INTRODUCTORY TIME SERIES

BY TANUJIT CHAKRABORTY, ISI KOLKATA

Content Page No

Time Series Analysis

TANUJIT CHAKRABORTY.

TIME SERIES ANALYSIS

By Tanujit Chakraborty.

Time Series Analysis
- Classical approach
- Probability modelling

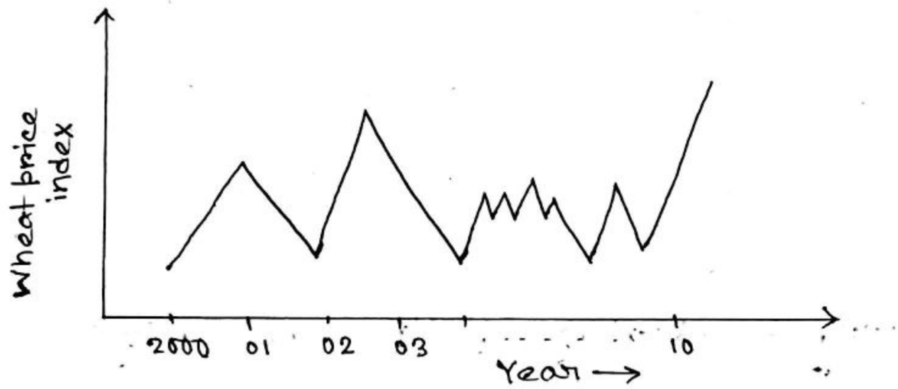
- **What is a time series?**

Ans:- A time series is a collection of observations made sequentially through time.

Example:-

1. Economic and financial time series:

These are the time series recorded in economics and finance.



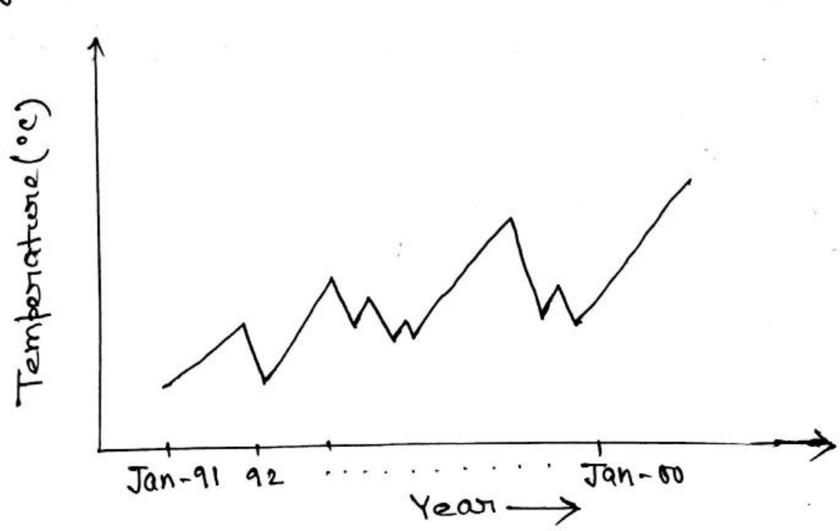
Year	W.P.I.
2000	P_0
2001	P_1
2002	P_2
⋮	⋮
2010	P_{10}

2. Physical time-series:

These are the time series occurs in physical science particularly in meterrology.

ex. Marine science and geophysics.

e.g. Rainfall on successive days, air limp recorded in successive days, etc.



Year	Temp. (°c)
Jan-1991	t_1
Jan-1992	t_2
⋮	⋮
Jan-2010	t_{10}

3. Marketing time series:
These are time series arise in marketing.
e.g. ~~No. of~~ sale figures in successive months or weeks, monthly receipts, advertising, etc.

4. Demographic time series:
These are time-series occur in the study of population changes.
e.g.: Popln. of India measured annually, monthly birth totals in india.

5. Share prices on successive days, export totals, in successive days, average income in successive months, company profit in successive months.

- Why we need seperate analysis for time series data?

Ans:- ~~The special~~
Observations on a phenomenon which is moving through time generate an ordered set known as a time series.
The special feature of time-series analysis is the fact that successive observations are usually not independent and that the analysis must take into ~~order~~ account the time order of the observations. When successive observations are dependent, future values may be predicted from past observations. If a time-series can be predicted exactly, it is said to be deterministic, otherwise called stochastic.

- What are the utility or advantages of time series analysis?

ANS:-
i) it helps in the analysis of past behavior of a variable, i.e. to determine the type and nature of the variations in the data.

ii) it helps in forecasting. It enables us to predict or estimate or forecast the behaviour of the phenomenon in future which is very essential for business planning.

iii) it helps in evaluation of current achievements. It helps to compare the actual current performance of accomplishments with the expected ones (on the basis of the past performances), and analyse the causes of such variations, if any.

iv) it helps in making comparative studies. It helps us to compare the changes in the values of different phenomenon at different times or places, etc.

- What are the objectives of time series analysis?

Ans :- The objectives of time series analysis may be classified as:

(a) Description: — When presented with a time series, the first step in the analysis is usually to plot the observations against time to give what is called a time plot, and then to obtain simple descriptive measures of the main properties of the series. For some series, the variation is dominated by such 'obvious' features, and a fairly simple model, which only attempts to describe trend and seasonal variation, may be perfectly adequate to describe the variation in the time series. For other series, more sophisticated techniques will be required to provide an adequate analysis and a more complex model will be constructed.

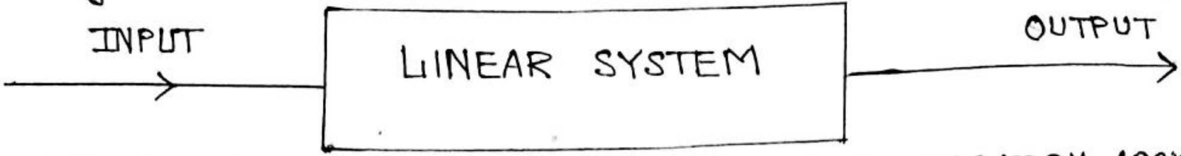
A time-plot will not only show up trend and seasonal variation but will also reveal any 'wild' observations or outliers that don't appear to be consistent with the rest of the data. The treatement of outliers is a complex object and robust methods are designed to be insensitive to outliers.

Other features to look for in a time ~~series~~ plot include sudden or gradual changes in the properties of the series; the analyst should also look out for a step change in the level of the series, any changes in the seasonal pattern, the possible presense of turning points, etc. If there is some sort of discontinuity in the series, then different models may need to be fitted to different parts of the series.

(b) Explanation : — When observations are taken on two or more variables, it may be possible to use the variation in one time series to explain the variation in other series. This may lead to a deeper understanding of the mechanism that generated in a given time series.

Although multiple regression models are occasionally helpful here, they are not really designed to handle time-series data, with all the correlations inherent there in, and alternative classes of models are considered. For example, it is of interest to see how sea-level is affected by temperature and pressure, or to see how sales are affected by price and economic conditions. A class of models, called transfer-function models or linear stochastic TS models representing a linear-system, enable us to model TS data in an appropriate way. A linear system converts an input series to an output series by a linear operation.

Given observations on the input and output to a linear system, the analyst wants to assess the properties of the linear system.



(c) Prediction:— Given an observed time series, one may want to predict the future values of the series. This is an important task in sales forecasting and in the analysis of economic and industrial time series.

(d) Control:— TS are sometimes collected or analysed so as to improve control over some physical or economic system. For example, when a TS is generated that measures the 'quality' of manufacturing process, the aim of the analysis may be to keep the process operating at a 'high'-level. Control problems are closely related to prediction in many situations. For example, if one can predict that a manufacturing process is going to move off target, then appropriate corrective action can be taken.

Control procedures vary considerably in style and sophistication. In SQC, the observations are plotted on control charts and the controller takes action as a result of studying the charts. A more complicated type of approach is based on modelling the data and using the model to work out an 'optimal' control strategy. In this approach a stochastic model is fitted to the series, future values of the series are predicted and then the input process variables are adjusted so as to keep the process on target.

Classification of Time Series

Discrete $\{X_t\}$

A TS is said to be discrete when observations are taken only at specific times, usually eaually spaced, even when the measured variable is of the continuous type.

Continuous $\{X(t)\}$

A TS is said to be continuous when observations are taken continuously through time even when the measured variable can take only a discrete set of values.

Time Series

Deterministic

A TS is said to be deterministic if it can be determined or predicted exactly.

e.g. $Y_t = \cos(2\pi f t)$

$Y_t = \mu + R\sin(\omega t + \phi)$

Non-deterministic / Stochastic / Statistical

A TS is said to be stochastic if the future is only partly determined by the past values, so that exact predictions are impossible and must be replaced by the idea that future values have a probability distribution which is conditioned by a knowledge of past values.

e.g. $Y_t = \mu + \epsilon_t : \{\epsilon_t\} \sim iid(0, \sigma^2)$

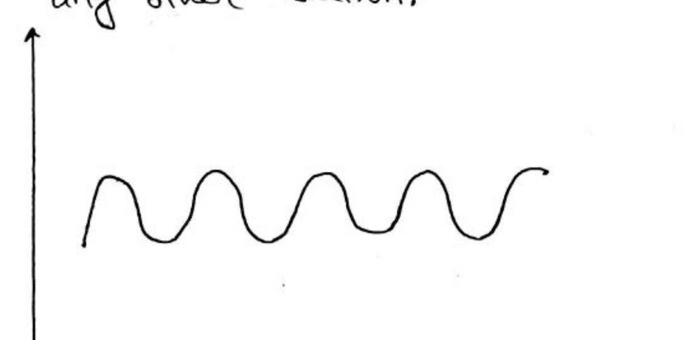
$Y_t = \alpha + \beta\epsilon_{t-1} + \epsilon_t : \{\epsilon_t\} \sim iid(0, \sigma^2)$

$Y_t = \alpha + \beta Y_{t-1} + \epsilon_t : \{\epsilon_t\} \sim iid(0, \sigma^2)$

Time series

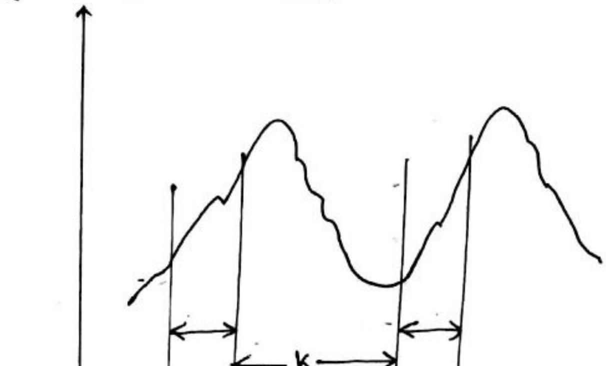
Stationary

A TS is said to be stationary if there is no systematic change in mean (no trend), if there is no systematic change in variance, if strictly periodic variations have been removed. In other words, the properties of one section of the data are much like than any other section.



Non-stationary or Evaluative

A TS is said to be non-stationary or evaluative if it is not a stationary one, i.e. if the properties of one section of the data is different from any other section.



- **The problems of time series analysis:** — The ultimate object of analysis of a time-series – as of statistical analysis as a whole – is to arrive at a deeper understanding of the causal mechanisms which generate it, because we wish to extrapolate into the future or out of sheer curiosity. It does not follow, however, that such understanding can be achieved by considering one series alone. For the present we curb our ambition to some extent by confining ourselves to the study of the type of behaviour of a single series and the setting up of models which can generate it; recognising that such models themselves may be only portions of a more basic structural system.

"The main objective in analysing time series is to understand, interpret and evaluate changes in economic phenomena in the hope of ~~the~~ more correctly anticipating the course of future events".

However, usually we shall come accross time series showing continual changes, overtime, giving us an overall impression of haphazard movement. A survey of the practical examples we have given, will reveal that the change is not totally haphazard and a part of it, at least, can be accounted for. The part which can be accounted for is the systematic part and the remaining part is the unsystematic or irregular.

Components of Time Series

- Systematic
  - Trend (T_t)
  - Periodic
    - Short term (period < 1 yr.) or Seasonal (S_t)
    - Long term (period > 1 yr.) or Cyclical (C_t)
- Random or irregular (ϵ_t / I_t)

$$Y_t = f(T_t, S_t, C_t, I_t)$$
$$= \phi(T_t, S_t, C_t, I_t) + (f - \phi)$$
$$= \phi(\cdots\cdots) + \epsilon$$

We choose ϕ such that ϵ will be minimum, where ϵ = error of discripency.

- A typical time series may be composed of four parts:—

(a) a trend, or long-term movement, (b) a seasonal effect, (c) cyclical movement, (d) a random or irregular component.

In a particular time series one or more of the systematic components may be absent. However, the "random" component is almost always present. Seperation of the different components of a time series is of importance, because it may be that we are interested in a particular component or that we want to study the series after eliminating the effect of a particular component.

(a) Secular trend:— This is the long term upward and downward movement of the series due to factors that influence the mean of the series. Some series may remain more or less at a constant level. sudden or frequent changes are imcompatible with the idea of the trend. thus trend may be defined as a slowly changing non-random component of a time series.

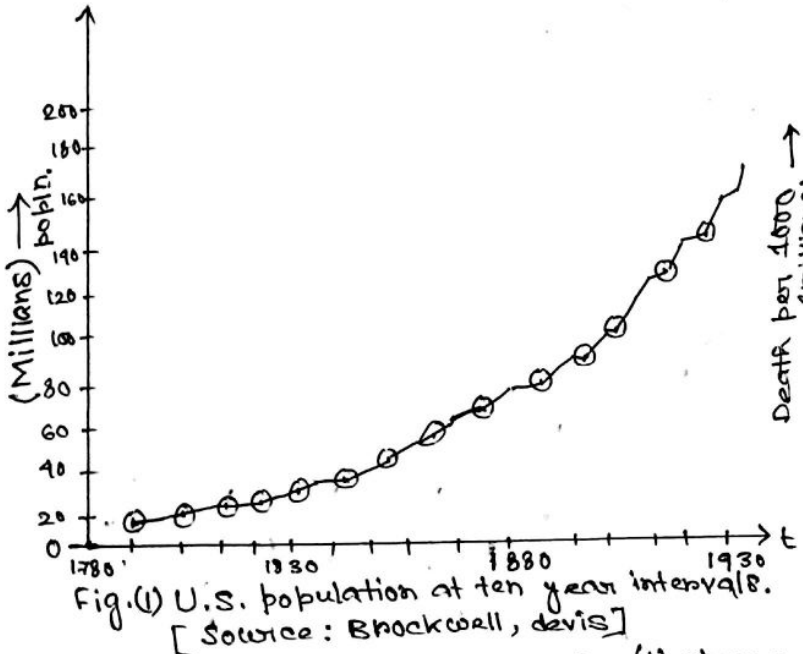
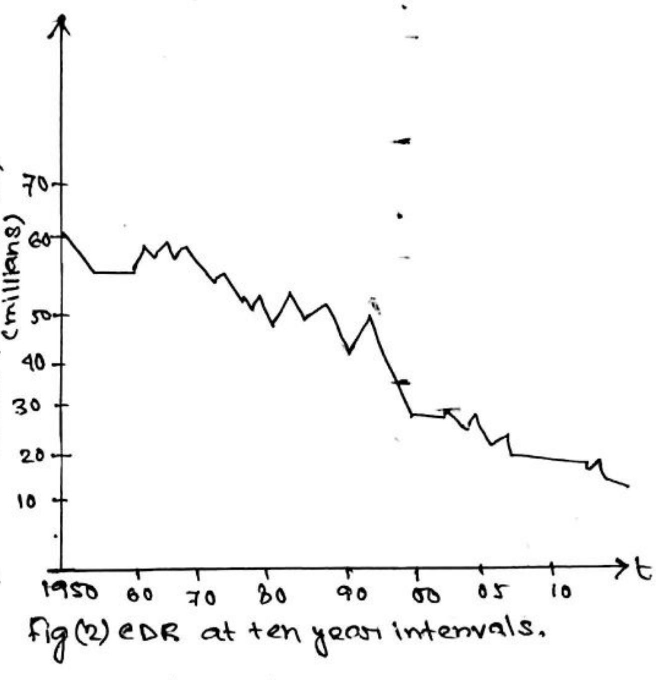


Fig.(1) U.S. population at ten year intervals.
[Source: Brockwell, devis]



Fig(2) CDR at ten year intervals.

The population data in Fig.(1) shows an upward trend. Not all time series shows upward trend. Some, like crude death rate, shown in figure (2), exhibit a downward trend.

(b) <u>Seasonal variations</u> :- This is a measure of the characteristic behaviours of the series during each season (specific interval of time) in the period, which may be one year. Many TS in business and industry show variations in values from one season to the other, due to climate conditions, ~~produce~~ variations, which repeat itself ~~at different~~ quaterly, monthly, annually. This yearly variation due seasonal effect is known as seasonal variation. ⑧

We may interested in seasonal variation either because of we wish ~~statistically~~ to eliminate seasonal variation from a time series or because we are intersted in the seasonal variation itself.

A seasonal structure may be constant from year to year [Fig. 1(a)] or may show slight deviations from one year to the next [Fig. 1(b)-(d)]

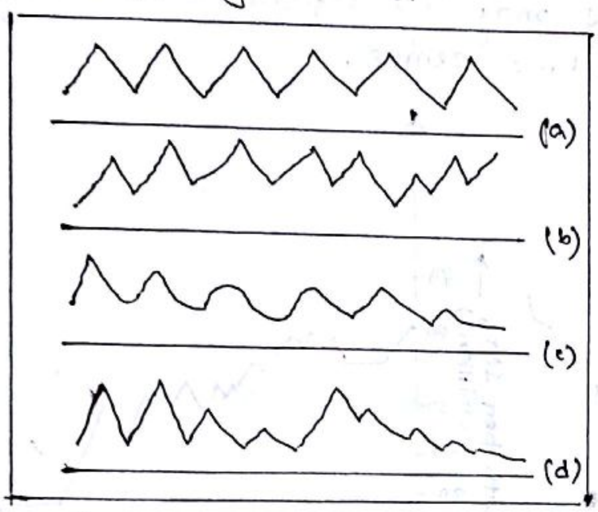


Fig. 1. (a)-(d)

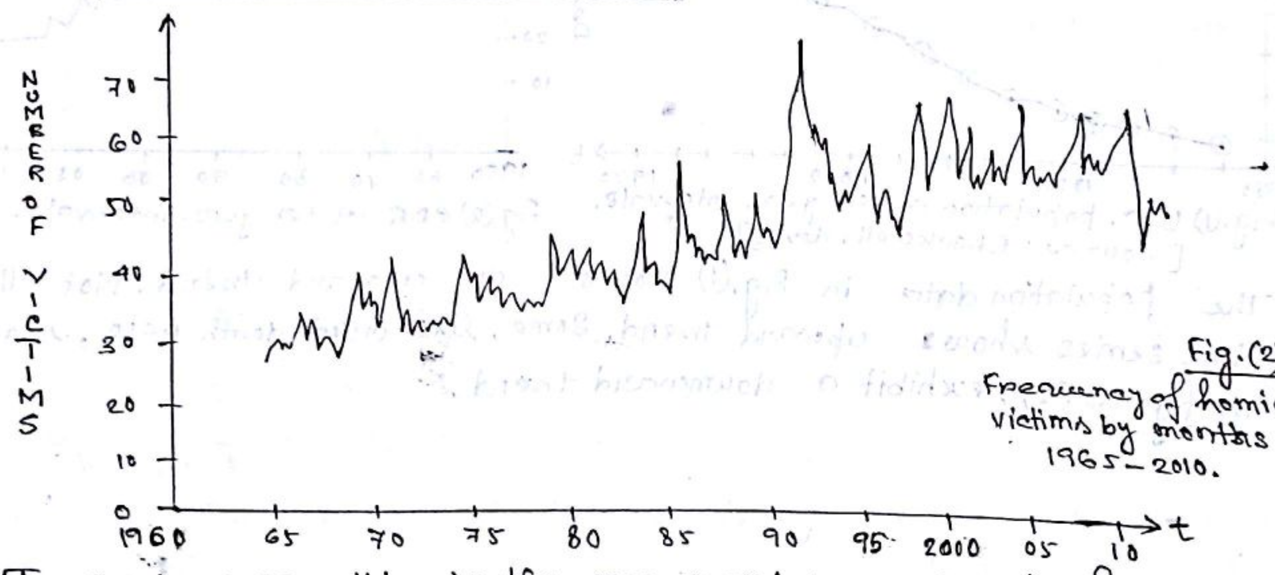


Fig.(2)
Frequency of homicide victims by months 1965-2010.

In short, seasonality is the non-random component of a time series which tend to repeat itself at regular interval of time. This is clearly visible in the data on canadian crimes (Fig.(2)). Data on crop-production, rainfall, trade are expected to show marked seasonal components.

(c) **Cyclical movements** :— The cyclical fluctuation means, the oscillatory movement of a time series, the period of oscillation being more than a year. One complete period is called a cycle. By cyclical component we mean, therefore, the slow oscillatory movement of a time series which repeats itself in each cycle. The length of the cycle and its intensity of fluctuation may vary from one cycle to the other. The length of many cycles average about 3 to 4 yrs though some are longer than 15 yrs. Moreover, the average length of a cycle is usually longer than that of seasonality and the magnitude of a cycle is usually more variable than that of seasonality.

Many data series include combinations of the preceding patterns. As an example, Figure 1.(3) shows trend, seasonality, and cyclical behaviours.

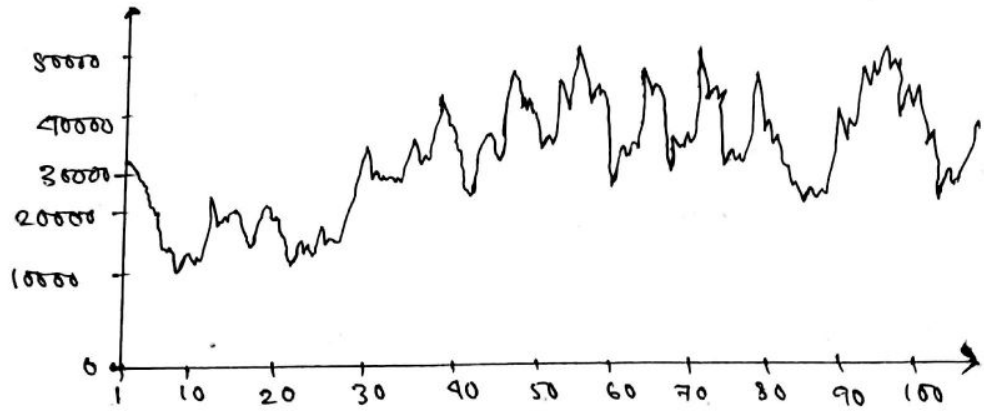


Fig. 1.(3) :- Montly tourist arrivals to the Sri-Lanka.

