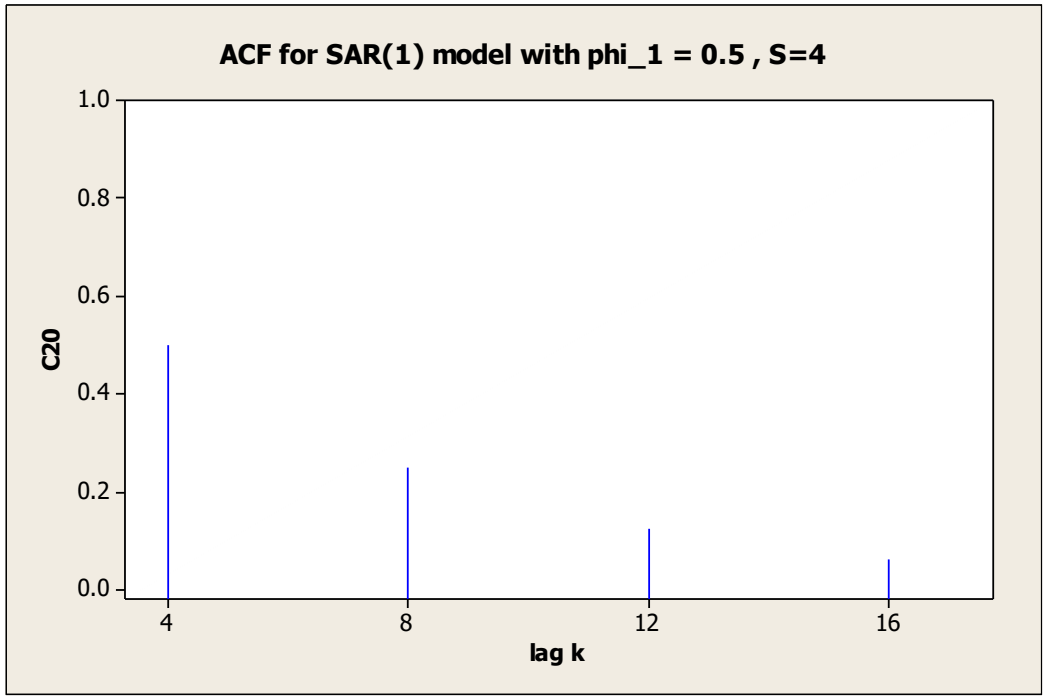


Figure (7.1): autocorrelation function for SAR(1), s=4

If we have SAR(1) model with seasonal period S=12, then the autocorrelation coefficients will appear at multiples of the number 12 (i.e. at 12, 24, 36, 48, …) .

7.2 Moving average seasonal models

we say that a stationary time series follows a seasonal moving average model of order one if we can express the current value of the series y_t as a linear function of the value of the random shock that occurred at current time ε_t and the one occurred at the same season in the previous year ε_{t-s} that is:

$$y_t = \varepsilon_t - \Theta_1 \varepsilon_{t-s}$$

It could also be written as:

$$y_t = (1 - \Theta_1 B^s)\varepsilon_t$$

Where Θ_1 represent the seasonal moving average parameter, this model is denoted as SMA(1). In the same manner, we can add seasonal moving

average parameters to this model to get SMA(Q), which can be expressed as:

$$y_t = \left(1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs}\right)\varepsilon_t$$

or,

$$y_t = \varepsilon_t - \Theta_1 \varepsilon_{t-s} - \Theta_2 \varepsilon_{t-2s} - \cdots - \Theta_Q \varepsilon_{t-Qs}$$

It can be proven that the autocorrelation function for the seasonal moving average model is very much similar the ACF of the usual moving average model, except that the autocorrelation coefficients appear at multiples of S, i.e. at the multiples of the seasonal period. For example, for the SMA(1) model, then there is only one non-zero autocorrelation value that occur at a time lag that is equal to seasonal

period, , see figure (7.2). For SMA(Q) models the number of non-negative autocorrelation coefficients will appear at multiples of S.

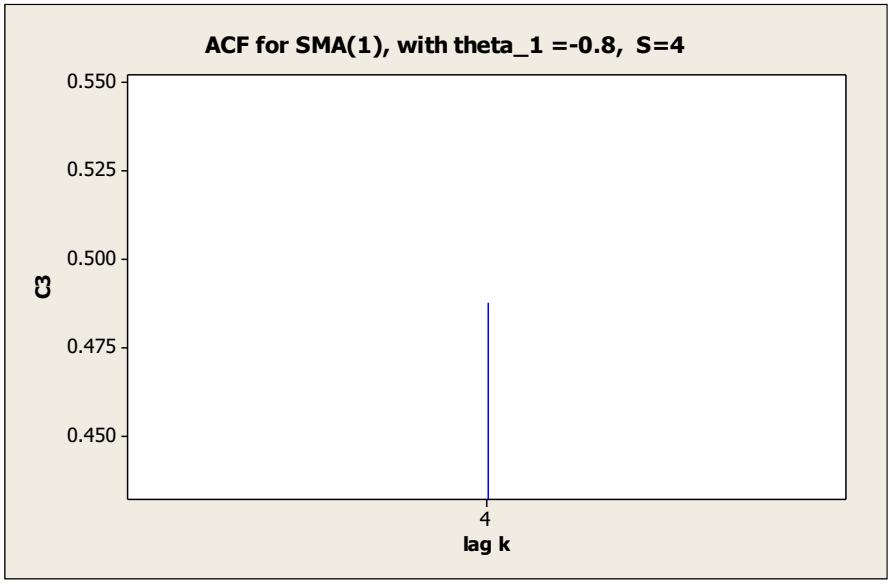


Figure (7.2): autocorrelation function for SMA(1)

7.3 Autoregressive Moving average seasonal models

It is possible to combine both the autoregressive and models in one group, such models are expressed as:

$$\Phi(B^s)y_t = \Theta(B^s)\varepsilon_t$$

Where,

$$\Phi(B^s) = \left(1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_p B^{Ps}\right)$$

$$\Theta(B^s) = \left(1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_q B^{Qs}\right)$$

And the symbol used to denote such models is SARMA(P,Q). In case the series was not stationary, then it is possible to apply the differences operator to the series as follows:

$$\Phi(B^s)\ \nabla_S^D\ y_t = \Theta(B^s)\varepsilon_t$$

Where ∇_S^D represent the seasonal differences at the seasonal period S, in this case we have the model SARIMA(P,D,Q), where,

P: the order of the seasonal autoregressive model

Q: the order of the seasonal moving average model

D: number of seasonal differences to render the series to be stationary at seasonal periods S.

