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 <u>different</u> 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 $(y_1, y_2, \dots, y_n)$ that are observed in the time interval $(1,2, \dots, n)$ is a realization drawn from multivariate random vector $(Y_1, Y_2, \dots, Y_n)$ that have cumulative distribution function $F(y_1, y_2, \dots, y_n)$ 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, 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 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 fifty 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 $F_{Y_{t_1}}(y_{t_1})$.

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 <u>single point</u> drawn from bivariate random variable $\left(Y_{t_1}, Y_{t_2}\right)$ (which represents all observations that can be observed at the two time points (t_1, t_2) and has a cumulative distribution function $F_{Y_{t_1}, Y_{t_2}}(y_{t_1}, y_{t_2})$.

<u>In general modern time series analysis assumes the existence of a (theoretical) stochastic process capable of generating an infinite</u>

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 $(Y_1, Y_2, \dots, Y_n)$.

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

$(Y_{t_1+k}, Y_{t_2+k}, \dots, Y_{t_m+k})$ for any time point t and any time shift k .

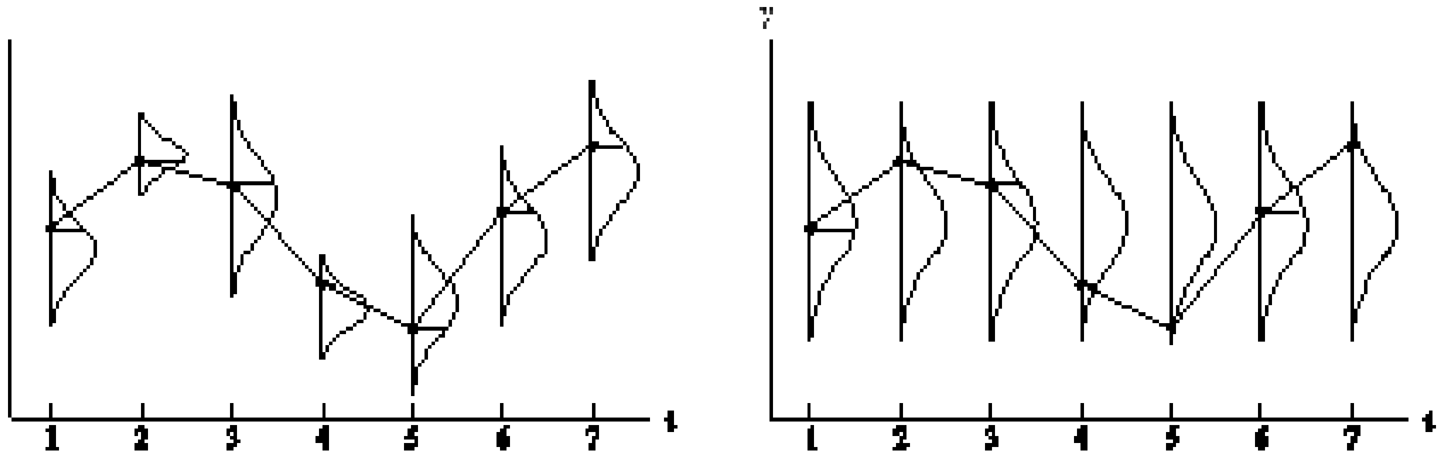
Mathematically we can write the condition of strict stationarity as:

$$(Y_{t_1}, Y_{t_2}, \dots, Y_{t_m}) =^d (Y_{t_1+k}, Y_{t_2+k}, \dots, Y_{t_m+k})$$

$$\Rightarrow P(Y_{t_1} \le c_1, Y_{t_2} \le c_2, \dots, Y_{t_m} \le c_m)$$

$$= P(Y_{t_1+k} \le c_1, Y_{t_2+k} \le c_2, \dots, Y_{t_m+k} \le c_m)$$

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 (or the time distance between them).