- **Distinction between seasonal and cyclical Pattern** → The major distinction between a seasonal and a cyclical pattern is that the former is of a constant length and recurs on a regular periodic basis, while the latter varies in length. Moreover, the average length of a cycle is usually longer than that of seasonality and the magnitule of a cycle is usually more variable than that of seasonality.

(d) **Irregular or random fluctuations** :— Apart from the regular variations, all the series contain another factor called the random or irregular which are not accounted for by secular trend, seasonal and cyclical variations. These fluctuations are purely random, erratic, unpredictable, and are due to numerous non-recurring and irregular circumstances which are beyond the control of human hand.

- Mathematical models for time series :— In mathematical model or simply a model is the representation of the system is an unknown function in terms of a known functions or variables.

Classical Models

- Additive: $(Y_t = T_t + S_t + C_t + I_t)$
- Multiplicative: $(Y_t = T_t \times S_t \times C_t \times I_t)$

where, Y_t is the original series, T_t, S_t, C_t, I_t are respectively, trend, seasonal, cyclical and random components.

- Assumptions of Additive and multiplicative models :—

In additive model, it will have positive or negative values and in longrun $\sum_t I_t$ will be zero. According to multiplicative model, I_t instead of assuming positive and negative values, fluctuates above or below unity and geometric mean of I_t in the long-run is unity.

It is noted that multiplicative model can be converted into an additive model by taking logarithms since

$$\log Y_t = \log T_t + \log S_t + \log C_t + \log I_t .$$

Here it is noted that in multiplicative model, $I_t > 0$. Hence $\log I_t$ is positive or negative accoring as $0 < I_t < 1$ or $I_t > 1$, which is desirable.

- A multiplicative model is more appropriate for explaining the variations of a business and economic data — Explain.

The assumptions underlying the additive model is that there is no interaction among the different components when at the multiplication scheme assumes the presence of such interaction. So that the suitablity of one scheme compared to the other depends on the nature of the time series. For instance, the multiplicative model may be more appropriate in the study of the time series on price, if it is the case that in an inflationary situation, the seasonal component is re-inforced by the rising trend in prices. Where as, in a deflationary situation the falling trend in prices dampens these components.

Alternative answer:-

The concept of the model $Y_t = T_t \times S_t \times C_t \times I_t$ is more useful than that of the model $Y_t = T_t + S_t + C_t + I_t$, this is because S_t, C_t and I_t tend to remain more nearly constant in magnitude relative to trend, rather than in absolute terms.
Furthermore, the movements are ordinarily more meaningful when considered relative to each other than when considered in absolute terms.
It is possible to compute a seasonal index which remains constant in absolute terms but a seasonal index relative to trend is changing because of alterations in relative importance or in trend values of the months. For example,
(i) $T_t = 1000$ units, $S_t = 100$ units, (ii) $T_t = 2000$ units, $S_t = 100$ units,
in the two cases the seasonal components are equal in absolute term but the seasonal movement relative to trend are
for (i) $\frac{S_t}{T_t} \times 100\% = 10\%$ and for (ii) $\frac{S_t}{T_t} \times 100\% = 5\%$,
which are changing because of alterations in trend values.

Occasionally series are encountered for which better results are obtained if the seasonal movement is considered constant in absolute rather than relative terms. This is especially likely to be the case when the seasonal movement typically falls almost at zero at one or more months.

- **Binary Process:-** This refers to a special type of time series arises when observations can take or assume only one of two values, usually denoted by '0' and '1'.
e.g. In computer science, the position of a switch is either 'on' or 'off' could be recorded as '1' or '0'.

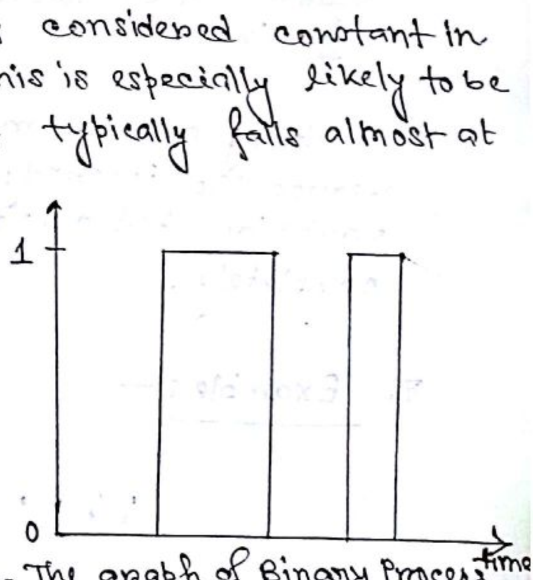


Fig:- The graph of Binary Process will be like this.

- **Point Process:-** Point process refers to a time series which occurs when we consider a series of events occurring randomly through time.
e.g. The datas of major railway disasters.

(ii) **Distinguish between seasonal and cyclical fluctuations in time-series data with a real life example for each.**

Seasonal Variation	Cyclical fluctuation
1. This is short term periodic movements where period is not longer than one year.	1. This is another type of periodic movement where period is greater than one year.
2. These fluctuations are found with definite periodicity and reappear almost at regular interval of time.	2. Cyclical fluctuations are not as regular as seasonal fluctuations and period of cycle as well as intensity of fluctuation varry from one cycle to another.
3. This variation is mostly related to changing seasons or religious or social customs.	3. cyclical fluctuations are caused by the joint interaction of many factors and are found to exist in almost all business and economic activity.
4. It has no distinct phases like that of cyclical variation.	4. It has four phases like prosperity decline, depression and recovery. They constitute a cycle.
5. For forecasting purpose seasonal variation is most effective.	5. Here the cycle is irregular and uneven in duration. Hence this is uneffective for forecasting.
6. Moving average method can reduce the intensity of seasonal variation but cannot wipe out completely.	6. If the period of moving average is a multiple of the period of cyclical movements, then the moving average method completely reduces cyclical variation.
7. Example:—	7. Example:—

[ISS EXAM'12] [8 Marks]
What are the different methods for measurement of trend
Discuss the method of fitting a straight line using least squares method.

Method of trend enumeration/Determination of trend component:

There are two reasons for attempting to describe the trend of a time series. First, it may be desired to eliminate the trend from the series, second, it may be desired to study the trend itself or to attempt to forecast the future behaviour of the trend. Methods are:—
~~The purpose for which~~

1. Free hand curve fitting,
2. Semi-average method,
3. Fitting of mathematical curves,
4. Moving average method.

1. Method of free-hand curve-fitting :— [quick graph estimator method]

The simplest method of describing a trend graphically is by inspection. Here, we first draw the line-diagram for the data. Then we draw a free-hand smooth curve which seems to fit the data best. The method, however is quite subjective and its use therefore calls for sound judgement. When a curve is fitted to a set of data, a criterion of fit is involved.

A free-hand smooth curve obtained on plotting the values Y_t against 't' enables us to form an idea about the general 'trend' of the series. Smoothing of the curve eliminates other components like regular and irregular fluctuations.

This method does not involve any complex mathematical techniques and can be used to describe all types of trend, linear and non-linear. Thus simplicity and flexibility are strong points of this method.

Its main drawbacks are:

i) This method is very subjective, i.e. the bias of the person handling the data plays a very important role and as such different trend curves will be obtained by different persons for the same set of data. As such 'trend by inspection' should be attempted only by skilled and experi-enced statistician.

ii) It does not enable us to measure trend.

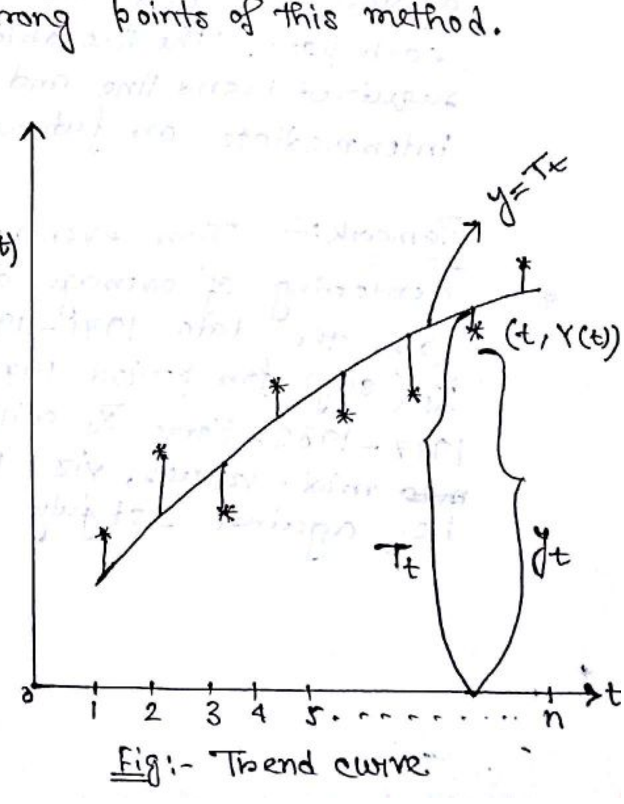
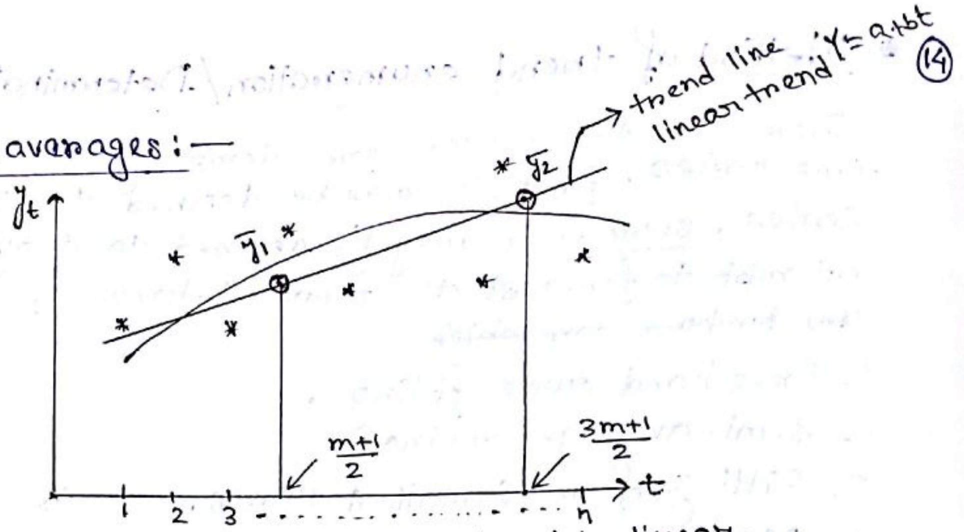


Fig:- Trend curve

2. Method of semi averages:—

t	y_t	
1, 2, 3, ..., m	$y_1, y_2, y_3, \ldots, y_m$	$\bar{y}_1$
$m+1$, ..., $n=2m$	$y_{m+1}, \ldots, y_{2m}$	$\bar{y}_2$



Assumption: The underlying trend is linear.

Two points are: $\left(\frac{m+1}{2}, \bar{y}_1\right)$ and $\left(\frac{3m+1}{2}, \bar{y}_2\right)$.

Equation of the straight line:

$$\frac{y-\bar{y}_1}{\bar{y}_2-\bar{y}_1} = \frac{t-\frac{3m+1}{2}}{\frac{3m+1}{2}-\frac{m+1}{2}} \Rightarrow (y-y_1) = \frac{\left(t-\frac{m+1}{2}\right)}{m}\cdot(\bar{y}_2-\bar{y}_1) \cdots\cdots (*)$$

In this method, the whole data is divided into two parts with respect to time. e.g. if we are given Y_t for t from 1971–1982, i.e. over a period of 12 years, the two equal parts will be the data from 1971 to 1976 and 1977 to 1982. In case of odd number of years the two parts are obtained by omitting the value corresponding to the middle year, e.g. for the data from 1971-1981; the two parts would be the values for 1971-75 and 1977-81, the value corresponding to middle year, viz. 1976 is omitted. Next we compute the arithmatic mean for each part and plot these two averages against the mid-values of the respective periods covered by each part. The line obtained on joining these two points is the required trend line and may be extended both ways to estimate intermediate or future values.

Remark:— For even number of years like 8, 12, 16, etc. the centering of average of each part would create problems, e.g. for the data 1971 - 1982 ($n=12$), let the two averages be $\bar{X}_1$ (say) for period 1971 - 1976 and $\bar{X}_2$ (say) for the period 1977 - 1982. Here $\bar{X}_1$ will be plotted against the mean of two mid-values, viz. 1973 and 1974 for the period 1971 - 1976, i.e. against 1st July 1973. Similarly, for the period 1977 - 1982.

Advantages of the Method of Semi Averages :—

(i) It is a simple method and readily comprehensible as compared to the 'method of least squares' on the 'moving average method'.
(ii) The trend figures are objective in the sense that any two persons will get the same trend line from a set of figures.

Disadvantages of the Method of Semi Averages :—

(i) This method assumes linear relationship between the plotted points — which may not exist. Moreover, the limitations of arithmatic mean as an average also stand in its way, i.e. in this method trend is affected appreciably by higher or extreme values.

3. Method of fitting mathematical curves :— The principle of least squares is the most popular and widely used method of fitting mathematical functions to a given set of data.

As the least squares method works only under some particular assumption, so here we assume that,

i) $y_t = T_t + I_t$

ii) $T_t = \begin{cases} a + bt \\ a + bt + ct^2 \\ \vdots \\ a + bt + ct^2 + \cdots + dt^p \\ ab^t \end{cases}$ → is linear in parameters.

Ques:- How one can determine the trend by fitting a polynomial app. degre

⇒ Polynomial curve fitting by the method of Least square :—

Let us suppose that we are here given with a time series data $Y_t : t = 1, 2, \ldots, n$ such that it consists of only trend and random error (i.e. there is no evidistance of presence of any periodic variation and if there by any inner periodic variation, that can be merged to error without any loss). Further trend (T_t) can be represented as a polynomial of degree p in t,

$$T_t = a_0 + a_1 t + a_2 t^2 + \cdots + a_p t^p$$ (which is linear in parameters $a_0, a_1, \ldots, a_p$).

Then our problem is to determine T_t, i.e. to determine $a_0, a_1, \ldots, a_p$ by minimising

$$S(a_0, a_1, \ldots, a_p) = \sum_{t=1}^{n} (Y_t - a_0 - a_1 t - \cdots - a_t t^p)^2 \text{ w.r.t. } a_0, a_1, \ldots, a_p$$

subject to $\sum_{t=1}^{n} \epsilon_t = \sum_{t=1}^{n} (Y_t - a_0 - a_1 \cdots - a_p t^p) = 0$

For minimum, we require, —

$\left(\frac{\text{1st order}}{\text{condition}}\right)$: $\frac{\partial S}{\partial a_j} = 0 \quad (j = 0(1)p)$

$\left(\frac{\text{2nd order}}{\text{condition}}\right)$: The $\overline{p+1} \times \overline{p+1}$ Hessian matrix

$$\begin{pmatrix} \frac{\partial^2 S}{\partial a_0^2} & \frac{\partial^2 S}{\partial a_0 \partial a_1} & \cdots & \frac{\partial^2 S}{\partial a_0 \partial a_t} \\ \vdots & \vdots & & \vdots \\ & & \cdots & \end{pmatrix} \text{ is p.s.d.}$$

Now, as $S(a_0, \ldots, a_p)$ is quadratic in $a_0, a_1, \ldots, a_p$.
$\therefore$ The (Hessian matrix) second order condition for minimum is automatically satisfied.

and $\frac{\partial S}{\partial a_j} = -2 \sum_{t=1}^{n} (Y_t - a_0 - a_1 t - \cdots - a_p t^p) t^j \;\forall\; j = 0, 1, \ldots, p.$

Therefore for $j = 0$,

$$\frac{\partial S}{\partial a_0} = 0 \Rightarrow \sum_{t=1}^{n} (Y_t - a_0 - a_1 t - \cdots - a_p t^p) = 0 \text{ i.e. } \sum_{t=1}^{n} \epsilon_t = 0$$

is satisfied.

But in general, we have for $j = 0(1)p$,

$$\frac{\partial S}{\partial a_j} = 0 \Rightarrow a_0 \sum_t t^j + a_1 \sum_t t^{j+1} + \cdots + a_p \sum_t t^{j+p} = \sum_t Y_t t^j$$

These $(p+1)$ equations in $a_0, a_1, \ldots, a_p$ are called the normal equations. Determining the pth degree polynomial trend

$$T_t = a_0 + a_1 t + \cdots + a_p t^p.$$

If $(\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_p)$ be a solution of

$$\begin{pmatrix} n & \sum t & \sum t^2 & \cdots & \sum t^n \\ \sum t & \sum t^2 & \sum t^3 & \cdots & \sum t^{n+1} \\ \vdots & & & & \\ \sum t^p & \sum t^{p+1} & \sum t^{p+2} & \cdots & \sum t^{2p} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} \sum y_t \\ \sum t y_t \\ \vdots \\ \sum t^p y_t \end{pmatrix}$$

The normal equation determining $a_0, a_1, a_2, \ldots, a_p$. Then we must have,

$$S(\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_p) \leq S(a_0, a_1, \ldots, a_p).$$

Give an example of a TS where an exponential trend model could be appropriate. Discuss how you will fit the exponential trend model.

EXAMPLE :—

1. The following TS data represents the production in India for the years 1971-1975:

Year	1971	1972	1973	1974	1975
Production	1·6	4·5	13·8	40·2	135·0

The values are likely to show a constant ratio of change over the years 1971-75. Hence, an exponential trend curve $Y_t = ab^t$ could be appropriate.

Here trend function is, $Y_t = 0{\cdot}5231 \times (2{\cdot}977)^t$

Trend values would be, 1·6, 4·6, 13·8, 41·1, 122·3.

2. The following TS data represents the monthly averages of tourist arrival in India for the years 1970 to 1975:

Year	1970	1971	1972	1973	1974	1975
Monthly average tourist arrival	23401	25083	28579	34157	35263	38773

The values are likely to show a constant ratio of change over the years 1970-75. Hence, an exponential trend curve $Y_t = ab^t$ would be appropriate.

Exponential curve :— The simplest exponential curve may be written as

$$Y_t = ab^t, \text{ where } a>0,\ b>0.$$

If we take logarithm both sides we get

$$\log Y_t = \log a + t \log b.$$

Hence, $\log Y_t$ is a linear function of t. Now, a curve $Y_t = ab^t$ indicates a constant ratio of change, since $Y_t / Y_{t-1} = b$. If $0<b<1$ the Y_t values gradually decays but if $b>1$, the Y_t values gradual increases and ultimately explodes. A fundamental virtue of this type of curve is that it represents a true picture of relative variations, of ratios between magnitudes. It is the exponential curve which best measure rates of change.

The curve may be constructed on semi-logarithmic paper, the logarithmic scale extending along the y-axis and the curve will be a straight line.

Fitting of Exponential curve: —

$$Y_t = ab^t$$

$$\Rightarrow \log Y_t = \log a + t \log b$$

$$\Rightarrow Y = A + Bt \quad \text{——— (1)}$$

where $Y = \log Y_t$, $B = \log b$, $A = \log a$. ——— (2)

(1) is a straight line in t and Y and thus the normal equations for estimating A and B are

$$\left.\begin{aligned} \sum Y &= nA + B\sum t \\ \sum tY &= A\sum t + B\sum t^2 \end{aligned}\right\} \quad \text{——— (*)}$$

These equations can be solved for A and B and finally on using (2), we get,

$$a = \text{antilog}(A), \quad b = \text{antilog}(B).$$

Second degree curve fitted to logarithms: —

Suppose the trend curve is $Y_t = ab^t c^{t^2}$ ——— (*)

Taking logarithms of both sides, we get

$$\log Y_t = \log a + t \log b + t^2 \log c$$

$$\Rightarrow Y = A + Bt + Ct^2 \quad \text{——— (1)}$$

where, $Y = \log Y_t$, $A = \log a$, $B = \log b$, $C = \log c$. ——— (2)

Now (1) is a second degree parabolic curve in Y and t and can be fitted by the above explained technique. Lastly, we may obtain,

$$a = \text{Antilog}(A), \quad b = \text{Antilog}(B) \text{ and } c = \text{Antilog}(C).$$

With these values of a, b and c the curve (*) becomes the best second degree curve fitted to logarithms.

Write the concept of the modified exponential and the Gompertz curves. Then write the fitting of Gompertz curve using the fitting of modified exponential curve.

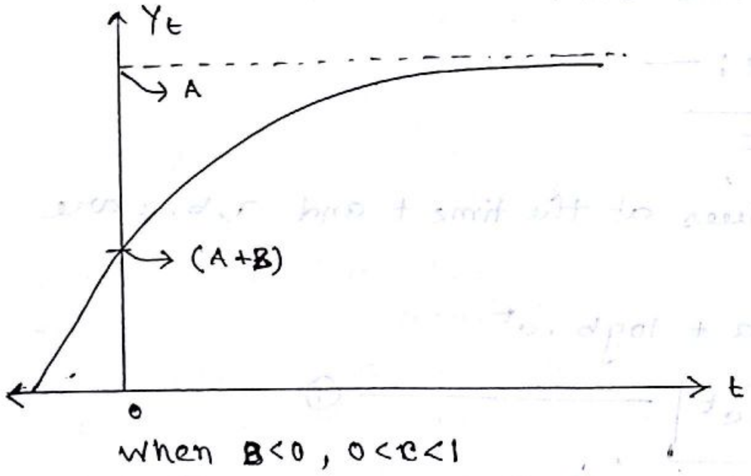
The modified Exponential :— Over a long period of time, TS are not likely to show either a constant amount of change or a constant ratio of change. It is much more likely that an increasing series (or, decreasing series) will show an increasing (or, a decreasing) amount of change but at a decreasing ratio of change. It is also possible that an increasing series may show a decline in the amount of increase.