It is possible as well to get a general form of the Box-Jenkins models that incorporate both normal and seasonal terms, and it is sometimes called "General multiplicative Box-Jenkins models":

$$\phi(B)\Phi(B^s)\,\nabla^d\,\nabla_s^D\,y_t = \theta(B)\,\Theta(B^s)\varepsilon_t$$

it is abbreviated as,

$$\boxed{\text{SARIMA(p, d, q)(P, D, Q)s}}$$

Example:

Write the mathematical formula for the model SARIMA(0,0,1)(0,0,1)4.

Solution:

We have the following values for the order indexes, q=1 ،Q=1 ،S=4, thus the model form is:

$$y_t = (1 - \theta_1 B)\,(1 - \Theta_1 B^4)\,\varepsilon_t$$

$$\Rightarrow y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \Theta_1 \varepsilon_{t-4} + \theta_1 \Theta_1 \varepsilon_{t-5}$$

7.3.1 some characteristics of the general multiplicative models

There are in fact very few general characteristics for the ACF and PACF functions that could be used to identify the multiplicative seasonal models. Table (7.1) shows some basic characteristics for ACF and

PACF for some multiplicative seasonal models, which are used to try to see if a specific multiplicative seasonal could be used to model the data.

Model	ρ_k	ϕ_{kk}
SARIMA(p,0,0)(P,0,0) ≡ SAR(p, P)	Approach zero gradually	Cut off completely after the time lag $p+sP$
SARIMA(0,0,q)(0,0,Q)	Cut off completely after the	Approach zero gradually

≡ SMA(q, Q)	time lag q+sQ	
SARIMA(p,0,q)(P,0,Q)	Approach zero gradually	Approach zero gradually

7.4 Example:

Data in table (7.2) represent amount of monthly produced electrical energy in the United States during the period of Jan. 1985 –Dec. 2014. Study this set of data, try to get a suitable mathematical model able to model it. Use your chosen model to forecast the amount of monthly

produced electrical energy for the year 2015. The actual monthly production for 2015 is shown below:

Month	*Amount of produced electricity*
1	399.96
2	400.26
3	401.52
4	403.26
5	403.94
6	402.80
7	401.30
8	398.93

9	397.63
10	398.29
11	400.16
12	401.85

Table (7.2): amount of monthly produced electrical energy in the United States during the period of Jan. 1985 –Dec. 2014

12	11	10	9	8	7	6	5	4	3	2	1	*Month/year*
345.82	344.4	343.08	343.2	344.85	346.65	348.4	348.92	348.2	347.66	346.06	345.25	1985
347.15	345.86	344.47	345.01	346.09	348.11	349.9	350.53	349.77	348.05	347.13	346.54	1986
349.18	347.96	346.65	346.52	347.84	349.9	351.61	352.14	351.32	349.72	348.7	348.38	1987
351.44	350.15	349.08	349.03	350.66	352.58	353.68	354.18	353.66	352.24	351.68	350.38	1988
352.84	351.44	350.29	350.02	351.53	353.98	355.3	355.89	355.59	353.8	353.24	352.89	1989
354.27	353.05	351.59	351.28	352.89	354.88	356.32	357.29	356.28	355.65	354.88	353.79	1990
355.07	353.79	352.32	352.3	353.89	356.12	358.1	359.09	358.51	357.06	355.68	354.87	1991
355.53	354.27	353.31	352.93	354.91	356.85	359.32	359.55	359	357.82	356.93	356.17	1992
356.84	355.4	354.12	354.1	355.46	357.42	359.52	360.19	359.27	358.36	357.27	356.86	1993
358.87	357.56	356.09	355.63	357.42	359.39	360.8	361.68	361.32	359.91	358.98	358.22	1994
360.61	359.4	357.97	358.11	359.11	361.7	363.22	363.77	363.23	361.77	360.79	359.87	1995
362.18	360.84	359.54	359.6	361.38	363.53	364.93	365.16	364.51	364.17	363.17	362.04	1996
364.33	362.45	360.71	360.31	362.2	364.34	365.59	366.69	366.25	364.47	364.09	363.04	1997
367.08	365.52	364.35	364.01	365.79	367.74	368.95	369.49	368.61	367.13	365.98	365.18	1998

368.04	366.68	365.35	364.94	366.86	369.28	370.33	370.77	370.96	369.6	368.98	368.12	1999
369.67	368.33	366.99	366.91	368.2	369.84	371.71	371.51	371.82	370.56	369.5	369.25	2000
371.18	369.69	368.42	368.16	369.63	371.57	373.18	373.82	373.37	372.53	371.49	370.52	2001
373.71	372.2	370.51	370.66	371.83	374	375.5	375.65	375	373.94	373.14	372.45	2002
375.93	374.64	373.1	373.2	374.31	376.72	378.18	378.5	377.74	376.48	375.62	374.87	2003
377.45	375.93	374.44	374.11	376.15	377.61	379.56	380.63	380.41	378.73	377.87	377	2004
379.92	378.29	376.98	376.66	378.73	380.78	382.2	382.47	382.2	381.14	379.76	378.47	2005
381.79	380.18	379.16	378.92	380.45	382.38	384.09	384.98	384.73	382.66	382.16	381.35	2006
383.89	382.42	381.14	380.9	382	384.49	386.05	386.58	386.4	384.56	383.81	382.93	2007
385.56	384.13	382.99	383.09	384.15	386.43	387.88	388.5	387.16	385.97	385.73	385.44	2008
387.31	386	384.39	384.79	385.92	387.78	389.45	390.19	389.44	388.77	387.42	386.94	2009
389.73	388.65	387.2	386.83	388.26	390.22	392.15	393.04	392.52	391.09	389.94	388.5	2010
391.83	390.24	388.96	389.04	390.19	392.42	393.72	394.21	393.34	392.49	391.82	391.24	2011
394.28	392.81	391.01	391.06	392.41	394.3	395.82	396.78	396.18	394.45	393.6	393.12	2012
396.81	395.11	393.66	393.51	395.15	397.2	398.58	399.76	398.35	397.31	396.8	395.54	2013
398.84	397.27	395.95	395.35	397.1	399.04	401.2	401.88	401.34	399.62	397.93	397.81	2014

Solution:

We start the analysis by plotting the time series for the amount of monthly produced electrical energy. Figure (7.3) shows this time series, it is evident that there is an upward trend in the total produced electricity. Also, we note the clear seasonal component, beside that we do not notice any change in variation of production over the years, so we do not need to use any transformation to stabilize the data variance.