So strict stationarity leads to the following:

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

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

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,

$$Var(Y_t) = \gamma(t,t) = \gamma(0), \quad t = 0, \pm 1, \pm 2, \ldots$$

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. :

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

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

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

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

$$Cov(Y_{t-k}, Y_t) = \gamma(k), \quad t = 0, \pm1, \pm2, \dots; k = \pm1, \pm2, \dots$$

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 $(Y_{t_1}, Y_{t_2}, \dots, Y_{t_m})$ 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, 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 nonstationarity, we will face many difficulties. The most important is the large number of parameters, such as expectations, variances and covariances 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 $(y_1, y_2, \dots, y_n)$, then the major parameters of the theoretical process are :

$$E(\boldsymbol{Y}) = ((Y_1), E(Y_2), \dots, E(Y_n))` = [\mu_1, \mu_2, \dots, \mu_n]`$$

$$Var(\boldsymbol{Y}) = \gamma(s,t) = \begin{bmatrix} \gamma(1,1) & \gamma(1,2) \dots & \gamma(1,n) \\ \gamma(2,1) & \gamma(2,2) \dots & \gamma(2,n) \\ \vdots & \vdots & \vdots \\ \gamma(n,1) & \gamma(n,2) \dots & \gamma(n,n) \end{bmatrix}$$

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 (linear dependence) 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:

$$Y_t = \beta_0 + \varepsilon_t, \quad t = 1,2,\dots,n$$

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

solution:

Calculate the expectation, variance and covariance of the process:

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

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

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

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

Example: If the series follow the following model:

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad t = 1,2,\dots,n$$

Where β_0, β_1 are fixed constants, and the variables $\varepsilon_1, \varepsilon_2, \dots$ 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:

$$Y_t = Y_{t-1} + \varepsilon_t, \quad t = 1,2,\dots,n$$

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

solution:

$$E(Y_t) = E(Y_{t-1}) + E(\varepsilon_t) = E(Y_{t-1}), \;\; t = 1,2,\dots n$$

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

$$\mathrm{Var}(Y_t) = \mathrm{Var}(Y_{t-1}) + \sigma^2 + 2Cov(Y_{t-1}, \varepsilon_t)$$

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

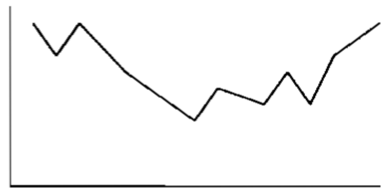
So that $\mathrm{Var}(Y_t) \neq \mathrm{Var}(Y_{t-1})$, i.e. the variance is not constant , and hence the process is not stationary.

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 testing 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.

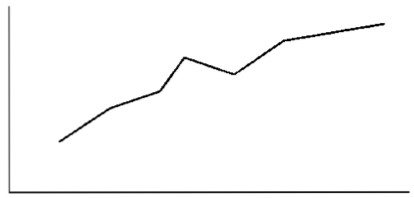
However, if we notice existence of a trend component and/or that the dispersion of the data change over time then we find this an indication of non-stationarity of the series, see figure bellow:

not Stationary in variance

Stationary series



not Stationary in mean (of second order)



Series 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 always preferable to calibrate the covariance by dividing by the product of standard deviation of the variables Y_t and 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:

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{Var(Y_s)\,Var(Y_t)}}$$
$$= \frac{E[(Y_s-\mu_s)(Y_t-\mu_t)]}{\sqrt{E(Y_s-\mu_s)^2\,E(Y_t-\mu_t)^2}} \;\; ; \;\; s,t = 0, \pm 1, \pm 2, \ldots$$

Since it measure 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:

$$E(\varepsilon_t) = 0, \quad t = 0, \pm 1, \pm 2, \ldots$$

$$Var(\varepsilon_t) = \sigma^2, \quad t = 0, \pm 1, \pm 2, \ldots$$

$$\gamma(k) = Cov(\varepsilon_t, \varepsilon_{t-k}) = 0, \qquad k \neq 0; \quad t = 0, \pm 1, \pm 2, \ldots$$

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

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:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad t = 1,2, \dots, n$$