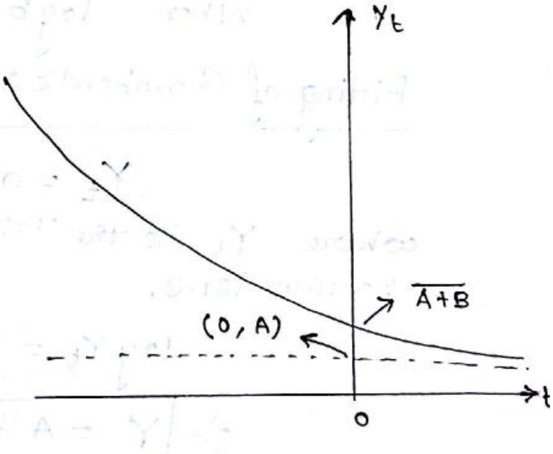
Decreasing absolute growth is not often encountered, but we shall discuss one such curve, the modified exponential. Since it serves as an excellent introduction to the more important Gompertz curves.

The equation of the modified exponential is,

$$Y_t = A + Bc^t \text{ , where } t \text{ is a positive numbers.}$$

This curves not only describes a trend in which the amount of growth declines by a constant percentage but the curve also approaches an upper limit A, called the asymptote. This is an important property of growth curves, since many Time series seem to approach an upper limit.





when $B<0$, $0<c<1$

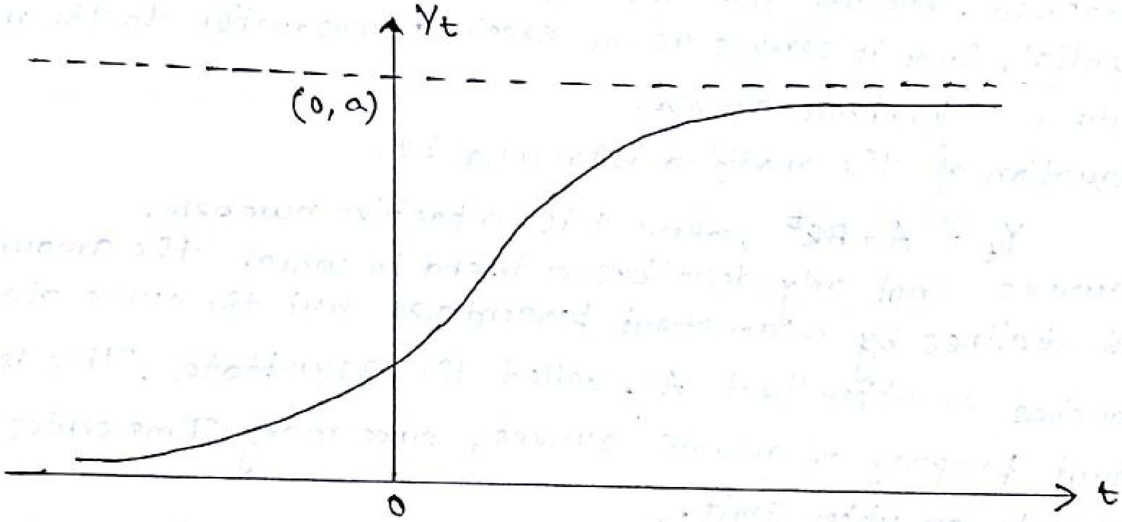
Note that, $\Delta Y_t = Y_{t+1} - Y_t = Bc^t(c-1)$. Hence, $\frac{\Delta Y_t}{\Delta Y_{t-1}} = c$.

This implies that the first differences of Y_t when plotted on a semi-logarithmic graph paper, lie on a straight line.

The Gompertz curve :— The equation for the Gompertz curve is $Y_t = ab^{c^t}$, may be put in logarithmic form:

$$\log Y_t = \log a + (\log b)c^t.$$

In the form which is of primary concern to us, the Gompertz curve describes a trend in which the growth increments of the logarithms are declining by a constant percentages. Thus, the natural values of the trend would show a declining ratio of increase, but the ratio does not decrease by either a constant amount or a constant percentage.



When $\log b < 0$ and $c < 1$, i.e. $0 < b < 1$, $0 < c < 1$.

Fitting of Gompertz curve :—

$$Y_t = ab^{c^t},$$

where Y_t is the TS values at the time t and a, b, c are parameters.

$$\log Y_t = \log a + \log b \cdot c^t$$

$$\Rightarrow \boxed{Y = A + Bc^t} \quad \text{——— (1)}$$

where $Y = \log Y_t$, $A = \log a$, $B = \log b$ ——— (2)

Now (1) is the equation of modified exponential curve and its constants A, B and c can be estimated by the method of three selected points. Finally the constants of the Gompertz curve ~~are~~ given by

$$a = \text{Antilog}(A), \text{ and } b = \text{antilog}(B).$$

Method of three selected Points:—

We take three ordinates Y_1, Y_2, Y_3 (say) corresponding to three equidistant values of t, (say) t_1, t_2 and t_3 respectively such that

$$t_2 - t_1 = t_3 - t_2$$

Substituting the values of $t = t_1, t_2$ and t_3 in $Y = A + Bc^t$, we get (*) respectively,

$$Y_1 = A + Bc^{t_1}, \quad Y_2 = A + Bc^{t_2}, \quad Y_3 = A + Bc^{t_3} \quad \text{——— (1)}$$

$$\Rightarrow Y_2 - Y_1 = Bc^{t_1}(c^{t_2 - t_1} - 1) \quad \text{——— (2)}$$

and $$Y_3 - Y_2 = Bc^{t_2}(c^{t_3 - t_2} - 1) \quad \text{——— (3)}$$

Dividing, we get

$$\frac{Y_3 - Y_2}{Y_2 - Y_1} = c^{t_2 - t_1} \qquad [\because t_2 - t_1 = t_3 - t_2]$$

$$\Rightarrow c = \left(\frac{Y_3 - Y_2}{Y_2 - Y_1}\right)^{1/(t_2 - t_1)} \quad \text{——— (4)}$$

Substituting the value of c in (2), we get —

$$Y_2 - Y_1 = B\left[\frac{Y_3 - Y_2}{Y_2 - Y_1}\right]^{t_1/\overline{t_2 - t_1}} \left[\frac{Y_3 - Y_2}{Y_2 - Y_1} - 1\right]$$

$$\therefore B = \frac{(Y_2 - Y_1)^2}{Y_3 - 2Y_2 + Y_1}\left[\frac{Y_2 - Y_1}{Y_3 - Y_2}\right]^{t_1/\overline{t_2 - t_1}} \quad \text{——— (5)}$$

Substituting B and c in (1), we get,

$$A = Y_1 - Bc^{t_1}$$

$$\Rightarrow A = Y_1 - \frac{(Y_2 - Y_1)^2}{Y_3 - 2Y_2 + Y_1} = \frac{Y_1 Y_3 - Y_2^2}{Y_3 - 2Y_2 + Y_1} \quad \text{——— (6)}$$

Substituting for A, B, c from (4), (5), (6) in (*) we get the equation of the modified exponential curve, fitted to the given time-series data Y_1, Y_2, Y_3 being ordinates of the free hand curve corresponding to the three selected points $t = t_1, t_2$ and t_3.

Remark:— If the number of years included is not a multiple of three, intervals that overlap slightly may be employed.

The selection of a curve to Represent Trend: — Various types of curves which may be fitted to represent the trend of economic data over a period of time have been described. But which of these many types is to be selected in a given case? Which will give the best standard of normality for each of the years covered?

[A] The first step in deciding what trend type to use should always consist of plotting the observed data. Examination of plotted data will frequently provide an adequate basis for deciding upon the type of trend to use.

i) If the approximate trend, when plotted on arithmatic paper, is a straight line, use a straight line or is a parabola, use a quadratic.

ii) If the approximate trend, when plotted on semi-logarithmic paper, is a straight line, use an exponential curve.

iii) If the approximate trend, when plotted on semi-logarithmic paper, resembles a modified exponential, use a Gompertz curve.

[B] When further guidance is needed, an approximate trend may be drawn by inspection and the tests applied to the smoothed curve:

i) If the first difference (ΔY_t) tend to be constant, use a straight line.

ii) If the second difference ($\Delta^2 Y_t$) tend to be constant, use a second degree polynomial.

iii) If $\frac{\Delta Y_t}{\Delta Y_{t-1}}$ tend to be constant, less than unity, use a modified exponential.

iv) If the first differences of the logarithms ($\Delta \log Y_t$) are constant, use an exponential curve.

v) If the second differences of the logarithms ($\Delta^2 \log Y_t$) are constant, fit a second-degree curve to logarithms.

vi) If the first differences of the logarithms ($\Delta \log Y_t$) are changing by a constant percentage, i.e. if $\frac{\Delta \log Y_t}{\Delta \log Y_{t-1}}$ tend to be constant. Use a Gompertz curve.

Remark:- The method of curve fitting by the principle of least squares is used quite often in trend analysis particularly when one is interested in making projection for future values.

Merits and Limitations of Trend fitting by the Principles of Least squares:—

Merits:-

1. Because of its mathematical or analytical character, this method completely eliminates the element of subjective judgement or personal bias on the part of the investigator.
2. Unlike the method of moving averages, this method enables us to compute the trend values for all the given time periods in the series.
3. The trend equation can be used to estimate or predict the values of the variable for any period t in future or even in the intermediate periods of the given series and the forecast values are also quite reliable.
4. The curve fitting by the principle of least squares is not only technique which enables us to obtain the rate of growth per annum, for yearly data, if linear trend is fitted.

Demerits:—

1. The method is quite tedious and time-consuming as compared with the other methods. It is rather difficult for a non-mathematical person (layman) to understand and use.
2. The addition of even a single new observation necessitates all calculations to be done afresh.
3. Future prediction or forecasts based on this method are based only on long run variations, i.e. trend and completely ignore the cyclical and seasonal and irregular fluctuations.
4. The most serious limitation of the method is the determination of the type of the trend curve to be fitted, viz., whether we should fit a linear or a parabolic trend or some other more complicated trend curve.

Analysing series which contain a trend : Applicability

The determination of trend in a TS depends on wheather the data exhibit seasonality. With seasonal data, it is a good idea to start by calculating successive yearly averages as these will provide a simple description of underlying trend. A traditional method of dealing with non-seasonal data which contain a trend, particularly yearly data, is to fit a simple function such as a polynomial curve, gompertz curve.

The determination of Monthly or Quaterly Trend Values :—

In some cases, it is preferable to obtain the trend from annual data, since the presence of a very violent seasonal movement may distort a trend fitted to monthly data.

(i) Let $T_t = a+bt$ be the trend equation for the annual data, with origin at $t=0$, with 't' unit in one year. If the annual data employed in the fitting process are total of monthly values, then 'a' is the trend value when $t=0$, to obtain the trend value for the month centering at the origin year, we divide a by 12. The constant 'b' defines the change due to trend over a 12 month period, we obtain the monthly trend increment in yearly totals. To obtain the unit of time in month, we replace t by $\frac{t}{12}$. Hence $T_t = \frac{a}{12} + \frac{b}{12} \cdot \frac{t}{12}$ is the monthly trend equation, with origin July 1 of the year of the origin, with t units in one month. Now, the monthly trend equation is $T_t = \frac{a}{12} + \frac{b}{12} \cdot \frac{(t+\frac{1}{2})}{12}$, where the origin is at July 15 of the year ~~[illegible]~~ of the origin and the units of 't' is one month.

Similarly, if the annual data employed in the fitting process are averages of monthly values, then the trend for monthly data is $T_t = a + \frac{b.(t+1/2)}{12}$, where the origin is at july 15 of the year of the origin, and the units of 't' is one month.

(ii) Let $T_t = a + bt + ct^2$ be the trend equation for the annual data, with origin at $t=0$ and with 't' units in one year.

If the annual data employed in the fitting are yearly totals, then the monthly trend equation is,

$$T_t = \frac{a}{12} + \frac{b}{12} \cdot \frac{t+1/2}{12} + \frac{c}{12} \cdot \left(\frac{t+1/2}{12}\right)^2,$$ the origin is at july 15 of the year of the origin and the unit of t is one month.

If annual data employed are averages of monthly values, the trend for monthly data is $T_t = a + b\left(\frac{t+1/2}{12}\right) + c\left(\frac{t+1/12}{12}\right)^2$, the origin is at July 15 of the year of the origin, the unit of t is one month.

(iii) Let $T_t = ab^t$ be the trend for annual data, with origin at $t=0$ with 't' units in one year.

If the annual data employed in the fitting are yearly totals, the monthly trend equation is $\log T_t = \frac{\log a}{12} + \frac{\log b}{12} \cdot \frac{t+1/2}{12}$

$$\Leftrightarrow T_t = a^{1/12} \left(b^{1/12}\right)^{\frac{t+1/2}{12}},$$ with origin at July 15 of the year of the origin, with 't' units in one month.

If annual data employed in the fitting are averages of monthly values, the monthly trend equation is

$$T_t = ab^{\frac{t+1/2}{12}}$$, with origin July 15 of the year of the origin with 't' units of one month.

4. <u>Method of moving Averages:</u>— Let us suppose that we are given with a time series data $\{y_t \mid t=1,2,\dots,n\}$. Then a k points weighted moving average value based on k values $Y_{t+1}, Y_{t+2}, \dots, Y_{t+k}$ of the given time series is defined as

$$M[\omega_1,\omega_2,\dots,\omega_k]\{Y_{t+1},\dots,Y_{t+k}\} = \sum_{j=1}^{k}\omega_j Y_{t+j}, \text{ where, } \left(\sum_{j=1}^{k}\omega_j = 1\right), \text{and,}$$

$M[\omega_1,\omega_2,\dots,\omega_k]$ is k-points <u>moving average operator</u> with $\omega_1,\omega_2,\dots,\omega_k$ as weights. Obviously,
$M[\omega_1,\omega_2,\dots,\omega_k]\{Y_{t+1}, Y_{t+2},\dots,Y_{t+k}\}$ will correspond to the time point $\left(t+\frac{k+1}{2}\right)$.

Accordingly, the moving average values corresponding to a time series value Y_t will be

$$M[\omega_1,\omega_2,\dots,\omega_k]\left\{Y_{t-\frac{k-1}{2}}, Y_{t-\frac{k-1}{2}+1},\dots,Y_{t-1}, Y_t, Y_{t+1},\dots\dots, Y_{t+\frac{k-1}{2}}\right\}$$
$$= \sum_{j=-\frac{k-1}{2}}^{\frac{k-1}{2}} \omega_j Y_{t+j}, \text{ when } k \text{ is odd.}$$

Now, when k is even, then a k-point weighted moving average corresponding to Y_t may be defined as

$$M[\omega_1,\omega_2,\dots,\omega_k]\left\{\left(\frac{Y_{t-k/2}+Y_{t-k/2+1}}{2}\right), \left(\frac{Y_{t-k/2+1}, Y_{t-\frac{k}{2}+2}}{2}\right), \dots\dots, \left(\frac{Y_{t-1}+Y_t}{2}\right), \left(\frac{Y_t+Y_{t+1}}{2}\right),\dots, \left(\frac{Y_{t+\frac{k}{2}-1}, Y_{t+k/2}}{2}\right)\right\}$$
$$= \sum_{j=1}^{k}\omega_j\left(\frac{Y_{t-\frac{k}{2}+j-1}+Y_{t-\frac{k}{2}+j}}{2}\right) = \frac{1}{2}\left[\sum_{j=1}^{k}\omega_j Y_{t-\frac{k}{2}+j-1} + \sum_{j=1}^{k}\omega_j Y_{t-\frac{k}{2}+j}\right]$$

In particular, when $\omega_1=\omega_2=\dots=\omega_k=\frac{1}{k}$, then the moving averages are called <u>simple moving average</u> of extent k, or, k-point simple MA.

27

Describe the MA method of trend fitting. What would be the effect on moving average series if the original series undergo a base and scale change?

- The simple moving average of period k of a time-series gives us a series of arithmatic means, each of k consecutive observations. We start with the first k observations. At the next stage, we leave the first and include the (k+1)st observation. This process is repeated until we arrive at the last k observations. Each of these means is centred against the time which is the mid-point of the time interval included in the calculation of the moving average. Thus, when the period of moving average k is odd, the MA value corresponds in time to a value which is actually observed. When the period is even, the MA falls mid-way between two observed values. In this case, we calculate a subsequent two-item moving average to make the resulting moving average values correspond to the observed value.

Let $Y_1, Y_2, \ldots, Y_n$ be a time series data. Then the 1st MA of period k is $m_1 = \frac{Y_1+Y_2+\cdots+Y_k}{k}$ and 2nd MA is

$$m_2 = \frac{Y_2+Y_3+\cdots+Y_{k+1}}{k}.$$

Note that, $m_2 - m_1 = \frac{Y_{k+1}-Y_1}{k}$, which is positive or negative according as $Y_{k+1} > Y_1$ or $Y_{k+1} < Y_1$. Hence, moving averages follow the increasing or decreasing pattern of the TS data, on the average. It is based on the data as given; if the general trend changes, the moving averages follow the new trend. Hence, it is a flexible measure of trend, adapting itself to changing conditions.

- If $\{y_t \mid t=1,2,\ldots,n\}$ be a given TS, then the k-point weighted MA based on k values $y_{t+1},\ldots,y_{t+k}$ of the given series is defined as

$$M_k[\omega_1,\ldots,\omega_k]\{y_{t+1},\ldots,y_{t+k}\}$$
$$= \sum_{j=1}^{k} \omega_j y_{t+j}, \text{ where } \sum_{j=1}^{k}\omega_j = 1.$$

$M[\omega_1,\ldots,\omega_k]$ is k point MA operator with ω_j as weights and this this MA value will correspond at time $\left(t+\frac{k+1}{2}\right)$.

$\omega_k = \frac{1}{k} \Rightarrow$ simple MA of extent k.

Let $\{x_t\}$ be a time series data, changing base and scale as follows: let $y_t = \frac{x_t - a}{b}$.

Moving average of k items:

$$m_1(y) = \frac{\sum_{t=1}^{k} y_t}{k} = \frac{\sum_{t=1}^{k} x_t - ka}{kb} = \frac{m_1(x) - a}{b}.$$

i.e. if the original series undergo a base and scale change, then the moving average series show a constant change in base and scale. That is, if the original series has the MA value $m_1(x)$, then changing the base an scale, we shall get for the new series the MA will be $\frac{m_1(x) - a}{b}$, when letting $y_t = \frac{x_t - a}{b}$.

Notion of Growth Curves: So far we have considered trend equations which are either polynomials or reducible to polynomial. In such cases the standard practice is to use least square method for fitting an appropriate curve. But in many cases we have to use trend equations which are neither polynomials nor reducible to polynomials. These occurs, say in fitting growth (decline) curves, studying the growth of a new industry or the growth of population in a country. We have to use growth curves, which are of these types. In such cases, we donot use the least squares method, other methods are recommended. Perhaps of even greater general utility, in the analysis of time series, are curves of a semi-logarithmic (exponential) type.

Ex. Exponential curves, Gompertz curves.

PROPERTIES OF MOVING AVERAGE METHOD :-

(1). A moving average method whose period is equal to that of a periodic series completely wipes out the periodicity present in the data. / If the timeseries data contains a periodic movement, then an appropriately chosen simple MA eliminates the periodicity present in the data. / What is the effect of MA on cyclical component?

Proof :- Let us consider a periodic series $\{Y_t\}$ whose period is K,

i.e. $Y_t = Y_{t+k} = Y_{t+2k} = \cdots \quad \forall \; t = 1, 2, \ldots$

Consequently, we have

$$\frac{Y_1 + Y_2 + \cdots + Y_K}{K} = \frac{Y_2 + Y_3 + \cdots + Y_{K+1}}{K} = \cdots\cdots \text{ and so on.}$$

But these are simple K-point moving averages of extent (period) k that the MA values are all equal showing no periodicity. Hence the proof.

(2). If in a TS the underlying trend is concave (convex) upward then an MA will overestimate (underestimate) the trend values.

Proof :- Let us suppose that $Y(t)$ be the given time series which is composed of only the trend component and random component $\ni$

$$Y(t) = T(t) + \epsilon(t)$$

If we take $(2K+1)$ point weighted moving average, $\cancel{M[\omega_1, \omega_2, \ldots, \omega_k}$ where $(2K+1)$ is sufficiently large, having weights

$$\{\omega_j \mid j = -K, -K+1, \ldots, 0, 1, 2, \ldots, K\} \text{ such that } \sum_{j=-K}^{K} \omega_j = 1.$$

Then

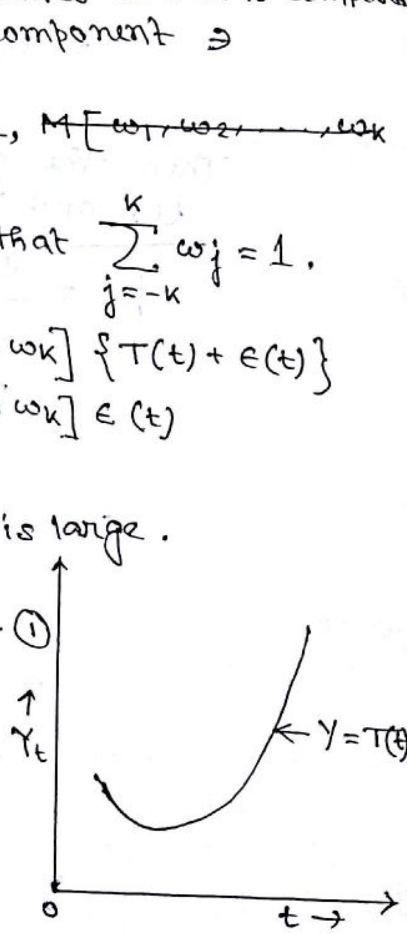
$$M[\omega_{-K}, \ldots, \omega_0, \ldots, \omega_K]\, Y(t) = M[\omega_{-K}, \ldots, \omega_K]\, \{T(t) + \epsilon(t)\}$$

$$= M[\omega_{-K}, \ldots, \omega_K]\, T(t) + M[\omega_{-K}, \ldots, \omega_K]\, \epsilon(t)$$

$$= M[\omega_{-K}, \ldots, \omega_K]\, T(t)$$

since $M[\omega_{-K}, \ldots, \omega_K]\, \epsilon(t) = 0$ as $(2K+1)$ is large.

i.e. $$\sum_{j=-k}^{k} \omega_j \, Y(t+j) = \sum_{j=-k}^{k} \omega_j \, T(t+j) \quad \cdots\cdots (1)$$



When the underlying trend $T(t)$ is concave upward then we have

$$T\left(\sum_{j=-K}^{K} \omega_j \,(t+j)\right) \le \sum_{j=-K}^{K} \omega_j \, T(t+j) = \sum_{j=-K}^{K} \omega_j \, Y(t+j) \quad [\text{By } (1)]$$

But the LHS is the trend value corresponding to a time point t and RHS is $(2K+1)$ point weighted moving averages ω_j, where $j = -K(1)K$ and $\sum_j \omega_j = 1$.