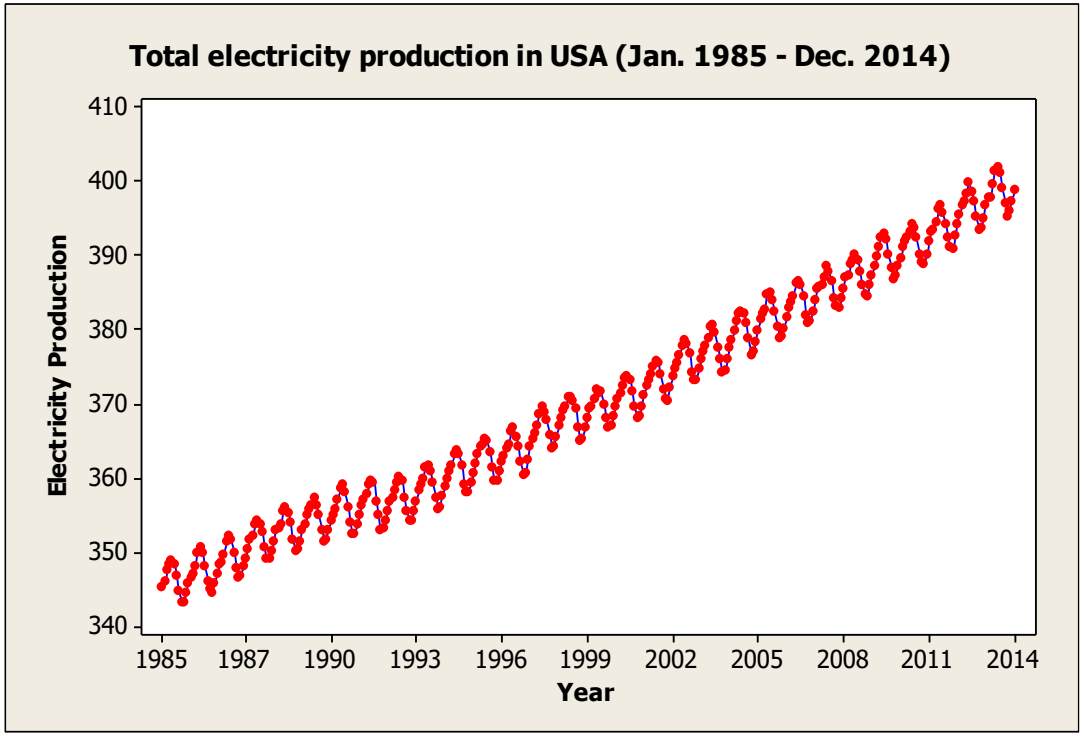


Figure 7.3: Monthly produced electrical energy in USA during1985 –2014

Surely, we will need to apply the differences operator to make the series stationary in the mean, also it is possible that we might need to take a

seasonal difference of order 12 if the series is not stationary at the seasonal periods, this will be apparent when we plot the autocorrelation and partial autocorrelation functions.

Identification:

As mentioned above we need to take the ordinary differences of order 1, i.e. $z_t = \nabla y_t$. We got the following ACF and PACF functions for z_t:

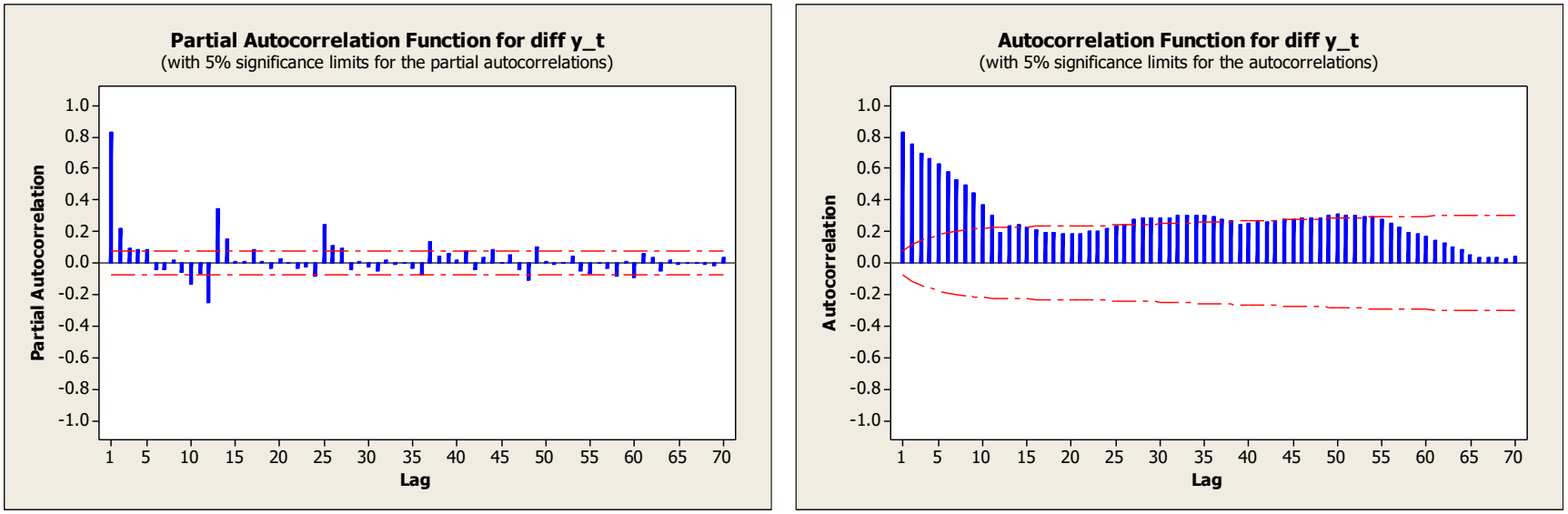


Figure 7.4: Autocorrelation and partial autocorrelation functions for the series $z_t = \nabla y_t$

Inspection of the estimated functions, we note that the autocorrelation function decay very slowly to zero, also that the first partial autocorrelation coefficient (0.83) is very large, this indicate that we might need to apply a second difference to the series. Also, we notice

that PACF coefficients at the seasonal periods (12, 24, 36,...) decay slowly, which again would indicate the need to take a seasonal difference at period s=12.