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,

$$\gamma(s,t) = Cov(\beta_0 + \beta_1 s + \varepsilon_s,\ \ \beta_0 + \beta_1 t + \varepsilon_t),$$

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

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:

$$Y_t = \varepsilon_t - \theta\varepsilon_{t-1}, \quad t = 1,2,\dots,n$$

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

solution:

$E(Y_t) = 0, \;\; t = 1,2,\dots,n$

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

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

$$= \sigma^2 + \theta^2\sigma^2 = \sigma^2(1 + \theta^2)\ ; t = 1,2, \ldots$$

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 $y_1, y_2, \dots, y_n$ of the time series. Based on these estimates, we can estimate the 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:

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

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:

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

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

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

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:

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

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
Number of sold units y_t (in thousands)	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:

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

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

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

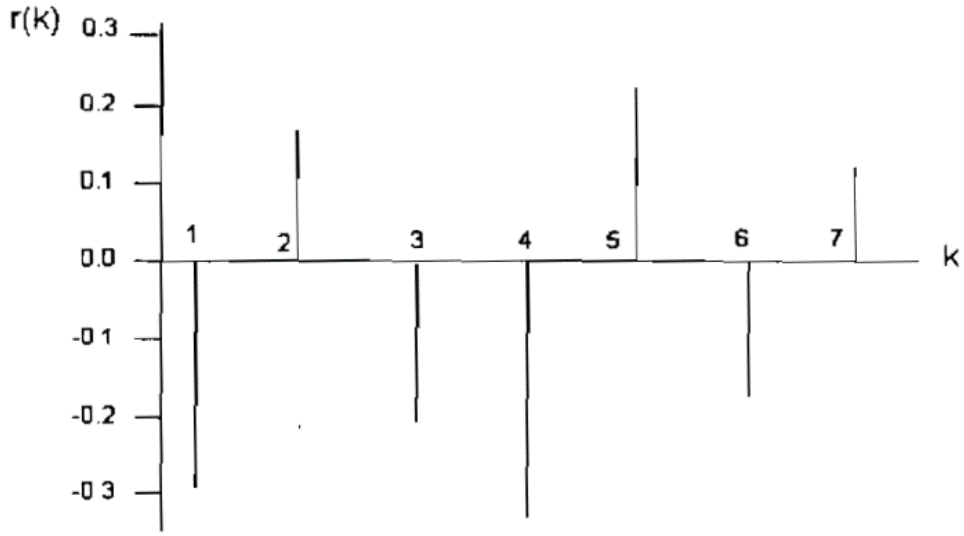
Also,

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

Similarly, the rest of the values are calculated:

$$r_3 = -0.21, \qquad r_4 = -0.33, \quad r_5 = 0.21, \quad r_6 = -0.17, \quad r_7 = 0.13$$

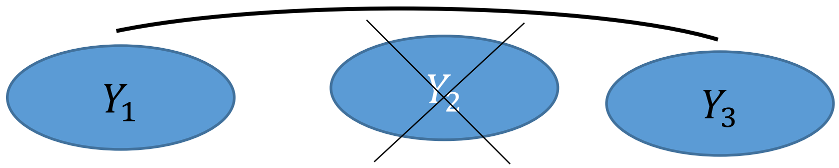
The auto-correlation function can be drawn such that, on the horizontal axis the time lags, k, and on the vertical axis auto-correlation coefficients, this figure is called the *correlogram*.



2.3 Partial autocorrelation function

The idea of this correlation arise as follows:

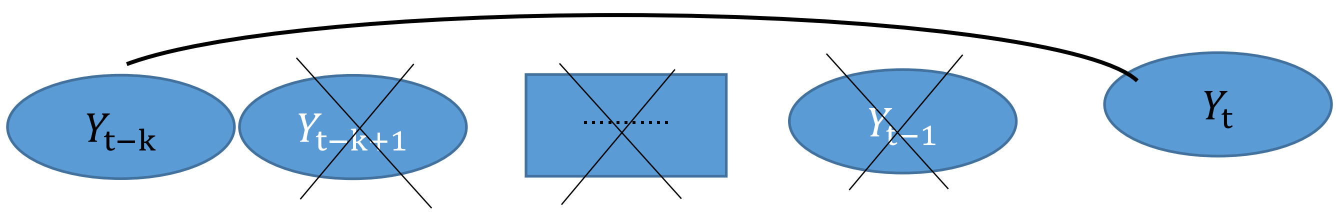
If two variables, say, Y_1 and Y_3 are found to be correlated , then this might be because of correlation between them and a