So, the $(2K+1)$ point weighted MA overestimates the trend values.

(3) Effect of moving average on random or irregular Component : Slutsky - Yule Effect :

Let us suppose a time-series data $\{Y_t : t = 1, 2, \cdots\}$ which is composed of only random or irregular components.

i.e. $Y_t = \epsilon_t$ where $E(\epsilon_t) = 0 \ \forall t$

and $\text{cov}(\epsilon_t, \epsilon_{t'}) = \begin{cases} \sigma^2 & \text{whenever } t = t' \\ 0 & \quad '' \quad t \neq t' \end{cases}$

Let us suppose that the TS $\{U_t\}$ is generated by an m-point moving average with weights $\omega_1, \omega_2, \cdots, \omega_m$, say.

i.e. $U_{t+\frac{m+1}{2}} = \sum_{j=1}^{m} \omega_j Y_{t+j}$, where $\sum_{j=1}^{m} \omega_j = 1$.

$$\Rightarrow U_t = \sum_{j=1}^{m} \omega_j Y_{t-\frac{m+1}{2}+j}$$

$$\Rightarrow U_t = \sum_{j=1}^{m} \omega_j \epsilon_{t-\frac{m+1}{2}+j}$$

Then the auto-correlation (or serial correlation) ρ_s between U_t and U_{t+s} is, —

$$\rho_s = \frac{\text{cov}(U_t, U_{t+s})}{\text{Var}(U_t)}$$

$$= \frac{E\left\{\sum_{j=1}^{m} \omega_j \epsilon_{t-\frac{m+1}{2}+j}, \sum_{j=1}^{m} \omega_j \epsilon_{t+s-\frac{m+1}{2}+j}\right\}}{E\left\{\sum_{j=1}^{m} \omega_j \epsilon_{t-\frac{m+1}{2}+j}\right\}^2}$$

$$= \frac{E\left(\sum_{j=1}^{m} \omega_j \epsilon_{r+j}, \sum_{j=1}^{m} \omega_j \epsilon_{r+s+j}\right)}{E\left(\sum_{j=1}^{m} \omega_j \epsilon_{r+j}\right)^2}$$

$$\rho_s = \begin{cases} \dfrac{\sum_{j=1}^{m-s} \omega_j \omega_{j+s} E(\epsilon^2_{r+j})}{\sum_{j=1}^{m} \omega_j^2 E(\epsilon^2_{r+j})} & \text{if } s < m \\ 0 & \text{if } s \geq m \end{cases}$$

$$= \begin{cases} \dfrac{\sigma^2 \sum_{j=1}^{m-s} \omega_j \omega_{j+s}}{\sigma^2 \sum_{j=1}^{m} \omega_j^2} & \text{if } s < m \\ 0 & \text{if } s \geq m \end{cases}$$

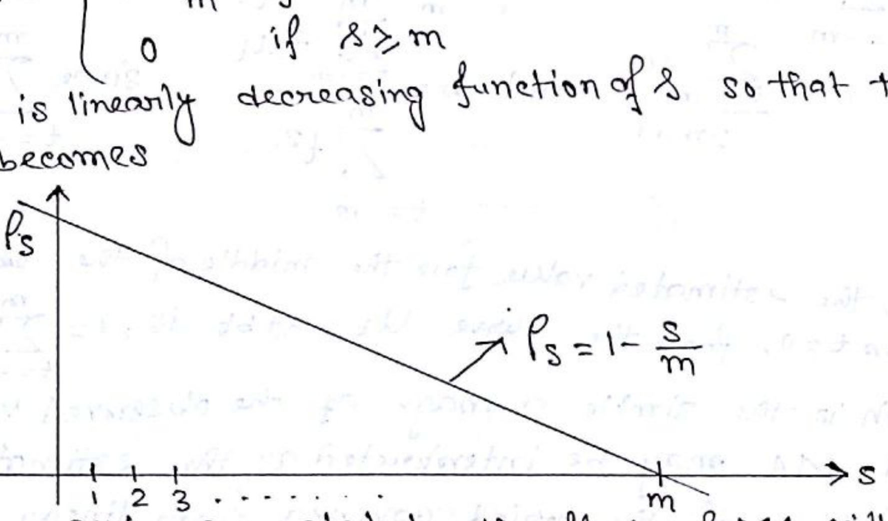
$$= \begin{cases} \dfrac{\sum_{j=1}^{m-s} \omega_j \omega_{j+s}}{\sum_{j=1}^{m} \omega_j^2} & \text{if } s < m \\ 0 & \text{if } s \geq m \end{cases}$$

Now if ω_j's are all equal, i.e. $\omega_1 = \omega_2 = \cdots = \omega_m = \frac{1}{m}$ then

$$\rho_s = \frac{m-s}{m}$$

$$\therefore \rho_s = \begin{cases} 1 - \frac{s}{m} & \text{if } s < m \\ 0 & \text{if } s \geq m \end{cases}$$

Therefore, ρ_s is linearly decreasing function of s so that the correlogram i.e. (s, ρ_s) plot becomes



This shows that a series generated by the effect of MA will not be uncorrelated so long as the period of the moving average m is greater than the distance or long s between u_t and u_{t+s}. Hence, u_t will be much smoother series than the random series I_t.

Thus the effect of taking a moving average of the random component could be to generate a spurious oscillatory series, provided the correlation between the successive members of the generated series is positive. This effect is generally known as Slutsky - Yule effect.

Relation between moving average method and polynomial trend:
Interpretation of Moving average method :—

Suppose that the number of terms is chosen to be odd from a TS data and is denoted by $(2m+1)$. Without loss of generality, we may denote the terms by $u_{-m}, u_{-(m+1)}, \dots, u_0, \dots, u_{m-1}, u_m$.

1. Justify the following statement: Application of k-point simple MA process on a TS is equivalent to fitting straight lines to successive k-values by method of least squares and finding the mid-values of the fitted lines.

Justification:— If we choose to fit to them a linear ~~trend~~ curve $T_t = a+bt$, we may determine the coefficients by the method of least squares, i.e. by solving the equations:

$$\frac{\partial}{\partial a}\sum_{t=-m}^{m}(u_t - a - bt)^2 = 0 \quad \text{and} \quad \frac{\partial}{\partial b}\sum_{t=-m}^{m}(u_t - a - bt)^2 = 0$$

$$\Leftrightarrow \sum_{t=-m}^{m} u_t = (2m+1)a + b\sum_{t=-m}^{m} t \quad \text{and} \quad \sum_{t=-m}^{m} t u_t = a\sum_{t=-m}^{m} t + b\sum_{t=-m}^{m} t^2$$

$$\Leftrightarrow a = \frac{\sum_{t=-m}^{m} u_t}{2m+1} \quad \text{and} \quad b = \frac{\sum_{t=-m}^{m} t u_t}{\sum_{t=-m}^{m} t^2}, \text{ since } \sum_{t=-m}^{m} t = 0.$$

Hence, the estimated value for the middle of the period covered i.e. for $t=0$, from the curve $u_t = a+bt$ is $a = \sum_{t=-m}^{m} u_t / (2m+1)$, which is the simple average of the observed values. A $(2m+1)$ point-MA may be interpreted as the estimated value for the middle of the period covered from linear curves (straight lines) fitted through the 1st $(2m+1)$ points, through 2nd to the $(2m+2)$th values, and lastly through the last $(2m+1)$ points.

2. Let $\{u_t\}$ be a given TS data, now if one fits a polynomial trend $P(t) = a_0 + a_1 t + \dots + a_p t^p$ to the set of K values of the TS, then fitting of polynomial trend will be equivalent to the weighted MA method where the weights are symmetric and sum of the weights is unity.

<u>Proof</u>:- $K = 2m+1$ (let).

If we choose to fit to them a polynomial of degree p, we may determine the coefficients by least squares, i.e. by solving the equations,

$$\frac{\partial}{\partial a_j} \sum_{t=-m}^{m} (u_t - a_0 - a_1 t - \dots - a_p t^p)^2 = 0 \quad \forall\, j = 0(1)p.$$

$$\Leftrightarrow \sum_{t=-m}^{m} t^j u_j = a_0 \sum t^j + a_1 \sum t^{j+1} + \dots + a_p \sum t^{j+p}, \quad j = 0(1)p.$$

Assume: p is even, since $\sum t^j = 0$ if j is odd, a_0 will occur in the following equations only:

$$\left.\begin{aligned}
\sum u_t &= (2m+1)a_0 + a_2 \sum t^2 + a_4 \sum t^4 + \dots + a_p \sum t^p \\
\sum t^2 u_t &= a_0 \sum t^2 + a_2 \sum t^4 + a_4 \sum t^6 + \dots + a_p \sum t^{p+2} \\
&\vdots \\
\sum t^p u_t &= a_0 \sum t^p + a_2 \sum t^{p+2} + a_4 \sum t^{p+4} + \dots + a_p \sum t^{2p}
\end{aligned}\right\} \longrightarrow (*)$$

If we write

$$\Delta = \begin{vmatrix} 2m+1 & \sum t^2 & \dots & \sum t^p \\ \sum t^2 & \sum t^4 & \dots & \sum t^{p+2} \\ \vdots & \vdots & & \vdots \\ \sum t^p & \sum t^{p+2} & \dots & \sum t^{2p} \end{vmatrix} \;\&\; \Delta_0 = \begin{vmatrix} \sum u_t & \sum t^2 & \dots & \sum t^p \\ \sum t^2 u_t & \sum t^4 & \dots & \sum t^{p+2} \\ \vdots & \vdots & & \vdots \\ \sum t^p u_t & \sum t^{p+2} & \dots & \sum t^{2p} \end{vmatrix},$$

then solving (*) for a_0, we shall get $a_0 = \frac{\Delta_0}{\Delta}$.

Then $a_0 = \frac{1}{\Delta}\left\{\Delta_{11} \cdot \sum u_t + \Delta_{21} \cdot \sum t^2 u_t + \dots \right\}$, where Δ_{ij} is the co-factor of the element in the i^{th} row and j^{th} column of Δ.

The weight of u_t i.e. the coefficient of u_t in a_0 is given by

$$C_t = \frac{1}{\Delta}\left\{\Delta_{11} + t^2 \Delta_{21} + \Delta_{31} t^4 + \dots \right\}.$$

Note that, Δ_{11}, Δ_{21}, etc. depend only on 'm' and are independent of the part of the series. Note that, $C_t = C_{-t}$; i.e. weights are symmetrical.

Again, $\sum_{t=-m}^{m} c_t = \frac{1}{\Delta}\left\{(2m+1)\Delta_{11} + \Delta_{21}\cdot\Sigma t^2 + \Delta_{31}\cdot\Sigma t^4 + \cdots\right\}$ (34)

$= \frac{1}{\Delta}\cdot\Delta$, expanding Δ by the first column.

$= 1$; that is, the sum of weights is unity.

Hence, a_0 can be expressed as $a_0 = c_{-m}u_{-m} + c_{-\overline{m-1}}u_{-\overline{m-1}} + \cdots + c_0u_0 + c_1u_1 + \cdots + c_mu_m$.

Now, a_0 is the estimated value of u_t for $t=0$. As we see, this is equivalent to a weighted average of the observed values, the weights being independent of which part of the series is taken.

Example:- Suppose we have a series and we wish to fit a curve which best approximates to sets of seven points; and suppose we regard a cubic as providing a satisfactory approximation. What are the weights of the MA?

Solution: Consider the seven values as $u_{-3}, u_{-2}, u_{-1}, u_0, u_1, u_2$ & u_3.

Our polynomial is $u_t = a_0 + a_1t + a_2t^2 + a_3t^3$, say,

Normal equations are:

$$\begin{aligned}
\Sigma u_t &= 7a_0 \quad\quad + 28a_2 \\
\Sigma t u_t &= \quad\quad 28a_1 \quad\quad + 196a_3 \\
\Sigma t^2 u_t &= 28a_0 + \quad\quad 196a_2, \\
\Sigma t^3 u_t &= \quad\quad 196a_1 \quad\quad + 1588a_3,
\end{aligned}$$

Since, for odd j, $\sum_j t^j = 0$.

Solving for a_0, we get $a_2 = \frac{1}{21}\{7\cdot\Sigma u_t - \Sigma t^2 u_t\}$

$$= \frac{1}{21}\{-2u_{-3} + 3u_{-2} + 6u_{-1} + 7u_0 + 6u_1 + 3u_2 - 2u_3\}$$

Remark:- It will be observed that in this example we should have obtained the same value for a_0 if we fitted quadratics instead of cubics, for a_0 does not depend on a_3 in the normal equation. In general, the case p odd includes the case of the next even value of p.

Discuss the method of Group average in fitting modified exponential trend equation to a TS data.

Method of Partial sums:– The types of equation we have considered above will explain trend in a majority of the cases. Occasionally, however, it may be necessary to consider more complicated trend equations. One such is the modified exponential equation:

$$T_t = K + ab^t$$

The curve approaches K as an upper limit if a is negative and approaches K as a lower limit if a is positive. To determine the constants of the curve, the whole range of t covered by the data is divided into three equal parts, each including, say, m points of time. Equating the totals,

$$S_1 = \sum_{1}^{m} Y_t, \quad S_2 = \sum_{m+1}^{2m} Y_t, \quad S_3 = \sum_{2m+1}^{3m} Y_t$$

then three equations are obtained, viz.

$$S_1 = \sum_{t=1}^{m} (K + ab^t) = mK + ab \times \frac{1-b^m}{1-b},$$

$$S_2 = mK + ab^{m+1} \times \frac{1-b^m}{1-b},$$

$$S_3 = mK + ab^{2m+1} \times \frac{1-b^m}{1-b}.$$

The three equations are now solved for the three unknowns: K, a and b. The values will be found to be

$$b = \left(\frac{S_2 - S_3}{S_1 - S_2}\right)^{1/m}$$

$$a = \frac{(S_1 - S_2)(1-b)}{b(1-b^m)^2}$$

and $$K = \frac{1}{m} \times \frac{S_1 S_3 - S_2^2}{S_1 - 2S_2 + S_3}.$$

Gompertz curve:–

$$T_t = Ka^{b^t},$$

or, $$\log T_t = \log K + b^t(\log a)$$

$\log T_t$ being of the modified exponential form.

Similar to the above case.

Remark on Moving Average Method :-

To summarise, moving average method gives a correct picture of the long term trend of the series if

i) the trend is linear or approximately linear.
ii) oscillatory movements affecting the data are regular in period and amplitude.

If trend is not linear, moving averages introduce bias in the trend values.

Moving average method is very flexible in the sense that the addition of a few more figures to the data simply results in some more trend values; the previous calculations are not affected at all.

The MA method has the following drawbacks:

i) It does not provide trend values for all the terms, e.g., for a moving average of extent $2k+1$, we have to forego the trend values for the first k and last k terms of the series.
ii) It can't be used for forecasting or predicting future trend, which is main objective of trend analysis.

Ques:- Which component of time series is mainly applicable in the following cases?

i) Fire in factory. Ans:→ Random Component.
ii) Decrease in employment in a sugar factory during the off season.→ Seasonal
iii) Fall in death rate due to scientific research. → Trend.
iv) An era of prosperity. → cyclical component.
v) An after Easter sale in a department store. → Seasonal component.
vi) A need for increased wheat production due to a constant increase population. ⟶ Long-term trend.

Measurement of Seasonal Variations :—

Any time series are subject to periodic fluctuations, seasonal in character, and these fluctuations are generally more important in their effects, upon business than the long time trend. Our present concern is with methods of isolating such seasonal variations.

We shall consider seasonal variation in monthly or quaterly data only, but the procedure for weekly or daily or hourly data will be quite similiar.

Estimation of Seasonal Component

Simple Average Method

(No trend)

$TS = \{X_t \mid t = 1, 2, \ldots, n\}$

Step I: Estimate trend by an appropriate method.

Step II: Detrend the data: $x - \hat{T}_t = u_t$, where the component u_t is: $u_t = S_t + I_t$

Merits & Demerits :

This method is based on the basic assumption that the data don't have any trend and cyclical components and consists in eliminating irregular component by averaging the monthly values over years. Since most of the economic TS have trends, these assumptions in general are not true. Though simple, this method is not of much practical utility.

Ratio to trend method

(Trend is determined by mathematical curve fitting)

Merits & Demerits :

Since the method irons out the cyclical or irregular components, but if the series exhibits pronounced cyclical swings, then MA method will be appropriate one.

The obvious advantage of this method over the MA method lies in the fact that 'ratio to trend' can be obtained for each month for which the data are available and as such, unlike the 'ratio to MA' method, there is no loss of data.

Ratio to MA method

(Trend is determined by moving average)

Merits & Demerits :-

Of all the methods of measuring seasonal variation the ratio to MA method is the most satisfactory, flexible, and widely used method. These indices do not fluctuate so much as the indices by the ratio to trend method.

This method does not completely utilize the data, e.g. in case of 12-month MA seasonal indices can't be obtained for the first 6 months and for the last 6 months.

(A) Averages of adjusted data / Method of monthly (or, quaterly) averages:

When the data do not contain cyclical movements or trend to any appreciable extent, the irregular variation may be eliminated by averaging the monthly / quaterly / weekly, etc. values over years or different time intervals. Let Y_{ij} be the observation for the jth month in the ith year. As the data $\{Y_{ij}\}$ contains no trend or cyclical component, the irregular variation is eliminated by computing the averages $\bar{y}_j = \sum_i Y_{ij} / (\text{no. of years})$, $j = 1(1)12$. Clearly, $\bar{y}_j$ measures the seasonal component for the jth month.

To express the averages $\bar{y}_j$ as seasonal indices, they are shown as percentages of the grand mean, i.e. seasonal index for the jth month $= \frac{\bar{y}_j}{\bar{\bar{y}}} \times 100$, where $\bar{\bar{y}} = \frac{1}{12} \sum_{j=1}^{12} \bar{y}_j$.
Therefore, the total of seasonal indices is 1200 for monthly data (or, 400 for quaterly data).

For additive model, the grand mean is subtracted from the monthly (or, quaterly) averages to obtain the seasonal indices which in this case will add upto zero.

(B). Ratio to trend method: A somewhat similar method of securing seasonal indices, which has certain distinctive advantages, involves averaging the ratio to trend. In the application of this method, a suitable equation of trend, linear or non-linear, is fitted to the data, the actual monthly data are expressed as percentages of the corresponding trend figures.

Let Y_{ij} be the observation for the jth month of the ith year. First determine a trend equation based on yearly data and obtain the monthly trend values. Let T_{ij} be the trend value obtained from the trend equation for the jth month in the ith week. Under the assumption that the TS model is multiplicative we compute the ratios to trend $r_{ij} = \frac{Y_{ij}}{T_{ij}} \times 100$.

The ratios or percentages r_{ij} measures the seasonal component with a part of the irregular component. These ratios show some variation from year to year in the relation of the figures for a given month. The different ratios for each month are then averaged so that irregular fluctuations may be removed; i.e. we compute.

$$r_j = \sum_i r_{ij} / (\text{number of years})$$

If $\sum_{j=1}^{12} r_j \neq 1200$, the seasonal indices of monthly variations are then obtained by adjusting the r_j's to add upto 1200. The adjustment factor $\lambda = \frac{1200}{\left(\sum_j r_j\right)}$. Then the required seasonal indices are given by $S_j = \lambda r_j$, $j = 1(1)12$. The way it is constructed S_j's must add upto 1200.

(C) Ratio to moving average method :- Moving average provides a useful method of defining seasonal variations. For the purpose of discussion we assume that we have a monthly series of observations: Y_{ij} be the observation for the j^{th} month in the i^{th} year. Assuming multiplicative model, the data Y_{ij} is composed of $T \times C \times S \times I$. So from monthly data, we compute moving averages, say y_{ij}, by taking a 12-month MA and which is again centred by taking 2-point MA. The ratios of the original values (Y_{ij}) to the MA values (y_{ij}) are expected to present the seasonal variation with a part of the irregular variations, i.e. the ratios $r_{ij} = \frac{Y_{ij}}{y_{ij}}$ measure the seasonal variation and a part of irregular variation: $\frac{T \times C \times S \times I}{T \times C} = S \times I$.

Theses ratios, one for each month except for 6 months at the beginning and 6 months at the end, are expressed as percentage. These percentages $\left(r'_{ij} = \frac{Y_{ij}}{y_{ij}} \times 100\right)$ show some variation from year to year in the relation of the figures for a given month. The values r'_{ij} for each month are averaged so that irregular fluctuations may be removed; that is, we compute $r_j = \sum_i r'_{ij} / \{\text{Number of Years}\}$. If $\sum_{j=1}^{12} r_j \neq 1200$, the seasonal indices of monthly variations are then obtained by adjusting the r_j's to add upto 1200. The adjustment factor is $\lambda = \frac{1200}{\sum_j r_j}$ and the required seasonal indices are given by $S_j = \lambda . r_j$, $j = 1(1)12$. The way it is constructed S_j's must add up to 1200.

For the additive model, the moving averages are substracted from the original values and the deviations for a month (quater) are averaged over the years. The monthly (quarterly) average deviations are finally adjusted so that the total seasonal values becomes zero.

Example :-

(1) Calcutta has shown no appreciable change in the total annual rainfall over the years. Discuss how you will find the seasonal variation in rainfall in Calcutta given the last-five years, monthly data.

Solution :- When the data do not contain trend or cyclical movements to any appreciable extent, but only seasonal and irregular fluctuations, then it will suffice to use the "Method of monthly average" (i.e. to average the data without making any previous adjustment).