Figure (7.4) nominate an initial model SARIMA(2,1,0)(0,0,1)12 , also following the notes in the previous paragraph, we applied a second difference to the data, i.e. $w_t = \nabla^2 y_t$, and obtained the following ACF and PACF:

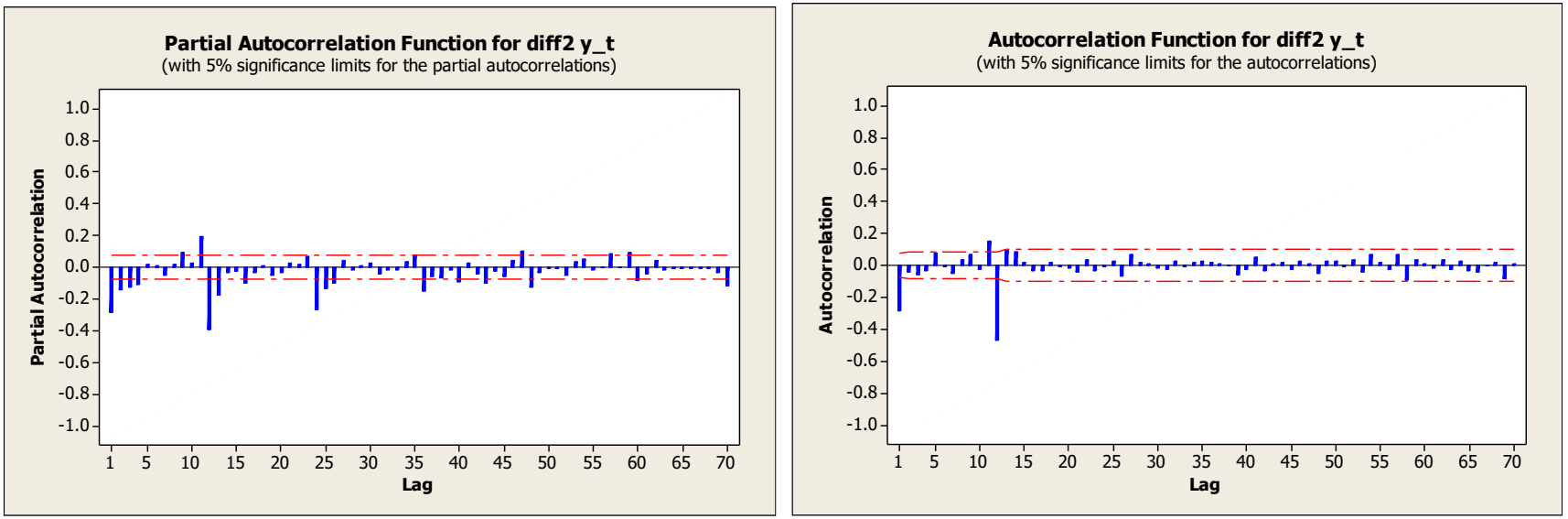


Figure 7.5: Autocorrelation and partial autocorrelation functions for the series $w_t = \nabla^2 y_t$

Inspection of the estimated functions in fig. (7.5), we note that the autocorrelation function cuts off after the first time lag, besides that, the partial autocorrelation function decay in an exponential format, this is

an indication that the data might follow a moving average of order one pattern. Also, that PACF coefficients at the seasonal periods (12, 24, 36,...) decay exponentially, and there exist a single significant value at the seasonal period s=12, which again would indicate that the data might follow a seasonal moving average of order one pattern . Thus Figure (7.5) nominate the model SARIMA(0,2,1)(0,0,1)12 .

Fitting the tentative models:

i) The model SARIMA(2,1,0)(1,0,0)12

Fitting the model with MINITAB, we got the following output:

Type		Coef	SE Coef	T	P
AR	**1**	**-0.3175**	0.0385	-8.25	**0.000**
AR	**2**	**-0.1331**	0.0384	-3.46	**0.001**
SAR	**12**	**0.9807**	0.0096	102.04	**0.000**

Differencing: 1 regular difference
Number of observations: Original series 672, after differencing 671
Residuals: SS = 103.064 (backforecasts excluded)
MS = 0.154 DF = 668

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	**12**	**24**	**36**	**48**
Chi-Square	172.2	183.6	194.0	202.7
DF	9	21	33	45
P-Value	**0.000**	**0.000**	**0.000**	**0.000**

As we can see from the output, all the model parameters are significantly different from zero, hence have to be retained in the model. However, when looking at the result of the Ljnug-Box statistic, that is used to test the hypothesis:

$$H_0: \rho_1 = \cdots = \rho_K = 0$$

$$H_1: at\ least\ two\ do\ not\ equal\ zero$$

This hypothesis tests the assertion that residuals of the model up to time lag k are uncorrelated, hence, upon accepting H_0 we will deduce that the model is suitable to that data. However, from the output above we notice that the P_values for the Ljung-Box test are all equal to zero, thus we reject H_0, and deduce that the model SARIMA(0,2,1)(1,0,0)12 could not capture all the autocorrelation structure of the data and thus it is unsuitable to model the data. We can also, plot the ACF and the PACF for the residuals of this model to check this point:

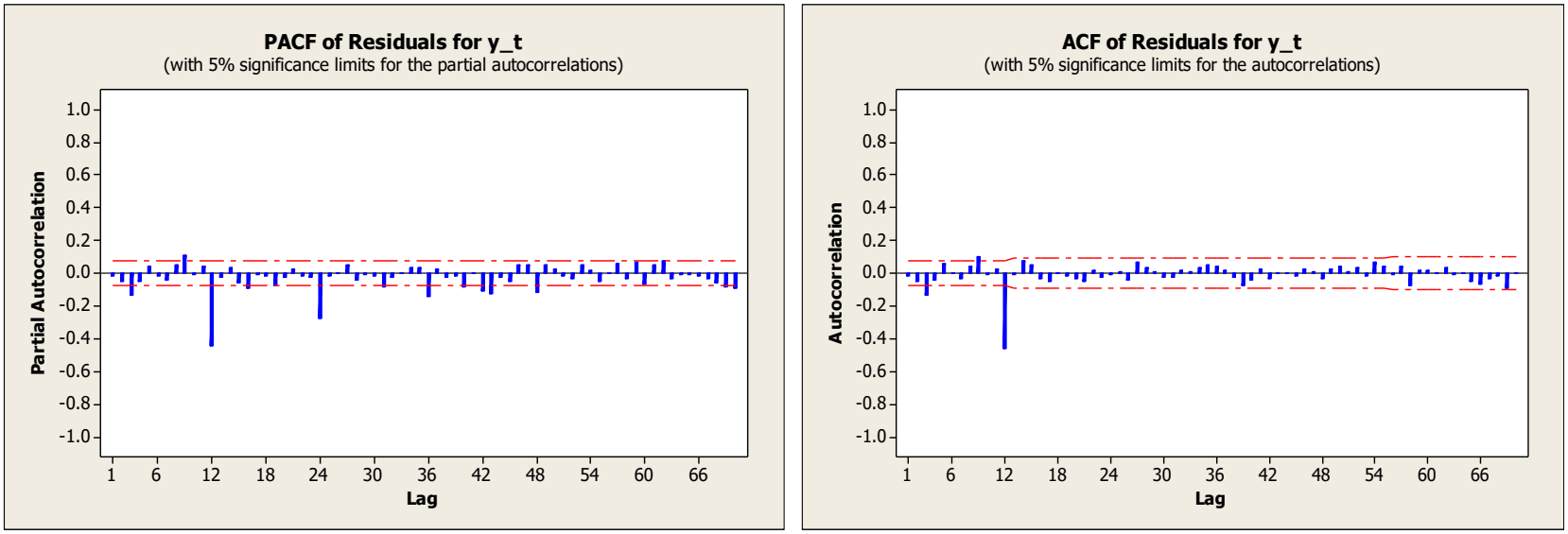


Figure 7.6: Autocorrelation and partial autocorrelation functions for the residuals of the model

$$SARIMA(0,2,1)(1,0,0)_{12}$$

We notice figure (7.6) that there is still some autocorrelation between the residuals of the model at time lag S=12 not explained by the model, also the PACF at time lags k=12, 24, 36 decay in an exponential

fashion. Hence, we search for another model that can model the data better.

ii) The model SARIMA(0,2,1)(0,0,1)12 :

Fitting this model using MINITAB, we got the following:

```
Type          Coef     SE Coef      T      P
MA    1     0.0142      0.0415   0.34  0.733
SMA   12   -0.5200      0.0364  -14.29  0.000


Differencing: 2 regular differences
Number of observations:  Original series 672, after
differencing 670
Residuals:    SS =  386.630 (backforecasts excluded)
```

```
                MS =   0.579   DF = 668


Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag                 12      24      36      48
Chi-Square      322.0   850.2  1207.7  1616.3
DF                 10      22      34      46
P-Value         0.000   0.000   0.000   0.000
```

We see notice that the moving average parameter in the non-seasonal part does not significantly differ from zero, thus it has to be removed from the model, also, we notice that all the P_values of the Ljung-Box test indicate that the model is not adequate in modelling the data, this

means that it could not model the correlation structure of the data. We, can also plot the ACF and the PACF for the model residuals to check upon this point:

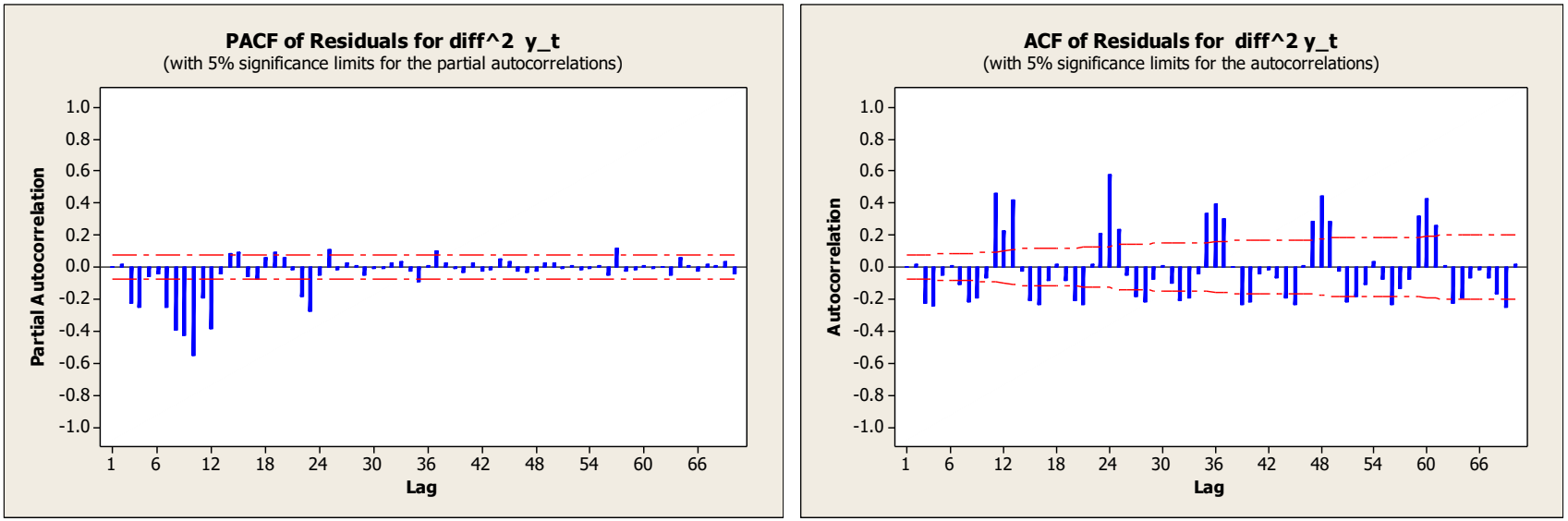


Figure 7.7: Autocorrelation and partial autocorrelation functions for the residuals of the model $SARIMA(0,2,1)(0,0,1)_{12}$

We notice from figure (7.7), the ACF of the residuals, there are still some high values of the autocorrelation coefficients at lags s=12,24,36,… . The same could be realized from the PACF at seasonal and non-seasonal lags. So, we deduce that the model could not model the correlation structure in the data properly. Hence, we search for another model that can model the data better.