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 $[Y_t - E(Y_t|Y_{t-1}, \dots, Y_{t-k+1})]$ and $[Y_{t-k} - E(Y_{t-k}|Y_{t-1}, \dots, Y_{t-k+1})]$

Where $E(Y_t|Y_{t-1}, \dots, Y_{t-k+1})$ and $E(Y_{t-k}|Y_{t-1}, \dots, Y_{t-k+1})$ 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:

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

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

$$+ \cdots + \phi_{kk} E(Y_{t-p} Y_{t-k}) + E(\varepsilon_t Y_{t-k})$$

So,

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

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 $\phi_{11}, \phi_{22}, \dots, \phi_{kk}$. 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) :

$$\phi_{kk} = \left\{ \begin{array}{lr} 1 & ,k = 0 \\ \rho_1 & ,k = 1 \\ \dfrac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & 1 \end{vmatrix}} & ,k = 2,3,\ldots \end{array} \right\}$$

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,

$$\phi_{kj} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j} \quad ,j=1,2,\dots,k-1$$

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 autocorrelation between Y_{t-k} and Y_t , however, there might be a nonlinear 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.:

$$\hat{\phi}_{kk} = r_{kk} \quad , k = 0,1, \dots$$

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 partial 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:

$$r_1 = \frac{\sum_{t=2}^{9}(y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^{9}(y_t - \bar{y})^2}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots, \quad r_8 = \frac{\sum_{t=9}^{9}(y_t - \bar{y})(y_{t-8} - \bar{y})}{\sum_{t=1}^{9}(y_t - \bar{y})^2}$$

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:

$$(158 - 206.89), (222 - 206.89), \dots, (169 - 206.89)$$

$$\Longrightarrow (-48.89), (15.11), (41.11), (9.11) \ \dots, (-37.89)$$

Then we get the required r_k coefficients as follow:

r_1

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

$$= 0.2651$$

$$r_2$$

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

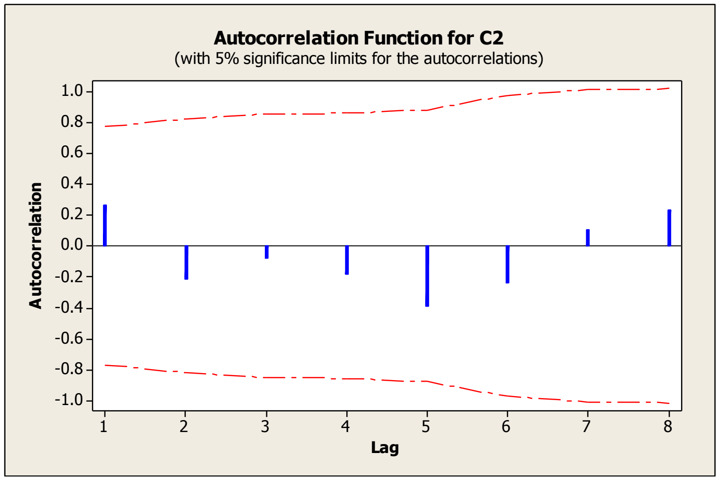
$$= -0.212$$

And the same for other coefficients,

$$r_3 = -0.076, \;\; r_4 = -0.183, \; r_5 = -0.387, r_6 = -0.242,$$

$$r_7 = 0.104, \qquad r_8 = 0.230$$

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:

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

Then:

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

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

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

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

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,

$$r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j} \quad , j = 1,2, \dots , k-1$$

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