Multiplicative model: Here the irregular variation may be eliminated by averaging the monthly data of rainfall over years, for each month. Let, Y_{ij} be the rainfall for the j^{th} month in the i^{th} year, $i=1(1)5$, $j=1(1)12$. Here $\bar{y}_j$ measures the seasonal component and is given by $\bar{y}_j = \frac{1}{5}\sum_{i=1}^{5} Y_{ij} \;\forall\; j=1(1)12$.

Seasonal index for the j^{th} month is obtained as $I_j = \frac{\bar{y}_j}{\bar{\bar{y}}} \times 100$, where $\bar{\bar{y}} = \sum_{i=1}^{5}\sum_{j=1}^{12} Y_{ij} / (5 \times 12) = \sum_{j=1}^{12} \bar{Y}_j / 12$, the grand mean; so that the total of the seasonal indices is $12{,}00$; i.e.

$$\sum_{j=1}^{12} I_j = 100 \sum_{j=1}^{12}\left(\frac{\bar{y}_j}{\bar{\bar{y}}}\right) = 100\left(\sum_{j=1}^{12}\bar{y}_j\right)/\bar{\bar{y}} = 1200$$

For Additive model, the grand mean $(\bar{\bar{y}})$ is subtracted from the monthly averages $(\bar{y}_j)$ to obtain the seasonal values, which in this case will add upto zero.

(2) The daily flow of traffic was observed for the 365 days of a year on a particular road of the city. The data revealed that
(i) the flow varied over the different days of a week and
(ii) the flow had a gradual increase over the weeks. Describe in details how the data should be analysed.

Solution:- When the data do not contain cyclical movement to an appreciable extent but the data exhibits secular trend, a seasonal index is computed by "Ratio-to-trend method". The given data of an illustration of such data consisting of secular trend, seasonal variation but the cyclical variation is given to be absent.

Let Y_{ij} be the data on the flow of traffic for the j^{th} day in the i^{th} week, $j=1(1)7$, $i=1(1)52$ (excluding the one extra day). In this method, the first step consists of determining a trend equation, for the data on weekly basis and obtain the daily trend equation. Let y_{ij} be the trend value obtained from the trend equation for the j^{th} day in the i^{th} week.

Under the assumption that the TS model is multiplicative, we compute the ratios to trend,

$$r_{ij} = \frac{Y_{ij}}{y_{ij}} \times 100.$$

These ratios r_{ij} measure the seasonal component apart from any random fluctuations. We take the average of r_{ij}'s over the weeks,

$$r_j = \sum_{i=1}^{52} r_{ij} / (\text{No. of years}) = \frac{\sum_{i=1}^{52} r_{ij}}{52},$$ so that irregular fluctuations may be removed, $j=1(1)7$. The seasonal indices of weekly variations are then obtained by adjusting the r_j's to add upto 700. Here the adjustment factor $\lambda = \frac{700}{\sum_j r_j}$.

Then the required seasonal indices are given by:

$$S_j = \lambda . r_j \,, \; j=1(1)7, \text{ so that } \sum_{j=1}^{7} S_j = \sum_{j=1}^{7} \lambda . r_j = \lambda \sum_{j=1}^{7} r_j = 700$$

For additive model, the seasonal component is first isolated by computing $r_{ij} = Y_{ij} - y_{ij}$; which are nothing but the observations in the residual series consisting of seasonal and irregular component. Then r_j's are obtained by averaging the r_{ij}'s over the weeks as in the first procedure. The seasonal indices are obtained by adjusting the r_j's to add up to zero. For that purpose define $\mu = \frac{1}{7}\sum_{j=1}^{7} r_j$. The seasonal indices are given by $S_j = (r_j - \mu)$, $j = 1(1)7$. It is clear that S_j's must add upto zero.

Problem (3). The sales of a company rose from Rs. 40,000 in March to Rs. 48,000 in april 1984. The companies seasonal indices for these two months are 105 and 140 respectively. The owner of the company expressed dissatisfaction with the April sales, but the sales manager said that he was quite pleased with the Rs. 8,000 increase. What argument should the owner of the company have used to reply to the sales manager?

The sales manager also predicted on the basis of the April sales that the total 1984 sales were going to be Rs. 5,76,000. Critiaise the sale's Manager's estimate and explain how the estimate of 4,11,000 may be arrived at.

Hints:- The sales manager did not take into account the seasonal indices of March and April. On the basis of March sales, the owner's estimates of sales of April 1984 keeping in view the seasonal indices are:

$$\text{Rs. } \frac{40,000}{105} \times 140 = \text{Rs. } 53333 \cdot 33$$

Since company's sales of April viz. 48,000 are (53,333·33 − 48,000) = 5333·33 less than the estimated sales, owner's dissatisfaction is justified. Sales Managers estimates of sales on the basis of april sales is:

$$\text{Rs. } 48000 \times 12 = \text{Rs. } 5,76,000.$$

Now, April's actual sales = Rs. 48,000

April's seasonal index = 140

$\therefore$ April's seasonal effect $= \frac{140}{100} = 1{\cdot}40$

Hence April's estimated sales = (April's actual sales) ÷ seasonal effect

$$= \frac{48{,}000}{1.4} \text{ Rs.}$$

Therefore, on the basis of April sales, the estimated annual sales for 1984 are:

$$\text{Rs. } \frac{48{,}000}{1.4} \times 12 = \text{Rs. } 4{,}11{,}428.57 \simeq \text{Rs. } 4{,}11{,}000$$

In the above discussion, we have used multiplicative model of TS.

Random Component in a Time Series :—

As the definition suggests, no formula, however approximate, can be obtained to measure the random component directly at any given point of the series. Usually the non-random components are determined and then a random residual which is left unaccounted for by these components is obtained. Even this becomes difficult when oscillations appear in the series. However, the Variate Difference method enables us to estimate the variance of the random component in a series.

The Variate Difference Method :—

The concept of a series which consists of a polynomial element plus a residual of a more or less random kind has given rise to a method which eliminates the polynomial element by differencing. Clearly, successive differencing will eventually entirely eliminate any element which is actually a polynomial in the time, and may be relied upon almost to eliminate any systematic element except, perhaps, exponential or cyclical terms.

Suppose the time series $\{X_t \mid t = \ldots, -2, -1, 0, 1, 2, \ldots\}$ is known to be of the form $x_t = \sum_{i=0}^{K-1} \beta_i t^i + \epsilon_t = \mu_t + \epsilon_t$, where β_i's are unknown and ϵ_t has mean '0' and variance v for all t.

Here, K is unknown, one way of estimating it is by a semi-empirical procedure known as <u>variate-difference method</u> which works as follows:

Now, $\Delta^h x_t = \Delta^h \mu_t + \Delta^h \epsilon_t$, where Δ is the forward difference operator.

and $\Delta^h \epsilon_t = (E-1)^h \epsilon_t = \epsilon_{t+h} - \binom{h}{1}\epsilon_{t+h-1} + \binom{h}{2}\epsilon_{t+h-2} \cdots + (-1)^h \epsilon_t$.

Since $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = v$, we find

$$E(\Delta^h \epsilon_t) = 0 \text{ and } Var(\Delta^h \epsilon_t) = v \sum_{i=0}^{h} \binom{h}{i}^2 = \binom{2h}{h} v.$$

Now we consider the samples $(x_1, x_2, \ldots, x_n)$ and from the ratios

$$Q_h = \frac{1}{n-h} \sum_{t=1}^{n-h} (\Delta^h x_t)^2 \Big/ \binom{2h}{h}, \quad h = 1, 2, 3, \ldots$$

Now, $E(\Delta^h x_t)^2 = E(\Delta^h \mu_t + \Delta^h \epsilon_t)^2 = E(\Delta^h \mu_t)^2 + E(\Delta^h \epsilon_t)^2$

$= (\Delta^h \mu_t)^2 + V(\Delta^h \epsilon_t)$, since $E(\Delta^h \epsilon_t) = 0$.

$= (\Delta^h \mu_t)^2 + \binom{2h}{h} \vartheta$, and

$$E\left[\frac{1}{n-h}\sum_{t=1}^{n-h}(\Delta^h x_t)^2\right] = \frac{1}{n-h}\sum_{t=1}^{n-h}(\Delta^h \mu_t)^2 + \binom{2h}{h}\vartheta$$

Therefore, $E[Q_h] = \frac{1}{(n-h)\binom{2h}{h}} \sum_{t=1}^{n-h} (\Delta^h \mu_t)^2 + \vartheta = R_h + \vartheta$, say.

It will be noted that R_h is non-negative and vanishes for all $h \geq k$. Thus $E\, Q_h = \vartheta$ for all $h \geq k$.

This, in a practical situation, one could calculate Q_h, $h = 1, 2, 3, \ldots$ and choose as the estimate of k the value of h of which Q_h appears to become constant (except for small random fluctuations) and use that Q_h as an estimate of ϑ.

Suppose in practical situation, Q_{K^*} remains constant within sampling limits, then the estimate of k is K^* and $E\, Q_{K^*} \simeq \vartheta$, i.e. an estimate of ϑ is $Q_{K^*} = \frac{1}{n-K^*}\sum_{t=1}^{n-K^*}(\Delta^{K^*} x_t)^2 \Big/ \binom{2K^*}{K^*}$

<u>Remark</u>:- More important, perhaps, is the fact that differencing a series which is not a polynomial plus a random residual may also give values of $S_h = \sum_{t=1}^{n-h} (\Delta^h x_t)^2$ which decrease up to a point and hence suggest that a polynomial model would be appropriate. In some studies it is found that sequences of S_h which could have indicated the suitability of a polynomial model, although the real model was entirely different. If, in fact, there are oscillatory movements in the series which extend over several time periods, the differencing process tends to regard them as trend and will eliminate their effect. <u>For this reason, we suggest that variate-differencing should be regarded as giving a lower limit to k, the order of polynomial fit, but not as providing a decisive ruling that a polynomial model is appropriate.</u>

Auto correlation and Correlogram :

An important guide to the properties of a time series is provided by a series of quantities called the sample auto correlation coefficient or serial correlation. They measure the correlation (if any) between observations of different distances (i.e. at different time points) apart and provide useful descriptive informations. This is also an important tool in model building and often provide valuable tools to the suitable probability model for a given set of data.

DEFINITION :— Given n observations $x_1, x_2, \dots, x_n$ on a time series, $(n-k)$ pairs of values (x_t, x_{t+k}) can be formed with a lag of period k (i.e. observations in each pair are seperated by k time interval). Regarding the first observation of each pair as one variable and the 2nd one as another variable, the correlation coefficient (r_k) between x_t and x_{t+k} defined as,

$$r_k = \frac{Cov(x_t, x_{t+k})}{\{Var(x_t)\, Var(x_{t+k})\}^{1/2}}$$

$$= \frac{\frac{1}{n-k}\sum_{t=1}^{n-k} x_t x_{t+k} - \frac{1}{(n-k)^2}\left(\sum_{t=1}^{n-k} x_t\right)\left(\sum_{t=1}^{n-k} x_{t+k}\right)}{\left\{\frac{1}{n-k}\sum_{t=1}^{n-k} x_t^2 - \frac{1}{(n-k)^2}\left(\sum_{t=1}^{n-k} x_t\right)^2\right\}^{1/2} \cdot \left\{\frac{1}{n-k}\sum_{t=1}^{n-k} x_{t+k}^2 - \frac{1}{(n-k)^2}\left(\sum_{t=1}^{n-k} x_{t+k}\right)^2\right\}^{1/2}}$$

and is called the serial correlation of order k, $k = 0, \pm 1, \pm 2, \dots$.

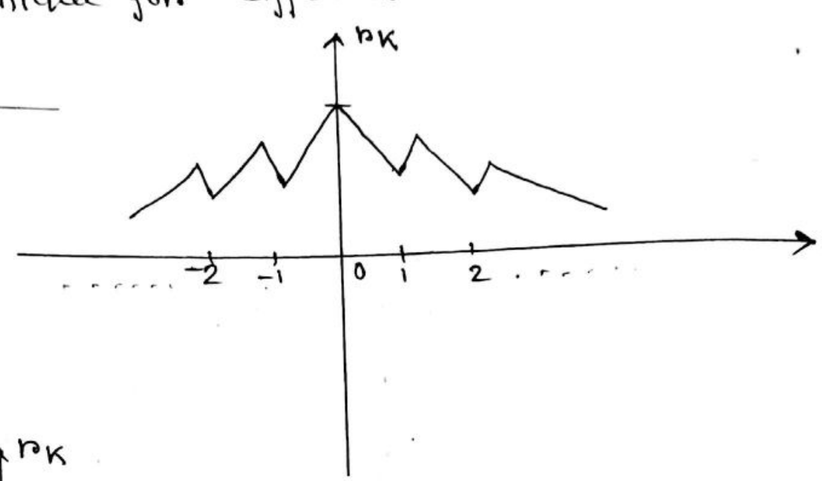
Obviously, $\boxed{r_0 = r = 1 \text{ and } r_{-k} = r_k}$ $\forall\, k = 1, 2, 3, \dots$.

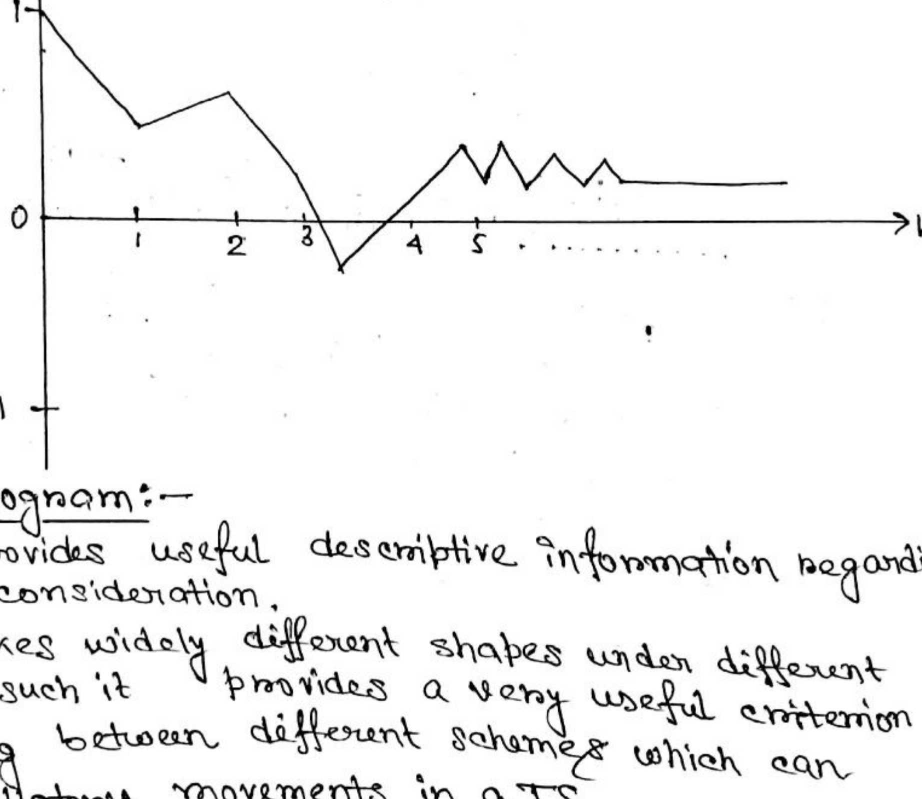
Further for varying values of k, r_k may be treated as a function of k. As such r_k is often called auto-correlation function (a.c.f.) of the given TS. If it is based on sample, then is called sample auto-correlation function. If it is based on population data, then it is called population auto correlation function.

Correlogram:-

DEFINITION:- A correlogram is a graph obtained by plotting the auto correlation coefficient (r_K) against the lag K for $K = 0, \pm1, \pm2, \ldots$ Correlogram is a useful aid in interpreting a set of auto correlation coefficients and visual inspection of the correlogram is often very helpful. The correlogram may alternatively be called the sample auto-correlation function (a.c.f.). Correlogram is unique for different series.

K	r_K
0	1
± 1	r_1
± 2	r_2
$\vdots$	$\vdots$





Usefulness of Correlogram:-

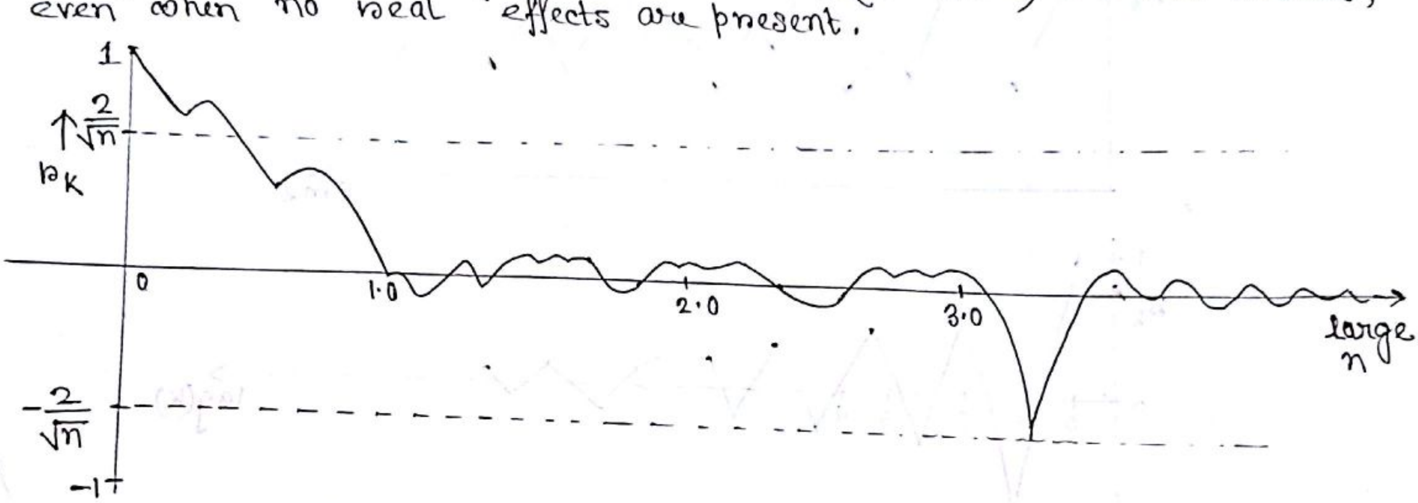
i) Correlogram provides useful descriptive information regarding the TS under consideration.

ii) Correlogram takes widely different shapes under different schemes and as such it provides a very useful criterion for discriminating between different schemes which can account for oscillatory movements in a TS.

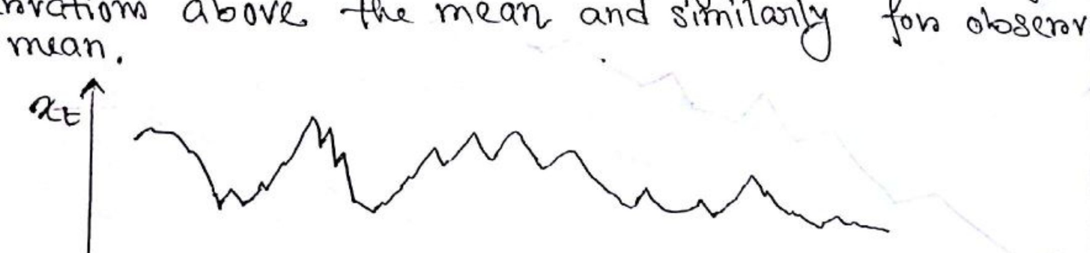
iii) Correlograms are important tool in model building, and often provide valuable clues to a suitable probability model for a given set of data.

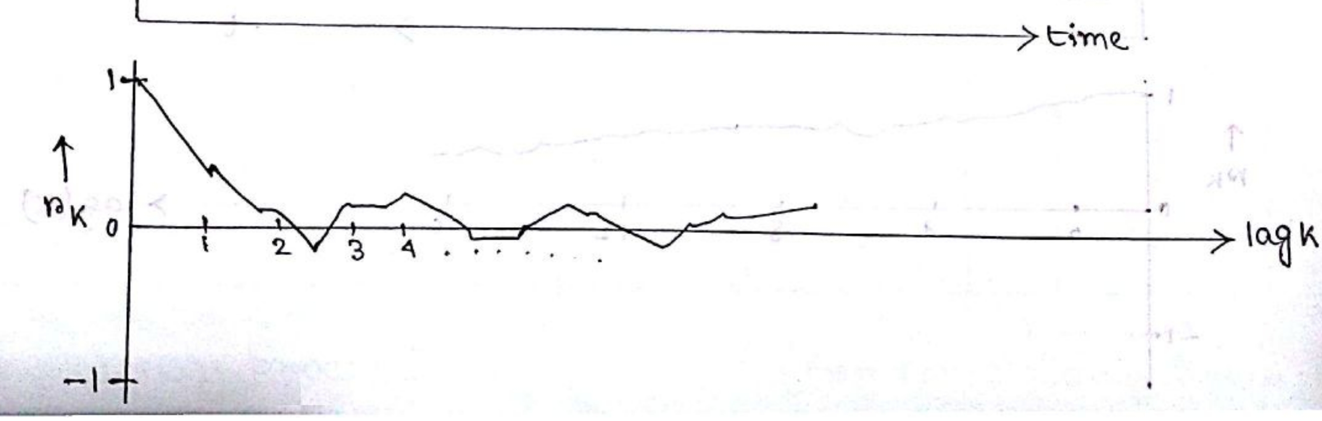
Interpreting the correlogram :—

(a) Correlogram of Random Series :- A TS is said to be completely random if it consists of a series of observations having the same distribution. Then for large n, it is expected that $r_K \simeq 0 \; \forall K=1,2,\ldots$ In fact, for a random series r_K is approximately $N(0,\frac{1}{n})$. Thus if a TS is random, then one can expect on an average 95 out of 100 values of r_K to lie between $\pm 2/\sqrt{n}$. This spotlights one of the difficulties in interpreting correlogram, in that a large number of coefficients quite likely to contain one (or more) unusual results, even when no real effects are present.