The pattern revealed at Figure (7.7) indicate that we should take a seasonal difference to the data. So, we propose the model

SARIMA(0,1,1)(0,1,1)12. Notice that we have removed the regular difference of order 2, this is because taking many (unnecessary) differences might distort the autocorrelation structure of the data, and when we decided to take a seasonal difference, this might relieve us from taking the second regular difference, we will study this model and see if it can convince us in modelling the data properly.

iii) The model SARIMA(0,1,1)(0,1,1)12 :

Fitting this model using MINITAB, we got the following:

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
MA	**1**	**0.3726**	0.0366	10.18	**0.000**
SMA	**12**	**0.8929**	0.0176	50.66	**0.000**

Differencing: 1 regular, 1 seasonal of order 12

Number of observations: Original series 672, after differencing 659

Residuals: SS = 56.8926 (backforecasts excluded)
MS = 0.0866 DF = 657

Modified Box-Pierce (**Ljung-Box**) Chi-Square statistic

Lag	**12**	**24**	**36**	**48**
Chi-Square	11.4	24.3	36.4	50.0
DF	10	22	34	46

P-Value	0.329	0.333	0.359	0.317

As we can see from the output, all the model parameters are significantly different from zero, hence have to be retained in the model. Also, the Ljnug-Box statistic, shows that all the P_values are greater than $\alpha = 0.05$, hence we accept the hypothesis $H_0: \rho_1 = \cdots = \rho_K = 0$, and deduce that the model is suitable for the data, since it could model all the observed autocorrelation structure in the data. Inspecting the ACF and PACF for the residuals of the model:

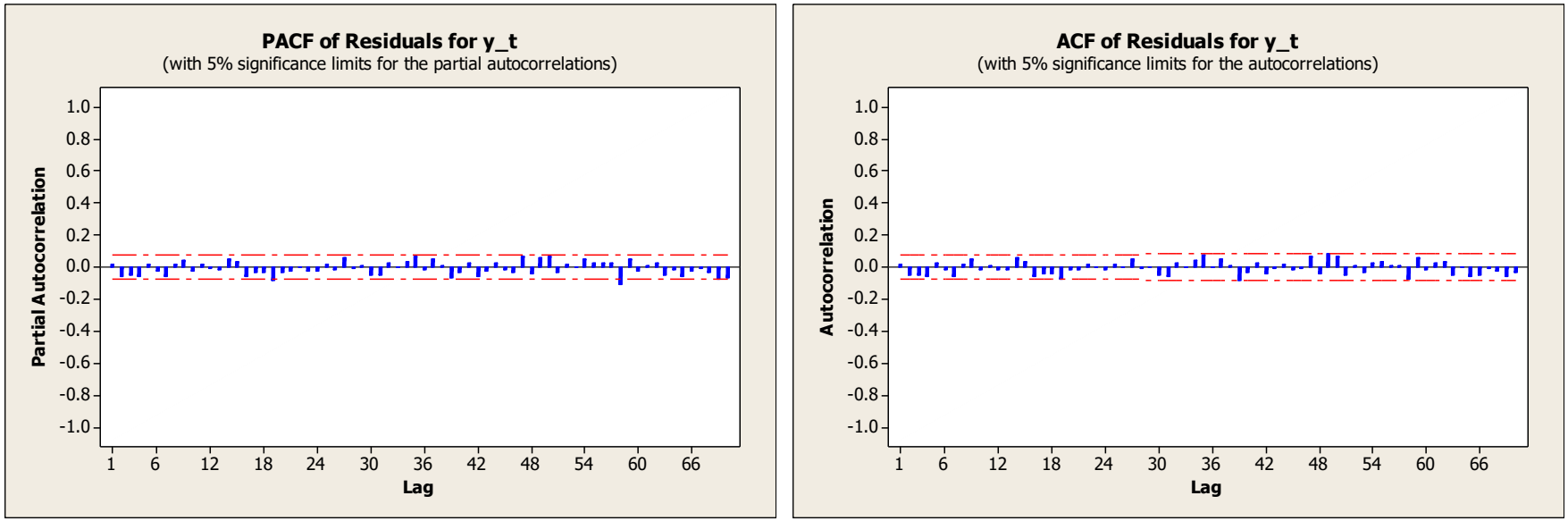


Figure 7.8: Autocorrelation and partial autocorrelation functions for the residuals of the model $SARIMA(0,1,1)(0,1,1)_{12}$

Which indeed indicate that the model is adequate, and that it could model all the autocorrelation in the data. The residuals of the model

show that it is an estimate of a white noise process, since all autocorrelation and partial autocorrelation coefficients do not significantly differ from zero, (which is a property of the white noise process).

Diagnostics:

Now, we have to perform diagnostic tests to see how these model residuals fulfill the conditions of the white noise process ε_t, because the model residuals $\hat{\varepsilon}_t$ are actually estimates of the white noise termsε_t.