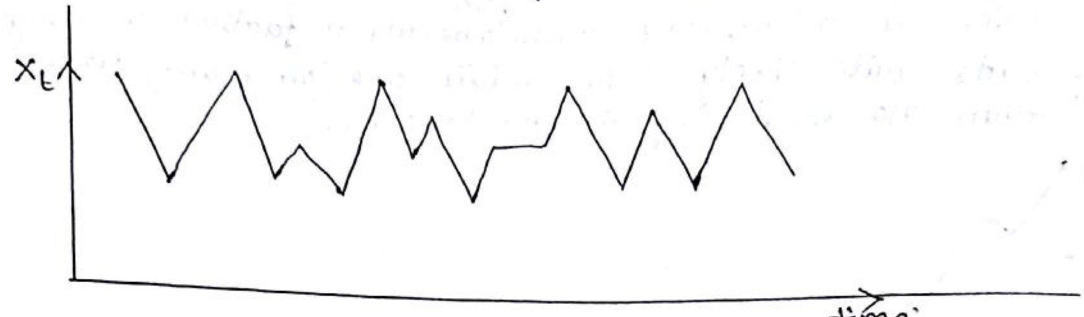


(b) Correlogram of a TS having short-term correlation :-
Stationary TS often exhibit short-term correlation characterized by a fairly large value of r_1 followed by one or two further coefficients, which, while greater than zero, tend to get successively smaller. Values of r_K for longer lags tend to be approximately zero. A TS that gives such a correlogram is one for which an observation above mean tends to be followed by one or more further observations above the mean and similarly for observations below the mean.

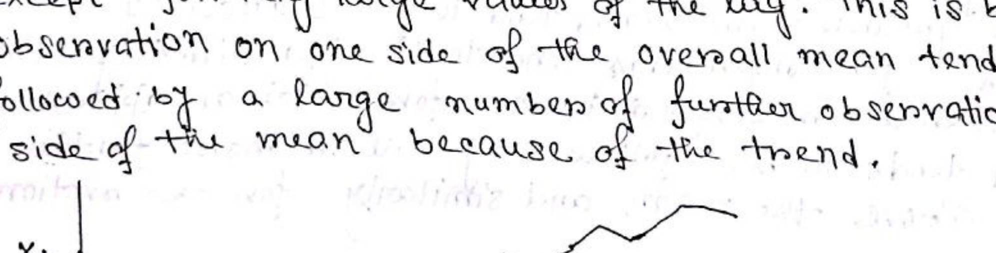


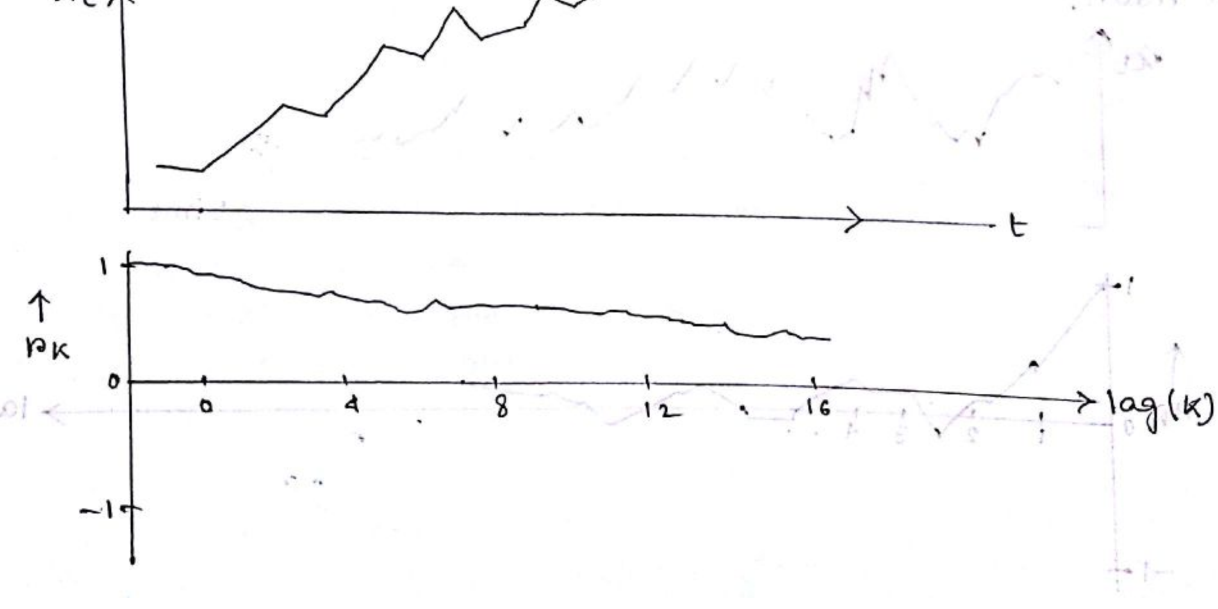


(c) Correlogram of an alternating series: If a TS has a tendency to alternate, with successive observations on different sides of the overall mean, then the correlogram also tends to alternate. This is beacuse of the successive values are on the opposite sides of the mean. The value of r_1 will be negative, but the value of r_2 will be positive, as observation at lag 2 will tend to be on the same side of the mean.



(d) Correlogram of non-stationary series: If a TS contains a trend, then the values of r_K will not come down to zero except for very large values of the lag. This is because an observation on one side of the overall mean tends to be followed by a large number of further observations on the same side of the mean because of the trend.

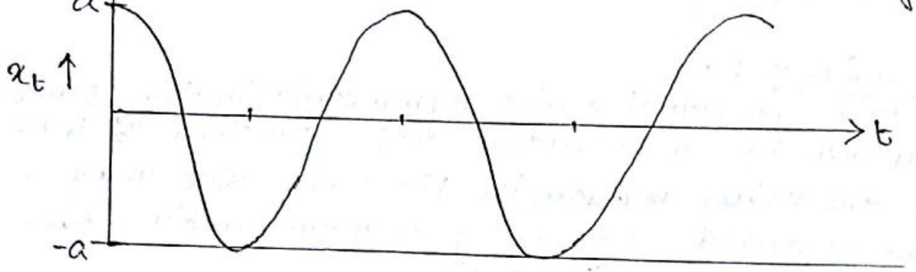


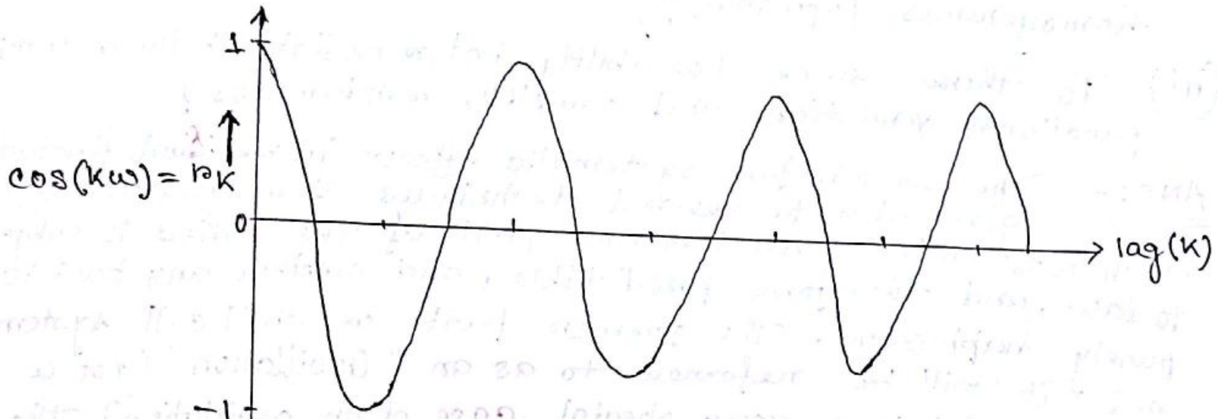


(e) Correlogram of seasonal series :- If a TS contains seasonal variation, then the correlogram will also exhibit oscillation at the same frequency. In particular, if x_t follows a sinusoidal pattern, then so does r_K. For example, if

$$x_t = a\cos t\omega ,$$

where a is a constant and the frequency ω is such that $0<\omega<\pi$, then it can be seen that $r_K \simeq \cos K\omega$ for large N.

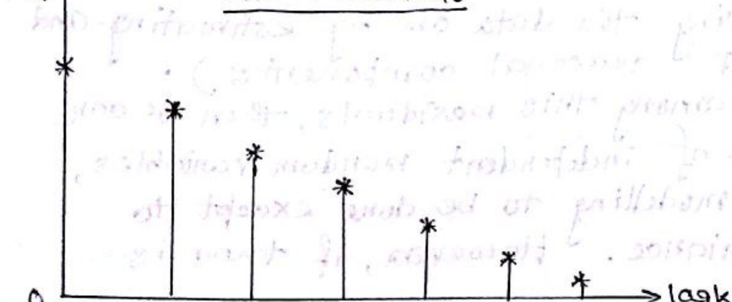




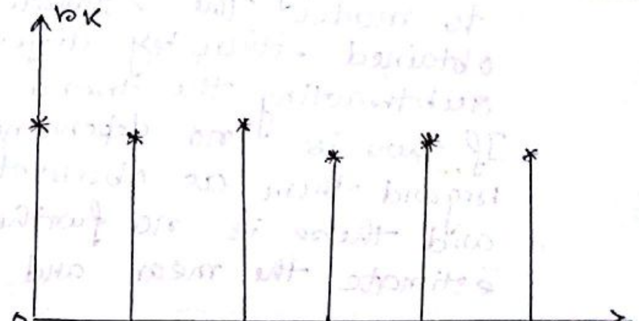
Remark:-

If the noise sequence (i.e. residuals of a TS after eliminating or removing trend and seasonal components) does have sample auto correlation (r_K) significantly different from zero. Then we can take advantage of this serial dependence to forecast future noise values in terms of the past values by modelling the noise sequence as a stationary TS.

For practical :-



$r_K \to 0$ as $K \to \infty$
(For stationary process)



$r_K \not\to 0$ as $K \to \infty$
(For non-stationary Process)

Residual Series :- The elimination of trend from TS data is called the detrending the TS, and elimination of seasonal variation from the TS data is called the deseasonalizing the TS data.

When seasonal variation and trend have been removed from the data, we are left with a series which will present, in general, fluctuations of a more or less regular kind, is called residual series.

- Ques arises :-
  (i) Is this residual series symmetric in the sense that its values can be represented as a function of time?
  (ii) Are the values random in the sense that they could occur, in the observed order, by random sampling from a homogeneous population?
  (iii) Is there some possibility intermediate between complete functional variation and complete randomness?

ANS:- The search for systematic effects in residual fluctuation gives rise to several techniques of analysis, the object of which is to detect whether any part of the series is subject to law, and therefore predictable, and whether any part is purely haphazard. The former part we shall call systematic and it will be referred to as an "Oscillation" (not a 'Cycle', which is a very special case of an oscillation). The remainder of the series we shall call the unsystematic component, and refer to its movements as "random".

Testing the estimated noise sequence (Test of randomness) :-

The main objective of estimating and extracting or removing the deterministic components, trend and seasonality from the given TS is to produce a series with no apparent deviations from stationarity, and in particular, with no apparent trend and seasonality. Assuming that this has been done, the next step is to model the estimated noise sequence (i.e. the residuals obtained either by differencing the data or by estimating and subtracting the trend and seasonal components).
If there is no dependence among this residuals, then we can regard them as observations of independent random variables, and there is no further modelling to be done except to estimate the mean and variance. However, if there is

significant dependence among residuals, then we need to look for a more complex stationary TS model for noise that accounts for this dependence. This will be to our advantage since dependence means in particular that the past observations of the noise sequence can assist in predicting future values.

In most cases, visual examination of a graph of TS is enough to see that the series is not random. However, it is occasionally desirable to assess whether an apparently stationary TS is "random". One type of approach is to carry out what is called a "Test of randomness", in which one tests whether the TS residuals $x_1, x_2, \ldots, x_n$ could have arisen in that order by chance by taking a simple random sample of size n from a population assumed to be stationary but with unknown characteristics. Various tests exist for this purpose. But we examine some simple tests for checking the hypothesis that the residuals are observed values of independent and identically distributed random variables. If they are, then our work is done, but if not, then we must use the theory of stationary processes to find a more appropriate model.

(a) The sample autocorrelation function test :—

For large n, sample auto correlation (r_K) of an i.i.d. sequence $Y_1, Y_2, \ldots, Y_n$ with finite variance are approximately i.i.d. $N(0, \frac{1}{n})$. Hence if $y_1, y_2, \ldots, y_n$ are the realization of such an i.i.d. sequence, then we reject the i.i.d. hypothesis at 5% level of significance if

$$|r_K| > \frac{1.96}{\sqrt{n}} \text{ for } K = 1, 2, 3, \ldots$$

(b) The turning point test :— If $y_1, y_2, \ldots, y_n$ is a sequence of observations, we say that there is a turning point at time t, $1 < t < n$, if $y_{t-1} < y_t > y_{t+1}$, or $y_{t-1} > y_t < y_{t+1}$.

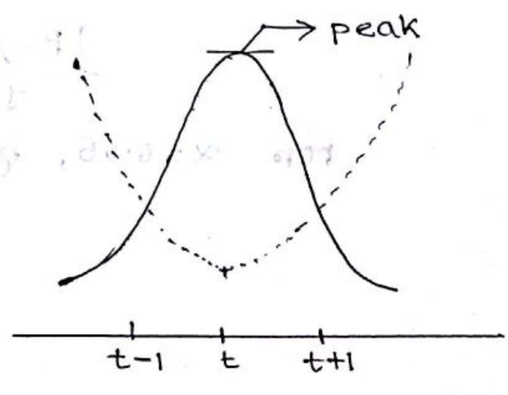
If T is the total number of turning points of an i.i.d. sequence of length n, then here the mean is,

$$\mu_T = E(T) = \frac{2}{3}(n-2), \text{ say}$$

and the variance of T is,

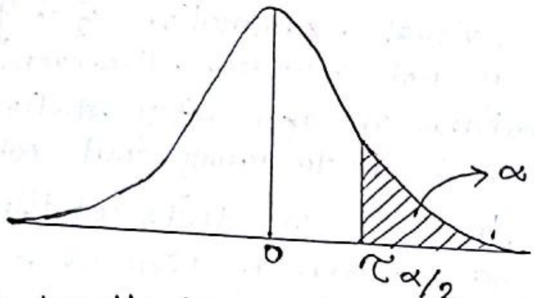
$$\sigma_T^2 = Var(T) = \frac{16n-29}{90}$$

and that for large values of n the distn. of T is approximately $N(\mu_T, \sigma_T^2)$.



Accordingly, we reject the i.i.d. hypothesis at level of significance α if

$$\frac{|T-\mu_T|}{\sigma_T} > \tau_{\alpha/2}$$, where $\tau_{\alpha/2}$ is the upper $100 \times \frac{\alpha}{2}$ % of $N(0,1)$ distribution.



$$P\left[T > \tau_{\alpha/2} \mid \tau \sim N(0,1)\right] = \frac{\alpha}{2}$$

In particular, at 5% level (i.e. at $\alpha = 0.05$), we reject the i.i.d. hypothesis if

$$\frac{|T-\mu_T|}{\sigma_T} > 1.96 \quad \text{for } \alpha = 0.05, \tau_{\alpha/2} = \tau_{0.025} = 1.96.$$

(c) <u>The rank test of Kendall</u>:– Let us define P to be the number of pairs (i,j) such that $y_j > y_i$ and $j > i$, $i = 1(1)\overline{n-1}$, (i.e. there is a positive concordance). There is a total of $\binom{n}{2} = \frac{n(n-1)}{2}$ pairs (i,j) such that $j > i$. For an i.i.d. sequence $\{Y_1, Y_2, \ldots, Y_n\}$, each event $\{Y_j > Y_i\}$ has probability $\frac{1}{2}$ and the mean of P is therefore,

$$\mu_P = \frac{1}{4} n(n-1),$$

variance of P is,

$$\sigma_P^2 = Var(P) = \frac{1}{72} n(n-1)(2n+5)$$

and for large n,

$$P \sim N(\mu_P, \sigma_P^2).$$

Accordingly, we reject the i.i.d. hypothesis at level α (i.e. $Y_1, Y_2, \ldots, Y_n$ are random) if

$$\frac{|P-\mu_P|}{\sigma_P} > \tau_{\alpha/2}.$$

For $\alpha = 0.05$, $\tau_{\alpha/2} = \tau_{0.025} = 1.96$.

Alternatively,

Describe Kendall's test for randomness.

ANS:- Given an observed series of observations $y_1, y_2, \dots, y_n$, can they have arisen by chance in that order by sampling independently on 'n' occasions from a population of unknown characteristics ?

Given the set of values $y_1, y_2, \dots, y_n$, in that order let us count the number of pairs in which $y_j > y_i$, $j > i$. If this is P, we note that there is $\frac{n(n-1)}{2}$ pairs and that the expected number of P in a random series is $\frac{n(n-1)}{4}$. The excess of P over this number; i.e. $\frac{n(n-1)}{4}$, indicates a tendency to positive trend, a definciency corresponding to a negative trend.

Hence the quantity $\tau = \frac{4P}{n(n-1)} - 1$ can be taken as a measure of randomness of the series $y_1, y_2, \dots, y_n$; and this quantity is the Kendall's rank correlation coefficient τ, between the order of the variables in time and their order in magnitude Y_t.

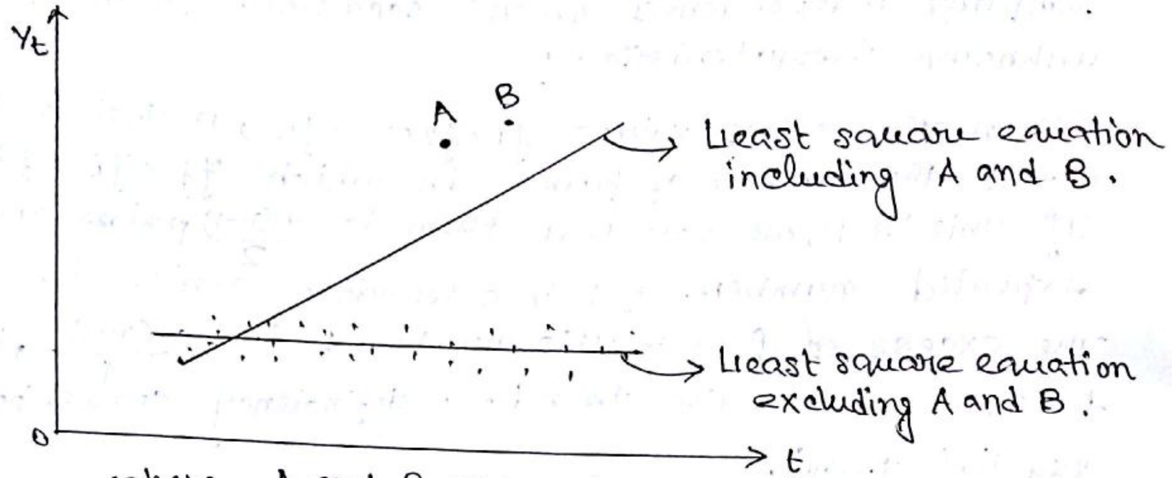
If Q is the complementary quantity to P, namely the number of values for which $y_j < y_i$, $j > i$, then we have
$$P + Q = \binom{n}{2} \text{ and } \tau = 1 - \frac{4Q}{n(n-1)}.$$

[For a random series, $E(\tau) = 0$ and $Var(\tau) = \frac{2(2n+5)}{9n(n-1)}$.
The distribution of τ tends rapidly to normality – in fact,
$$\frac{\tau - E(\tau)}{\sqrt{V(\tau)}} = \frac{\tau}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}} \overset{a}{\sim} N(0,1), \text{ as } n \to \infty.]$$

Problems of "extreme outliers" in data set while fitting Polynomial trend :

Suppose few points are wide apart from the majority points on either end.
Suppose the trend line (linear) to be fitted and is given by —

$$Y_t = a + bt.$$



where, A and B are extreme outliers.

where, $\hat{b} = \sum_{t=-k}^{k} t y_t \Big/ \sum_{t=-k}^{k} t^2$, $n = 2k+1$.

$$= \frac{\underbrace{-k y_{-k}}_{\text{dominates}} + \cdots + (-1).y_{-1} + \underbrace{0.y_0}_{\text{no influence}} + 1.y_1 + \cdots + \underbrace{k.y_k}_{\text{dominates}}}{\sum_{t=-k}^{k} t^2}$$

Slope of the least square line will be dominated by the extreme outliers and not by the central values as

$$\hat{b} = \sum_{t=-k}^{k} t y_t \Big/ \sum_{t=-k}^{k} t^2, \text{ where } n = 2k+1.$$

Note: — $Y_t = a + bt$ (linear trend)

$$\Rightarrow \frac{dY_t}{dt} = b = \text{constant}$$

$Y_t = ab^t$ (exponential trend)

$$\Rightarrow \ln Y_t = \ln a + t \ln b$$

$$\Rightarrow \frac{1}{Y_t} \cdot \frac{dY_t}{dt} = \ln b = r = \text{constant} = \text{Relative growth rate} = R_t.$$

i.e. $R_t = r$

In discrete case $= \frac{Y_{t+1} - Y_t}{Y_t} = (b-1) = r$ (unit of $t = 1$)

e.g. If $b = 1.02$, then $R_t = 2\%$.

The Probability Models for Time Series : Stationary TS

Introduction : A time-series may be defined as a collection of random variables which are ordered in time and defined at a set of time-points. Let $(\Omega, \mathcal{A}, P)$ be a probability space. A time-series is a real valued function $X_t(\omega)$ defined on Ω and on the index set T.

If future values of a TS are exactly determined by some mathematical function such as

$$X_t = a+bt,\ X_t = ab^t,\ X_t = e^{a+bt} \text{ or } X_t = \cos(2\pi f t), \text{etc.}$$

the time series is said to be deterministic. If the future values can be described only in terms of a probability distribution, the TS is said to be non-determinstic or simply a statistical TS. Here, it is not possible to forecast the exact value at a future time point. It is with such statistical time series that we are concerned in this topic.

If we remove from a time-series the deterministic elements attributable to trend and seasonal variation we shall, in general, be left with a series oscillating about some constant value, is called Residual Series. Here, we shall restrict ourselves in this topic to oscillatory behaviour which is taken to be purely stochastic.

Stochastic Processes :—

Most physical processes in the real world involve a random element in their structure and a "STOCHASTIC PROCESS" can be described as a statistical phenomenon that evolves in time according to probabilistic laws, e.g., the length of a queue, the number of accidents in a particular town in successive months, the air temperature at a particular site on successive days, etc. Since it is impossible to make more than one observation at any given time, we only have a single observation on the r.v. at time t [i.e. $X_t(\omega)$] and a single out come of the process $\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\}$.

- **DEFINITION**:– A stochastic process is a family (collection) of random variables $\{X(t)\}$ or $\{X_t\}$ on a probability space $(\Omega, \mathcal{A}, P)$ that are ordered in time and defined at a set T of time point t, which may be continuous or discrete. If time is continuous, i.e. in case

$T = \{t \mid -\infty < t < \infty\}$, the stochastic process is said to be a continuous-parameter stochastic process and is denoted by $\{X(t)\}$.

If time is discrete, i.e. in case

$T = \{t \mid t = 0, \pm 1, \pm 2, \pm 3, \ldots\ldots\}$, the stochastic process is said to be a discrete-parameter stochastic process or, a stochastic sequence and is denoted by $\{X_t\}$.

$X(t)$ or X_t is said to refer as the state of the process at time t.

Any observed TS is to be regarded as an observation or realisation of a stochastic process.

Ensemble and Realization of a Stochastic Process:— In TS analysis, we may regard the observed TS as just one example of the infinite set of TS that might have been observed. This infinite set of time series is sometimes called the ensemble. Every member of the ensemble is a possible realization of the stochastic process.

Stochastic Process and Time Series:— The observed TS can be thought of as one particular realization of a stochastic process $X(t)$ and will be denoted as $x(t)$ for $(0 \le t \le T)$ if time is continuous, and by x_t for $t = 1, 2, \ldots, N$ if time is discrete. Time Series analysis is essentially concerned with evaluating the properties of the underlying probability model from this observed time series, even though this single realization is the only one we will ever observe. As such use the terms stochastic process 'vis-a-vis' time series simultaneously.

Properties / Characteristics of Stochastic Process :—

i) **Mean:-** The mean function $\mu(t)$ of a stochastic process $\{X(t)\}$, continuous in time, is defined for all t, by

$$\mu(t) = E[X(t)],$$

provided the corresponding expectation exists for all t,
Similarly for a stochastic process $\{X_t\}$, discrete in time. Mean-function is defined by

$$\mu(t) = E(X_t) \text{ for all } t,$$

provided the expectation exists for all t.

ii) **Variance:-** The variance function $\sigma^2(t)$ of stochastic process $\{X(t)\}$, continuous in time is defined for all t, by

$$\sigma^2(t) = Var(X(t))$$
$$= E[X(t) - \mu(t)]^2$$

When $\{X_t\}$ is discrete in time, the variance function σ_t^2 of stochastic process,

$$\sigma_t^2 = Var(X_t)$$
$$= E[X_t - \mu_t]^2$$

iii) **Auto-covariance function:-** The auto-covariance function (a.c.v.f.) $\gamma(t_1, t_2)$ of stochastic process $X(t)$ is defined for all t_1, t_2 to be as the covariance between $X(t_1)$ and $X(t_2)$, i.e.

$$\gamma(t_1, t_2) = Cov[X(t_1), X(t_2)]$$
$$= E[\{X(t_1) - \mu(t_1)\}\{X(t_2) - \mu(t_2)\}]$$

clearly, the variance function is a special case of a.c.v.f. when $t_1 = t_2$, i.e. $\sigma^2(t) = \gamma(t, t)\ \forall t$.

iv) **Auto correlation function:-** The size of an auto-covariance coefficient $\gamma(\tau)$ depends on the units in which $X(t)$ is measured. Thus, for interpretative purposes, it is helpful to standardize acv.f. to produce a function called the auto-correlation function (a.c.f.).
The auto correlation at lag 'τ', the correlation between X_t and $X_{t+\tau}$, is

$$\rho(\tau) = \frac{cov(X_t, X_{t+\tau})}{\sqrt{Var(X_t)\,Var(X_{t+\tau})}} = \frac{\gamma(\tau)}{\gamma(0)}.$$

. This quantity measures the correlation between $X(t)$ and $X(t+\tau)$.
Since for a stationary process, the variance $V(X_t) = \sigma^2 = V(X_{t+\tau})$ for all t and $\gamma(0) = Cov(X_t, X_t) = Var(X_t) = \sigma^2$.

$$\therefore \rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$$

$$\therefore \rho(0) = \frac{\sigma^2}{\sigma^2} = 1.$$

<u>Stationary Stochastic process</u>: — A special class of stochastic processes, called stationary processes, is based on the assumption that the process is in a particular state of <u>statistical equilibrium</u>. A stochastic process is said to be <u>strictly stationary</u> if its properties are unaffected by a change of time origin, if the joint probability distn. associated with k observations: $X(t_1), X(t_2), \ldots, X(t_k)$ is the same as that associated with k observations: $X(t_1+\tau), X(t_2+\tau), \ldots, X(t_k+\tau)$.

<u>DEFINITION</u>:- A stochastic process (TS) $\{X(t)\}$ is said to be "<u>STRICTLY STATIONARY</u>" if the joint distn. of $X(t_1), \ldots, X(t_k)$ is the same as the joint distribution of $X(t_1+\tau), \ldots, X(t_k+\tau)$ $\forall$ integers $k \geq 0$ and τ. i.e. if

$$\{X(t_1), X(t_2), \ldots, X(t_k)\}' \stackrel{d}{=} \{X(t_1+\tau), X(t_2+\tau), \ldots, X(t_k+\tau)\}',$$

where, $X_1(t_1) \stackrel{d}{=} X_1(t_1+\lambda)$ $\forall$ λ, where '$\stackrel{d}{=}$' is used to indicate that the two random vectors have the same distribution functions.

In other words, shifting the time origin by an amount τ has no effect on the joint distributions, which must, therefore, depend only on the intervals between $t_1, t_2, \ldots, t_k$.

In particular, if $K=1$, then <u>strict stationarity</u> implies that the distribution of $X(t)$ is same for all t, so we have

$$\mu = E(X_t),\ \sigma^2 = \text{var}(X_t)\ \forall t,$$ provided they exist; i.e. the stochastic process has a constant mean μ and a constant variance σ^2.

For $K=2$, the joint distribution of $X(t_1)$ and $X(t_2)$ depend only on the time difference $(t_2-t_1)=\tau$, which is called the lag. Thus the auto covariance function $\gamma(t_1,t_2)$ depends only on (t_2-t_1) and may be written as

$$\begin{aligned}\gamma(t_1,t_2) = \text{Cov}[X(t_1), X(t_2)] &= E[\{X(t_1)-\mu\}\{X(t_2)-\mu\}] \\ &= E[\{X(t_1)-\mu\}\{X(t_1+\tau)-\mu\}] \\ &= \gamma(\tau), \text{ say} \quad [\because t_2 = t_1+\tau]\end{aligned}$$

$\gamma(\tau)$ is called the auto-covariance coefficient at lag τ.

Second Order Stationarity or Weakly stationarity :

We have seen that for a process to be strictly stationary, the whole probability structure must depend only on time differences. In practice, it is often useful to define stationarity in a less restricted way than that described earlier.

A stochastic process is called "WEAKLY STATIONARY" of order 'r' if the moments of the process up to order 'r' depend only on time differences.

DEFINITION:- A stochastic process $\{X_t\}$ is called second-order stationary or weakly stationary if its mean is constant and its auto-covariance function depends only on the lag, so that

i) $E(X_t) = \mu$
ii) $Cov[X_t, X_{t+\tau}] = \gamma(\tau)$, for all t → independent of t and depends only on the lag τ.

No assumption is made about higher moments than those of second order. By letting $\tau = 0$, we have $Cov(X_t, X_t) = \gamma(0)$, i.e. $Var(X_t)$ is constant.

The second-order stationarity plus an assumption of normality in a stochastic process $\{X_t\}$ gives that $\{X_{t_1}, \dots, X_{t_n}\}$ has multivariate normal distribution and multivariate normal distribution is completely characterized by its first and second order moments. For normal processes it follows that second-order stationarity implies strict stationarity.

Remark:-

1. Stationary (strictly) $\Rightarrow$ Weak stationarity
$\not\Leftarrow$
2. Weak stationarity $\Leftrightarrow$ Co-variance stationary.

Properties of a.c.f. of a stationary process: —

Suppose a stationary stochastic process $\{X_t\}$ has mean μ and variance σ^2, auto-covariance $\gamma(\tau)$ and auto-correlation $\rho(\tau)$. Then
$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \frac{\gamma(\tau)}{\sigma^2}$$

Note that, $\rho(0) = 1$.

Property 1.: $\gamma(0) \geq 0$

Proof: This is because
$$\gamma(0) = \text{Cov}[X(t), X(t)] = \text{Var}(X(t)) = \sigma^2 \geq 0$$

Property 2.: $\rho(\tau) = \rho(-\tau)$, i.e. the auto correlation function $\rho(\tau)$ is an even function of the lag 'τ'.

Proof:- This property simply says that the correlation between $X(t)$ and $X(t+\tau)$ is the same as that between $X(t)$ and $X(t-\tau)$.

$$\begin{aligned}\gamma(\tau) &= \text{cov}[X_t, X_{t+\tau}] \\ &= \text{cov}[X_{t-\tau}, X_t], \text{ since } \{X_t\} \text{ is stationary.} \\ &= \text{cov}[X_t, X_{t+(-\tau)}] \\ &= \gamma(-\tau)\end{aligned}$$

Hence, $\dfrac{\gamma(\tau)}{\gamma(0)} = \dfrac{\gamma(-\tau)}{\gamma(0)} \Leftrightarrow \rho(\tau) = \rho(-\tau)$

Property 3.: $|\rho(\tau)| \leq 1$.

Proof:- This is the 'usual' property of correlation, namely, that it lies between ± 1. It is proved by noting that
$\text{Var}[\lambda_1 X(t) + \lambda_2 X(t+\tau)] \geq 0$ for any constants λ_1 & λ_2.

$\Rightarrow \lambda_1^2 \text{Var}[X(t)] + 2\lambda_1\lambda_2 \text{Cov}[X(t), X(t+\tau)] + \lambda_2^2 \text{Var}[X(t+\tau)] \geq 0$

$\Rightarrow (\lambda_1^2 + \lambda_2^2)\sigma^2 + 2\lambda_1\lambda_2\gamma(\tau) \geq 0$, since $X(t)$ is stationary.

When $\lambda_1 = \lambda_2 = 1$, we find $\gamma(\tau) \geq -\sigma^2$, so that $\rho(\tau) \geq -1$

When $\lambda_1 = 1, \lambda_2 = -1$, we find $\sigma^2 \geq \gamma(\tau)$, so that $\rho(\tau) \leq 1$

i.e. $|\gamma(\tau)| \leq \sigma^2$

and $|\rho(\tau)| \leq 1$.

The standardized nature of a correlation coefficient means that the value of $\rho(\tau)$ does not depend on the units in which the TS is measured, as can readily be demonstrated by multiplying all values in a series by the same constant and showing that the resulting autocorrelations are unchanged.

Another way :- This is a 'usual' property of correlation coefficient.

Note that, $Var(X_t \pm X_{t+\tau}) \geq 0$

$$\Rightarrow V(X_t) + V(X_{t+\tau}) \pm 2Cov(X_t, X_{t+\tau}) \geq 0$$

$$\Rightarrow \sigma^2 + \sigma^2 \pm 2\gamma(\tau) \geq 0, \text{ since } \{X_t\} \text{ is stationary.}$$

$$\Rightarrow 1 \pm \rho(\tau) \geq 0$$

$$\Rightarrow -1 \leq \rho(\tau) \leq 1.$$

Property 4. $\rho(\tau)$ is invariant of the change of units in which the Time series is measured.

Property 5. The auto correlation function $\rho(\tau)$ does not uniquely identify the underlying model.

Lack of uniqueness. Although a given stochastic process has a unique covariance structure, the converse is not in general true.

Property 6. The auto-covariance function $\gamma(\tau)$ of a real valued stationary process $\{X(t)\}$ is p.d.f. (n.n.d.).

Proof:- WLG, let us assume that $\mu = 0 = E\{X(t)\}$

Then for a set of real numbers $a_1, a_2, \ldots, a_n$ and the index set $T = \{t_1, t_2, \ldots, t_n\}$, it may be noted that —

$$0 \leq Var\left\{\sum_{j=1}^{n} a_j X(t_j)\right\}$$

$$= E\left\{\sum_{j=1}^{n} a_j X(t_j)\right\}^2 \quad [\because E[X(t_j)] = 0 \ \forall \ j = 1(1)n]$$

$$= E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j X(t_i) X(t_j)\right\}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E\{X(t_i), X(t_j)\}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \, Cov\{X(t_i), X(t_j)\}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \, Cov\{X(t_i), X(t_i + \overline{t_j - t_i})\}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(t_j - t_i) \quad [\because \text{since } X(t) \text{ is stationary}]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(t_i - t_j) \quad [\because \text{since a.c.f. is symmetri}$$

$$\geq 0$$

$\therefore$ The correlation function $\gamma(\cdot)$ of stationary process is positive semi definite.

- **Definition:-** A real valued function $f(t)$ defined as $t \in \tau$ is said to be positive semi-definite iff

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j f(t_i - t_j) \geq 0 \quad , \text{ where}$$

$a_1, a_2, \ldots, a_n$ are set of real numbers and $t_1, t_2, \ldots, t_n$ are such that $(t_i - t_j) \in T \ \forall \ i, j = 1(1)n$.

Cor. $|\rho(\tau)| \leq 1$.

Proof:- For $n = 2$, we have

$$\sum_{i=1}^{2}\sum_{j=1}^{2} a_i a_j \gamma(t_i - t_j) \geq 0$$

$$\Rightarrow a_1^2 \gamma(0) + a_2^2 \gamma(0) + 2a_1 a_2 \gamma(t_1 - t_2) \geq 0 \quad [\because \gamma(\tau) = \gamma(-\tau)]$$

$$\Rightarrow \frac{a_1^2 + a_2^2}{2} \geq -a_1 a_2 \frac{\gamma(t_1 - t_2)}{\gamma(0)}$$

$$= -a_1 a_2 \rho(t_1 - t_2)$$

$$= -a_1 a_2 \rho(\tau) \text{ , letting } t_1 - t_2 = \tau.$$

For $a_1 = a_2 = 1$, we have,

$$\rho(\tau) \geq -1$$

for $a_1 = -1$, $a_2 = 1$, we have,

$$\rho(\tau) \leq 1.$$

$$\therefore -1 \leq \rho(\tau) \leq 1$$

i.e. $|\rho(\tau)| \leq 1$.

SOME USEFUL LINEAR STOCHASTIC TIME SERIES MODELS

A model is a representation of a system under study, and a mathematical model is one in which the system is represented by symbols that can be manipulated by using mathematical rules.

As we know that a stochastic process is a statistical phenomenon that evolves in time according to probabilistic laws and the Time series to be analysed, may be thought of as a particular realisation, produced by the underlying probability mechanism, of the system under study.

Therefore, a model that describes the probability structure of a sequence of observations is called a stochastic process. A TS is regarded as a sample realisation from an infinite population of such samples, which could have been generated by the process (model).

A major objective of statistical investigation is to infer properties of the population from those of samples. For example, to make forecast is to infer the probality distribution of a future observation from the population, given a sample of past values. To do this we need ways of describing stochastic processes and time series and we also need classes of stochastic model that are capable of describing situations occuring in practice.

A very special and also very important class of stochastic processes is the stationary process, it is based on the assumptions that the process is in a particular state of statistical equilibrium.

A less restrictive requirement for a stationary process, called weak stationarity of order 's', is that the moments upto some order 's' depends wholely on the time differences (or, lag τ). For example, the existance of a fixed mean μ and an a.cv. matrix $\underset{\sim}{\Gamma}_n$ of the form, —

$$\underset{\sim}{\Gamma}_n = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_0 \end{pmatrix}$$

with $\gamma_0 = \sigma^2$, constant and $\gamma_K = \text{cov}(X_t, X_{t+K})$ for all $K = 1, 2, \ldots, n-1$.

where, γ_K's are the a.cv. function of a stationary stochastic process for observations $(X_1, X_2, \ldots, X_n)$ made at n successive times. It is sufficient to ensure stationarity upto 2nd order.

Obviously, $\underset{\sim}{\Gamma}_n = \underset{\sim}{\Gamma}_n'$ and $\underset{\sim}{\Gamma}_n$ is p.d. matrix.

Thus, 2nd order stationarity plus an assumption of normality, are sufficient to produce STRICT stationarity, since normal distribution is fully characterised by its moments of 1st and 2nd order.

NOTE:- Whenever we use the term 'STATIONARY', we shall mean weakly stationary unless told otherwise.

Some Useful Stochastic Processes :

1) IID Noise / A purely random Process : — A discrete-time process $\{Z_t\}$ is called an i.i.d. noise or purely random process if it consists of a sequence of uncorrelated random variables, say, $\{Z_t\}$ which are mutually independent and identically distributed with mean zero and constant variance, i.e.

$$E(Z_t) = 0, \quad \text{Var}(Z_t) = \sigma_Z^2 = E(Z_t^2) < \infty \ \forall\ t.$$

i.e. i.i.d. noise or purely random process of RV's are

$$\{Z_t\} \sim \text{IID}(0, \sigma_Z^2).$$

Since the r.v.'s Z_t are uncorrelated, it follows that

$$\gamma_Z = \text{cov}(Z_t, Z_{t+K}) = \begin{cases} \sigma_Z^2 & \text{if } K = 0 \\ 0 & \text{iff } K \neq 0 \end{cases}$$
$$= \gamma(K) \ \forall\ t.$$

The auto covariance function is $\gamma(k) = \begin{cases} \sigma_Z^2 & \text{if } k=0 \\ 0 & \text{if } k \neq 0 \end{cases}$

and the auto correlation function is

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

As the mean and auto covariance function do not depend on time and the process is second order stationary. In fact, the identical assumption implies that the process is also strictly stationary.

A purely random process is sometimes called white noise. Process $\{Z_t\}$ of this type are useful in deriving more complicated purposes (processes).

2) Random shocks / White noise Process :— A discrete time process is called a white noise process if it consists of a sequence of r.v.'s $\{Z_t\}$ which are uncorrelated each with mean zero and variance $\sigma_Z^2 = E(Z_t^2) < \infty$, i.e. $\{Z_t\} \sim WN(0, \sigma_Z^2)$

The acv. function is given by

$$\begin{aligned} \gamma(Z_t, Z_{t+k}) &= \operatorname{cov}(Z_t, Z_{t+k}) \\ &= \begin{cases} \sigma_Z^2 & \text{if } k=0 \text{ for all } t. \\ 0 & \text{ow} \end{cases} \\ &= \gamma(k), \text{ say.} \end{aligned}$$

so that a.c.f. is

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \begin{cases} 1 & \text{if } k=0 \text{ for all } t. \\ 0 & \text{ow} \end{cases}$$

Therefore, the white noise process is stationary as its acv. function depends only on lag k; the mean and variance being constant.

Note:- For each the IID noise process, and white noise process one may further assume that the distribution of Z_t is univariate $N(0, \sigma^2)$.

3) Random Walk: — A process $\{X_t\}$ is said to be a random walk if

$$X_t = X_{t-1} + Z_t ,$$

where $\{Z_t\}$ is (discrete-time) purely random process (i.i.d. noise) with mean μ and variance σ_Z^2.

The process is customarily started at zero when $t=0$, so that $X_1 = Z_1$, $X_2 = Z_1 + Z_2$, ...

and lastly, $X_t = \sum_{i=1}^{t} Z_i \;\forall\, t$

Then we find that $E(X_t) = \sum_{i=1}^{t} (Z_t) = t\mu$ and $Var(X_t) = t\sigma_Z^2$

$\gamma(X_t, X_{t+k}) = Cov(X_t, X_{t+k}) = t\sigma_Z^2 \;\forall\, t.$

As the mean and variance change with 't', the process $\{X_t\}$ is non-stationary.

However, the first differences of a random walk, given by $\nabla X_t = X_t - X_{t-1} = Z_t$ from a purely random process, which is therefore stationary.

Note:- The best-known examples of time series, which behave like random walks, are share prices on successive days.

4) Linear Filter Model: — A time series $\{X_t\}$ in which successive values are highly dependent can frequently be regarded as generated from a series of independent shocks $\{Z_t\}$ which are random drawing from a fixed and identical distribution. Usually, a normal distribution with mean zero and variance σ_Z^2, i.e. $\{Z_t\}$ is a white noise process. Based on this idea let us consider a linear stochastic model for $\{X_t\}$.

$$X_t = \mu + Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \cdots\cdots$$
$$= \mu + \Psi(B) Z_t \text{ , where,} \quad (1)$$

$\Psi(B) = 1 + \psi_1 (B) + \psi_2 B^2 + \cdots\cdots$ and B is $\ni$

$B Z_t = Z_{t-1} \quad \forall\, t$, B is backward shifting operator.

In fact, $\psi_0 = 1, \psi_1, \psi_2, \ldots\ldots$ are called weights and μ is a parameter which determines the 'level' of the process and the operator $\Psi(B)$ is the linear filter (operator) that transforms Z_t with X_t and is called the transfer function of the filter (i.e. $\Psi(B)$ may be thought of as a linear filter which when applied to the white noise input series $\{Z_t\}$ produces the output $\{X_t\}$).

$$\text{input} \xrightarrow[\{Z_t\}]{\text{white noise}} \boxed{\begin{array}{c}\text{Linear Filter}\\ \Psi(B)\end{array}} \xrightarrow{\text{output}} \{X_t\}$$

The sequence of weights, $\psi_0 = 1; \psi_2, \ldots, \psi_n, \ldots$ are finite or infinite. If the sequence is finite or infinite but absolutely summable, i.e. $\sum_{j=0}^{\infty} |\psi_j| < \infty$, [i.e. $\Psi(B)$ is convergent and in which case $|B| < 1$] The filter $\Psi(B)$ is said to be 'stable' and the process $\{X_t\}$ is stationary. The parameter μ is then the mean about which the process varies, otherwise $\{X_t\}$ is non-stationary and μ has no specific meaning except as a reference point for the level of the process.

Taking μ as zero or writing $X_t - \mu$ as X_t the model ①, viz.

$$X_t = \Psi(B) Z_t$$

is called a linear filter model.