The following figure shows results of diagnostic checks of the model residuals:

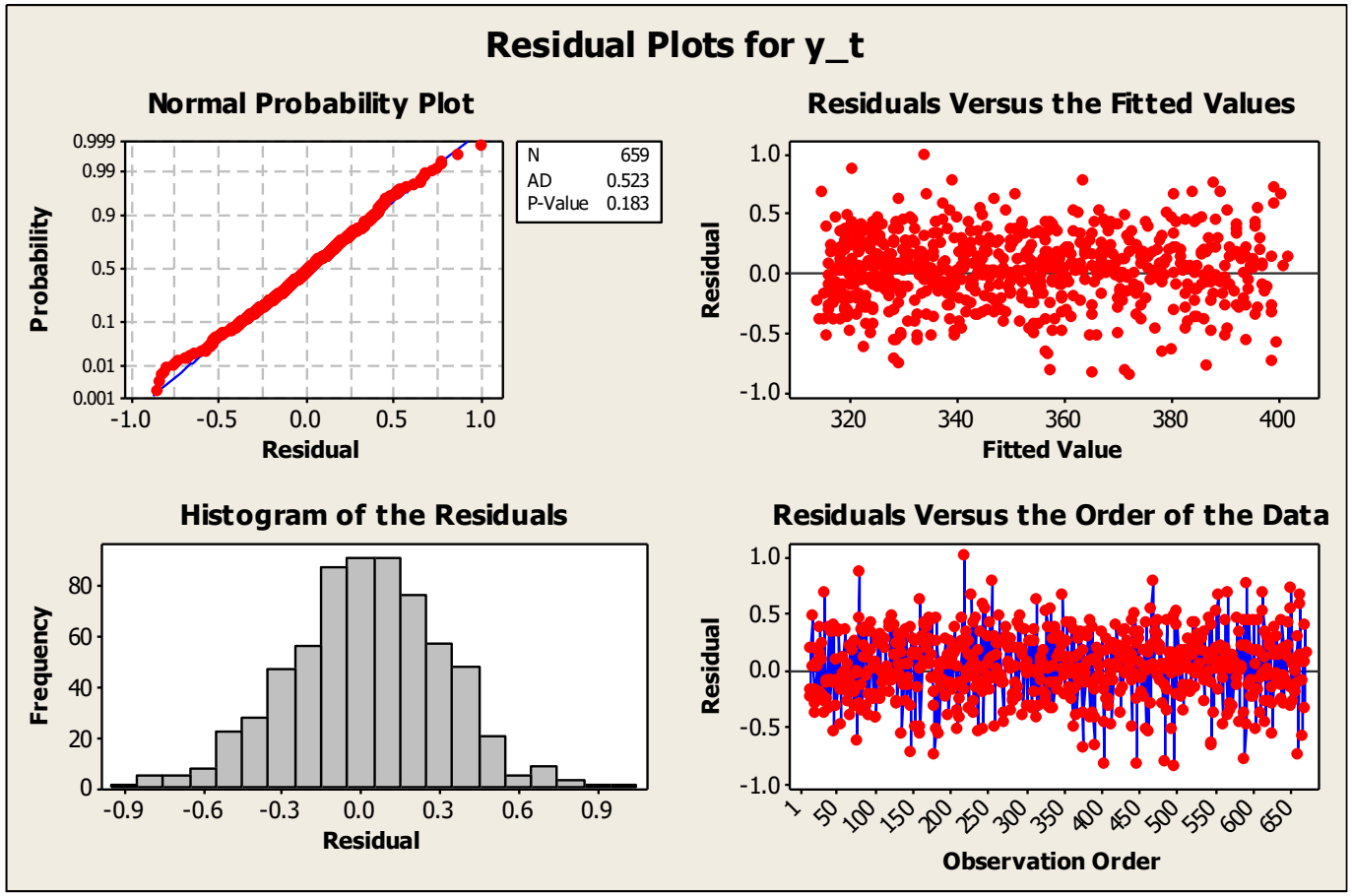


Figure (7.9): Diagnostic plots for the residuals of SARIMA(0,1,1)(0,1,1)12model

a) Residuals follow the normal distribution

From figure (7.9), the normal probability plot, shows that the percentiles lie on a straight line, which indicate that the residual percentiles agrees to a large extent with those of the normal distribution. The figure also show the result of a nonparametric goodness of fit test with the normal distribution, it is the Anderson-Darling (AD) test for testing the hypothesis:

H_0:residuals follow the normal distribution

The P_value is 0.183, which indicate that we accept H_0, also note that the histogram of the residuals takes a shape very similar to the normal distribution.

b) Variance of the residuals is constant:

The plot at the top right side of the figure indicate that the variance of the residuals does not change over time.

c) Mean of the residuals is zero:

We can conduct a t-test for testing the hypothesis that residuals mean is zero, the MINITAB output provide us with the following output:

One-Sample T: RESI

Test of mu = 0 vs not = 0

Var	N	Mean	StDev	SE Mean	95% CI	T	P
RESI	659	0.024	0.2931	0.0114	(0.001375, 0.046210)	2.08	**0.038**

Since the P_value of the test is 0.038, thus we reject the zero mean hypothesis of the residuals, also note that the 95% CI for the residual

mean does not contain zero, thus we conclude that we have to make amendments to the model. Let us include a constant term δ to the model, Doing so, we obtained the following output:

```
Final Estimates of Parameters

Type            Coef    SE Coef       T      P
MA   1        0.3958     0.0500    7.91  0.000
SMA  12       0.9448     0.0291   32.46  0.000
Constant  0.0025838  0.0009135    2.83  0.005


Differencing: 1 regular, 1 seasonal of order 12
Number of observations:  Original series 348, after differencing 335
Residuals:     SS =  27.2551 (backforecasts excluded)
               MS =  0.0821  DF = 332


Modified Box-Pierce (Ljung-Box) Chi-Square statistic
```

Lag	12	24	36	48
Chi-Square	8.0	22.8	28.6	37.9
DF	9	21	33	45
P-Value	0.536	0.355	0.688	0.764

From the above output, we notice that all the results indicate that the model is appropriate, and that constant parameter δ should also be retained in the model. Now, let us perform again the test that residuals mean is zero:

One-Sample T: RESI1

```
Test of mu = 0 vs not = 0


Var   N      Mean     StDev   SE Mean         95% CI         T     P
RES   659  -0.0048   0.2857   0.0156   (-0.0356, 0.0258)-0.31 0.755
```

Since the P_value is 0.755, so we accept the hypothesis of zero mean for the residuals.

a) Randomness of the residuals:

Using the Runs test, which is a non-parametric test for testing the hypothesis that the residuals are random versus that they are not random, the MINITAB provide us with the following results:

Runs Test: RESI1

```
Runs test for RESI1

Runs above and below K = 0

The observed number of runs = 174
The expected number of runs = 168.487
166 observations above K, 169 below
P-value = 0.546
```

Since the P_value of the test is 0.546 which means that we accept the hypothesis of the residuals randomness.

1- Residuals are uncorrelated:

We have already mentioned the result of the Ljung-Box test, which in fact is a test for the uncorrelation of the residuals, and we have accepted this hypothesis.

2- Stationarity analysis:

Since the model contains only moving average terms, then it is stationary.

3- Invertibility analysis:

The estimated parameters are $\hat{\theta}_1 = 0.3958$. $\hat{\Theta}_1 = 0.9448$, thus we see that the invertibility conditions are satisfied:

$$|\hat{\theta}_1| < 1 \Longrightarrow |0.3958| < 1 \ , |\hat{\Theta}_1| < 1 \Longrightarrow |0.9448| < 1$$

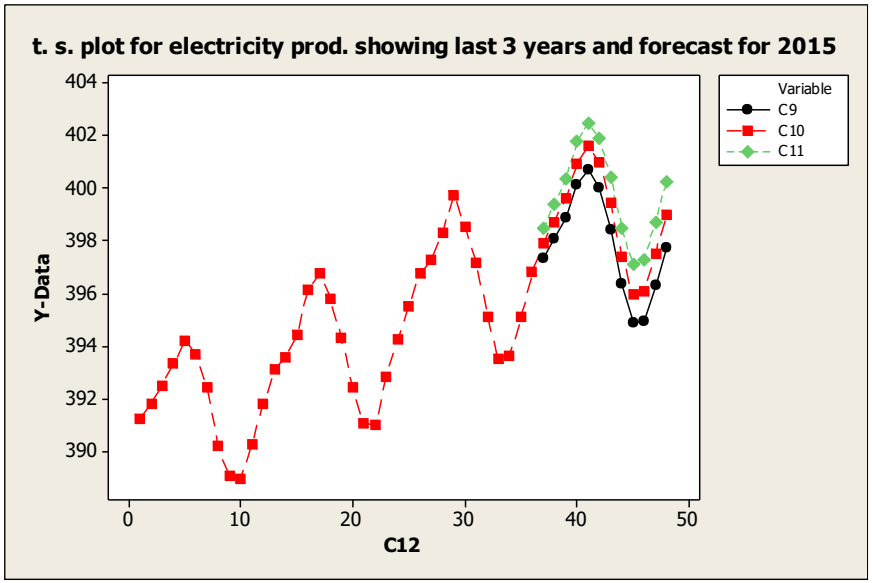
Thus the model SARIMA(0,1,1)(0,1,1)12 has passed all diagnostics tests, hence, it is suitable to model generated electricity amounts in USA during Jan. 1985 till Dec. 2014, and it has the form:

$$y_t = 0.00227 + y_{t-1} + \varepsilon_t - 0.3823\varepsilon_{t-1} - 0.9058\varepsilon_{t-12} + 0.3463\varepsilon_{t-13}$$

Where y_t represent generated electricity amounts in month t, and the white noise estimated variance is $MS = 0.0858$.

- Using the model to forecast the generated electricity amount for the next 12 months:

  The following figure shows the forecasts for the 2015 together with the observed actual values:



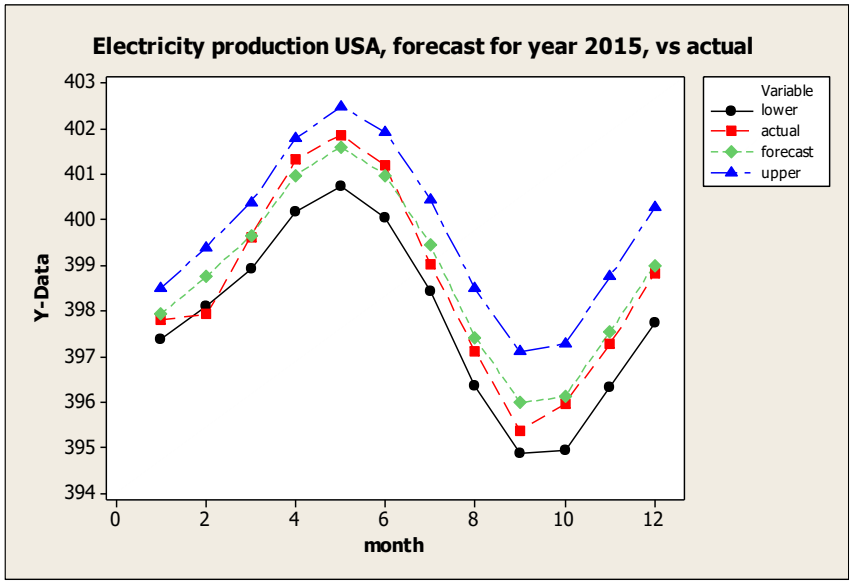


Figure (7.11): Forecast for the generated electricity amount for the year 2015 using SARIMA(0,1,1)(0,1,1)12 Model

The following table also shows these forecasts together with 95% confidence limits:

Table (7.2): Forecast for the generated electricity amount for the year 2015 using SARIMA(0,1,1)(0,1,1)12 Model

Month	Lower limit	Actual value	Forecast	Upper limit
1	397.369	397.81	397.931	398.493
2	398.094	397.93	398.750	399.406
3	398.913	399.62	399.652	400.391
4	400.168	401.34	400.981	401.794
5	400.734	401.88	401.615	402.496
6	400.046	401.20	400.991	401.935
7	398.450	399.04	399.454	400.457
8	396.368	397.10	397.427	398.487
9	394.878	395.35	395.991	397.103
10	394.952	395.95	396.115	397.277
11	396.331	397.27	397.542	398.753
12	397.750	398.84	399.008	400.266

As we can see from table (7.2), the proposed model could produce forecasts that are very near to the actual values of the production amounts, also it was able to model seasonality in the data with high accuracy. Also, notice that the confidence limits contain the actual values, except the production amount for February, as the actual value lies outside the limits, but bearing in mind that this is a 95% C.I., then one would expect about 5% of the values to be outside the confidence limits, and hence this does not down grade the postulated model. Also,

note that these limits are very narrow, indicating that the model is very highly reliable.