Under suitable conditions we can write the above, in the form

$$\begin{aligned} Z_t &= \Psi^{-1}(B) X_t \\ &= \pi(B) X_t \\ &= (1 - \pi_1 B - \pi_2 B^2 - \cdots\cdots) X_t \\ &= X_t - \pi_1 X_{t-1} - \pi_2 X_{t-2} - \cdots\cdots, \end{aligned}$$

which is the inverted form of

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

If there are finite numbers of weights $\pi_0 = 1, \pi_1, \pi_2, \ldots\ldots$ or an infinite number of weights but the series $\pi(B)$ is convergent for $|B| < 1$, then the process is said to be INVERTIBLE.

Auto-covariance generating function of a Linear Process:-

A basic data analysis tool for identifying models is the auto-correlation function. Therefore, it is important to know the a.c.f. of a linear process. The acv.f. of the linear process

$$X_t = \Psi(B) Z_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \text{ with } \psi_0 = 1 \text{ and}$$

$$\{Z_t\} \sim WN(0, \sigma_Z^2), \text{ is given by}$$

$$\gamma_K = \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+K} \quad \left[= \operatorname{cov}(X_t, X_{t+K}) \right]$$

In particular, for $K=0$, we find that the variance is

$$\gamma_0 = Var(X_t) = \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j^2$$

It follows that the stationary condition of absolute summability of the coefficients ψ_j, $\sum_{j=0}^{\infty} |\psi_j| < \infty$, implies that the series on the right of $\gamma_0 = \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j^2$ converges and hence guarantees that the linear process will have a finite variance.

Another way of obtaining the auto-covariance of a linear process is via the auto-covariance generating function

$$\gamma(B) = \sum_{K=-\infty}^{\infty} \gamma_K B^K$$

in which it is noted that γ_0, the variance of the process, is the coefficient of $B^0 = 1$, while γ_K, the acv. function of lag K, is the coefficient of both B^j and $B^{-j} = F^j$ [$FX_t = X_{t+1}$ is the forward shifting operator] so that

$$\boxed{\gamma(B) = \sigma_Z^2 \Psi(B)\Psi(B^{-1}) = \sigma_Z^2 \Psi(B)\Psi(F)}$$

In the development, when treated as a variable in a generating function, B will be supposed capable for taking complex values. In particular, it will often be necessary to consider the different situations occuring when $|B|<1$, $|B|=1$, or, $|B|>1$, i.e. when the complex number B lies inside, on, or, inside the unit circle.

Stationarity condition :— The convergence of the series $\gamma_0 = \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j^2$ ensures that the process $\{X_t\}$ has a finite variance. Also that the auto covariances and autocorrelations must satisfy a set of conditions to ensure stationarity. For a linear process $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$. These conditions are guaranted by the single condition that $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

This condition can also be embodied in the condition that the series $\Psi(B)$, which is the generating function of the ψ weights, must converge for $|B| \le 1$, i.e. on or within the unit circle.

Invertibility condition :— We now consider a restriction applied to the π weights to ensure what is called invertibility. The invertibility condition is independent of the stationary condition (i.e. it is also applicable to the non-stationary linear models).

The linear process $X_t = \Psi(B) Z_t$ is invertible and has the representation,

$$Z_t = \Pi B(X_t)$$
$$= (1 - \pi_1 B - \pi_2 B^2 - \cdots) X_t$$

if the weights π_j are absolutely summable, i.e. if $\sum_{j=0}^{\infty} |\pi_j| < \infty$ which implies that the series $\pi(B)$ converges on or within the unit circle.

To sum up, a linear process $X_t = \Psi(B) Z_t$ is stationary if $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and is invertible if $\sum_{j=0}^{\infty} |\pi_j| < \infty$ where

$$\pi(B) = \psi^{-1}(B)$$
$$= 1 - \sum_{j=1}^{\infty} \pi_j B^j$$

Remark :— Moving average process and auto-regressive process are the special cases of a general linear process.

5) Moving Average Process :— A (linear) process $\{X_t\}$ is said to be an MA process of order q, abbreviated as MA(q), if

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}$$

$$\text{or, } X_t = \sum_{i=0}^{q} \beta_i Z_{t-i},$$

where, $\{Z_t\}$ is a purely random process with mean zero and variance σ_Z^2 and $\{\beta_i\}$ are constants.
The Z's are usually scaled so that $\beta_0 = 1$, so that we take $X_t = \sum_{i=0}^{\infty} \beta_i Z_{t-i}$ with $\beta_0 = 1$.

Mean :—
$$E(X_t) = E\left(\sum_{i=0}^{q} \beta_i Z_{t-i}\right)$$
$$= \sum_{i=0}^{q} \beta_i E(Z_{t-i})$$
$$= 0 \quad \forall t.$$

Variance :—
$$Var(X_t) = Var\left(\sum_{i=0}^{q} \beta_i Z_{t-i}\right)$$
$$= \sum_{i=0}^{q} \beta_i^2 Var(Z_{t-i})$$
$$= \sigma_Z^2 \sum_{i=0}^{q} \beta_i^2$$
, since Z_t's are i.i.d and variance $= \sigma_Z^2$.

Auto-covariance function :—

$$\gamma(k) = Cov(X_t, X_{t+k})$$
$$= Cov\left[\sum_{i=0}^{q} \beta_i Z_{t-i}, \sum_{i=0}^{q-k} \beta_i Z_{t+k-i}\right]$$
$$= \begin{cases} \gamma(-k) & \text{if } k < 0 \\ \sigma_Z^2 \sum_{i=0}^{q-k} \beta_i \beta_{i+k} & \text{if } k = 0, 1, 2, \ldots, q \\ 0 & \text{if } k > q \end{cases}$$

since $Cov(Z_s, Z_t) = \begin{cases} 0 & \text{for } s \neq t \\ \sigma_Z^2 & \text{for } s = t \end{cases}$

Auto correlation function :- The a.c.f. of the above MA(q) process, i.e. of the process

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}, \text{ is given by,}$$

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \begin{cases} \rho(-k) & \text{for } k<0 \\ 1 & \text{for } k=0 \\ \left\{\sum_{i=0}^{q-k} \beta_i\beta_{i+k}\right\} \Big/ \left\{\sum_{i=0}^{q} \beta_i^2\right\} & \text{for } k=1,2,3,\ldots,q \\ 0 & \text{for } k>q . \end{cases}$$

Note that the a.c.f. of MA(q) process "cuts off" at lag q, which is a special feature of MA process. Further an arbitary constant, μ, say, may be added to the RHS of

$$X_t = \beta_0 Z_t + \cdots + \beta_q Z_{t-q},$$ to give a process with mean μ. This does not affect the a.c.f. and has been omitted for simplicity.

Stationarity :- As $\gamma(k)$ does not depend on t and the mean is constant, the process is second order stationary for all values of $\{\beta_i\}$. Furthermore, if the Z_i's are normally distributed, then so are the X_t's and we have that the MA(q) process is strictly stationary.

Stated otherwise, let us write the MA(q) process as

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q} \quad \text{with } \beta_0 = 1$$

$$= (1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_q B^q) Z_t, \text{ where } B^j X_t = X_{t-j} \; \forall j.$$

$$= \theta(B) Z_t, \text{ say}$$

Since the series $\psi(B) = \theta(B) = 1 + \beta_1 B + \cdots + \beta_q B^q$ is finite, no restriction are needed on the parameters $\{\beta_i\}$ & σ_Z^2 of the MA(q) process to ensure stationarity, i.e. an MA(q) process is always stationary.

Invertibility :- The invertibility condition for an MA(q) process $\{X_t\}$ of the form

$$X_t = Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q},$$
$$= \theta(B) Z_t, \text{ may be obtained as}$$

$$Z_t = \theta^{-1}(B) X_t$$

Hence, if $\theta(B) = \prod_{j=1}^{q} (1 - H_j B)$, where $\frac{1}{H_1}, \frac{1}{H_2}, \ldots, \frac{1}{H_q}$ are the roots of the characteristic equation $\theta(B) = 0$.

Then, on expanding in partial fractions, we obtain

$$\pi(B) = \Psi^{-1}(B) = \theta^{-1}(B) = \sum_{i=1}^{q} \left(\frac{M_i}{1 - H_i B}\right)$$

which converges, or, equivalently, the weights

$\pi_j = -\sum_{i=1}^{q} M_i H_i^j$ are absolutely summable if $|H_i| < 1$ $\forall\ i = 1(1)q$, since the roots of $\theta(B) = 0$ are H_i^{-1} $(i = 1(1)q)$, it follows that the invertibility condition for an MA(q) process is that the roots of the characteristic equation $\theta(B) = 0$ lie outside the unit circle.

⟨i⟩ MA(1) Process [First order MA process] :-

The first order MA process is defined as

$$X_t = Z_t + \beta_1 Z_{t-1}, \text{ where } \beta_0 = 1.$$
$$= (1 + \beta_1 B) Z_t$$
$$= \theta(B) Z_t, \text{ say},$$

where, $\theta(B) = (1 + \beta_1 B)$

i.e. $Z_t = \theta^{-1}(B) X_t = \pi(B) X_t$, where $\pi(B) = (1 + \beta_1 B)^{-1}$

i.e. $\pi(B) = 1 - \beta_1 B + \beta_1^2 B^2 - \cdots\cdots$

Now for the process to be invertible it is required that the series

$\pi(B) = 1 - \beta_1 B + \beta_1^2 B^2 - \cdots\cdots$ is convergent.

Obviously, $\pi(B) = (1 + \beta_1 B)^{-1}$ is convergent for all B with $|B| \le 1$ if $|\beta_1| < 1$. Also, if the root of the characteristic equation $\theta(B) = 0$, i.e. $(1 + \beta_1 B) = 0$ lie outside the unit circle.

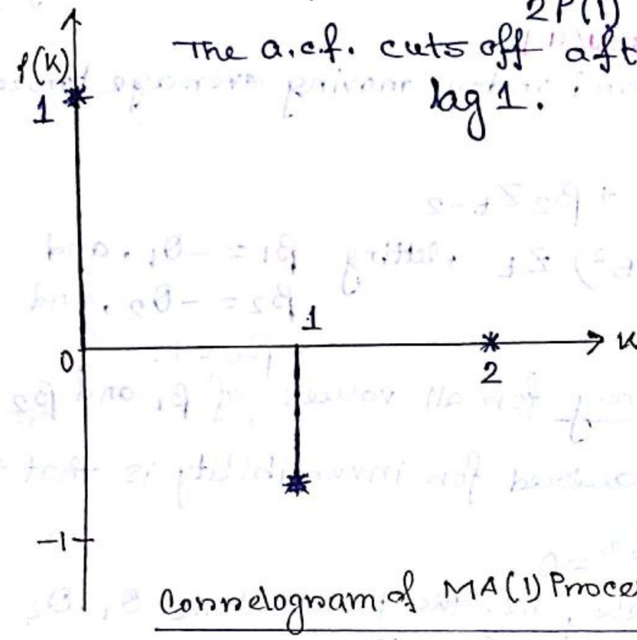
Clearly, $\rho(k) = \begin{cases} 1 & , k=0 \\ \frac{\beta_1}{1+\beta_1^2} & , k=\pm 1 \\ 0 & , k \geq 2 \end{cases}$ and the process is of course <u>stationary</u> for all values of β_1.

Note that, $\rho(1) = \frac{\beta_1}{1+\beta_1^2} \Rightarrow \beta_1^2 - \frac{\beta_1}{\rho(1)} + 1 = 0 \Rightarrow \rho(1)\beta_1^2 - \beta_1 + \rho(1) = 0$

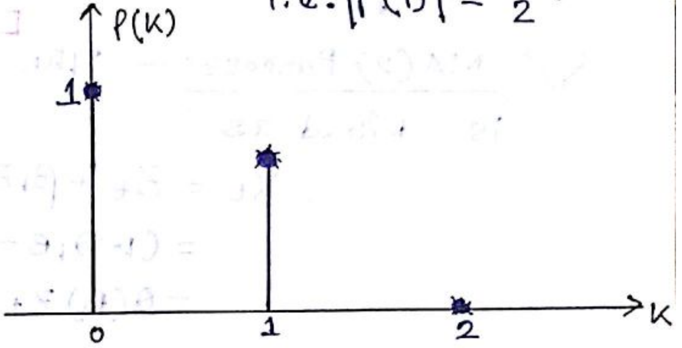
$\therefore \beta_1 = \frac{+1 \pm \sqrt{1-4\rho^2(1)}}{2\rho(1)}$. since β_1 is real, $1-4\rho^2(1) \geq 0$

i.e. $|\rho(1)| \leq \frac{1}{2}$.

The a.c.f. cuts off after lag 1.



Correlogram of MA(1) Process when $\beta_1 < 0$.



Correlogram of MA(1) Process when $\beta_1 > 0$

<u>Remark</u>:- Consider the following MA(1) process.

i) $X_t = Z_t + \beta_1 Z_{t-1}$; ii) $X_t = Z_t + \frac{1}{\beta_1} Z_{t-1}$.

The a.c.f. of model (ii) is $\rho_2(K) = \begin{cases} 1 & , k=0 \\ \frac{\beta_1}{1+\beta_1^2} & , k=\pm 1 \\ 0 & \text{ow} \end{cases}$, same as the a.c.f of model (i).

[The fact is also clear from the equation $\rho(1) = \frac{\beta_1}{1+\beta_1^2}$
$\Leftrightarrow 1+\beta_1^2 - \frac{\beta_1}{\rho(1)} + 1 = 0$. The constant term is 1, the product of the roots of the equation is unity, i.e. if β_1 is a solution, so is $\frac{1}{\beta_1}$.]

Thus we can't identify an MA process uniquely from a given a.c.f.

Now, we can write

(i) $X_t = (1+\beta_1 B) Z_t$, where $BZ_t = Z_{t-1}$.

$\Rightarrow Z_t = (1+\beta_1 B)^{-1} X_t = (1 - \beta_1 B + \beta_1^2 B^2 - \cdots\cdots) X_t$

$= X_t - \beta_1 X_{t-1} + \beta_1^2 X_{t-2} - \cdots\cdots$

(ii) $X_t = (1 + \frac{B}{\beta_1}) Z_t$

$\Rightarrow Z_t = X_t - \frac{1}{\beta_1} X_{t-1} + \frac{1}{\beta_1^2} X_{t-2} - \cdots\cdots$

If $|\beta_1| < 1$, the series for (i) converges, where as that for (ii) does not. Thus if $|\beta_1| < 1$, the model (i) is said to be invertible, where as model (ii) is not. If β_1 satisfies the invertible condition $|\beta_1| < 1$, then $\left|\frac{1}{\beta_1}\right| > 1$ and $\frac{1}{\beta_1}$ will not satisfy the condition. The imposition of the invertibility condition ensures that there is a <u>unique</u> MA(1) process for a given a.c.f.

⟨ii⟩ <u>MA(2) Process:</u>— The second order moving average process is defined as

$$X_t = Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2}$$
$$= (1-\theta_1 B - \theta_2 B^2) Z_t \text{, letting } \beta_1 = -\theta_1 \text{, and } \beta_2 = -\theta_2 \text{, and } \beta_0 = 1.$$
$$= \theta(B) Z_t .$$

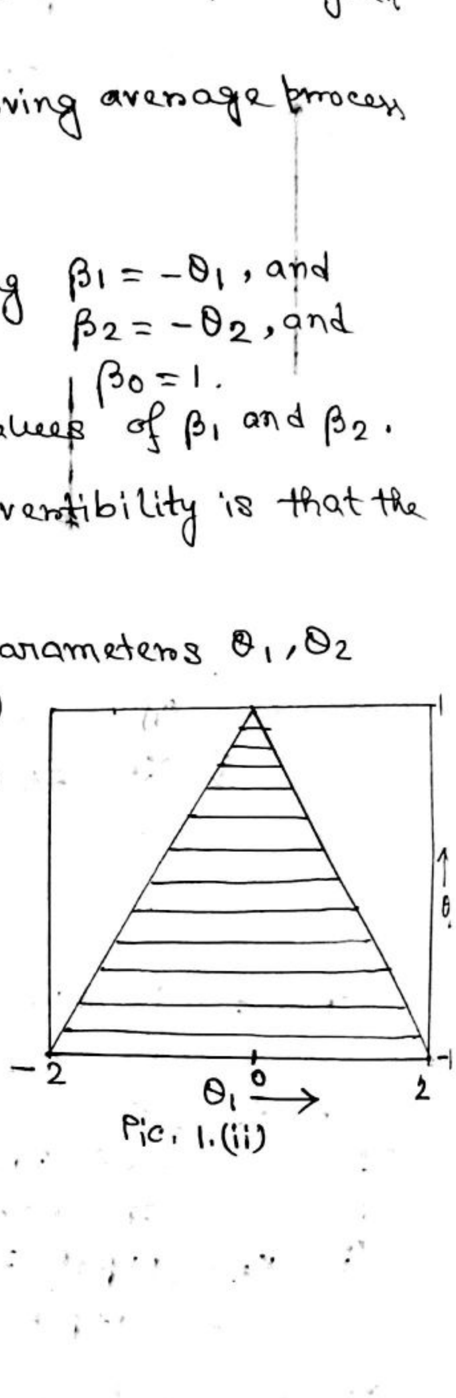
MA(2) process is <u>stationary</u> for all values of β_1 and β_2.

For MA(2) process, it is required for invertibility is that the roots of the equation

$$1-\theta_1 B - \theta_2 B^2 = 0$$

lie outside the unit circle, i.e. the parameters θ_1, θ_2 should lie outside the region (Pic. 1.ii)

$$\theta_1 + \theta_2 < 1 \ , \ \theta_2 - \theta_1 < 1 \ , \ |\theta_2| < 1 .$$



Pic. 1.(ii)

The a.c.f. is

$$\rho(k) = \begin{cases} 1 & \text{if } k=0 \\ \left(\dfrac{\sum_{i=0}^{2-k} \beta_i \beta_{i+k}}{\sum_{i=0}^{2} \beta_i^2} \right) & \text{if } k=1,2 \\ 0 & \text{if } k \geq 3 \end{cases}$$

i.e. $\rho(1) = \dfrac{\beta_1 + \beta_1\beta_2}{1+\beta_1^2+\beta_2^2}$ and $\rho(2) = \dfrac{\beta_2}{1+\beta_1^2+\beta_2^2}$; $\rho(k) = 0 \ \forall \ k \geq 3$

Thus the a.c.f. has a "cut off" at lag 2.

Example:- For the a.c.f. of the MA(2) process given by

$$X_t = \epsilon_t + 0.7\epsilon_{t-1} - 0.2\epsilon_{t-2}$$, where $\{\epsilon_t\}$ is a purely random process. Also draw the correlogram.

Solution:- As $\{\epsilon_t\}$ is a purely random process, then

$$E(\epsilon_t) = 0,$$

$$Cov(\epsilon_t, \epsilon_{t+k}) = \begin{cases} \sigma_\epsilon^2 & \text{if } k=0 \\ 0 & \text{if } k \neq 0 \end{cases}$$

Then $E(X_t) = 0$ and $Var(X_t) = V(\epsilon_t) + (0.7)^2 V(\epsilon_{t-1}) + (-0.2)^2 V(\epsilon_{t-2})$

$$= (1 + 0.49 + 0.04)\sigma_\epsilon^2 = 1.53\,\sigma_\epsilon^2.$$

Now, $Cov(X_t, X_{t+1}) = Cov\left[\epsilon_t + 0.7\epsilon_{t-1} - 0.2\epsilon_{t-2},\ \epsilon_{t+1} + 0.7\epsilon_t - 0.2\epsilon_{t-1}\right]$

$$= 0.7\,Cov(\epsilon_t, \epsilon_t) + (0.7)(-0.2)\,Cov(\epsilon_{t-1}, \epsilon_{t-1})$$

$$= (0.7 - 0.14)\sigma_\epsilon^2$$

$$= (0.56)\sigma_\epsilon^2.$$

and $Cov(X_t, X_{t+2}) = Cov\left[\epsilon_t + 0.7\epsilon_{t-1} - 0.2\epsilon_{t-2},\ \epsilon_{t+2} + 0.7\epsilon_{t+1} - 0.2\epsilon_t\right]$

$$= (-0.2)\,Cov(\epsilon_t, \epsilon_t)$$

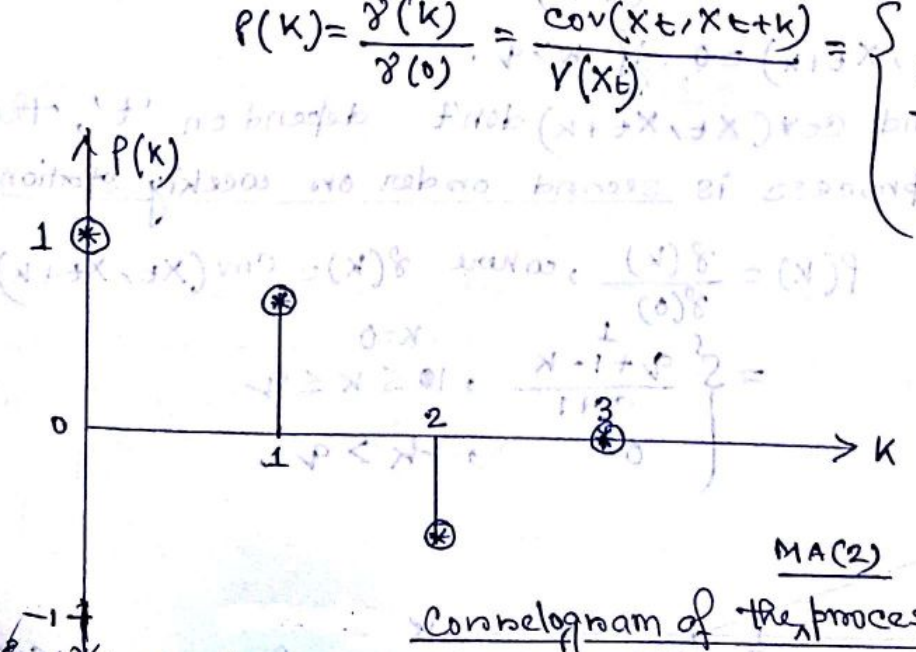
$$= (-0.2)\sigma_\epsilon^2$$

Also, $Cov(X_t, X_{t+k}) = 0 \quad \forall\, k = 3, 4, 5, \ldots$

Therefore, $Cov(X_t, X_{t+k}) = \begin{cases} \sigma_\epsilon^2, & k=0 \\ (0.56)\sigma_\epsilon^2, & k=1 \\ (-0.2)\sigma_\epsilon^2, & k=2 \\ 0, & k \geq 3 \end{cases}$?

and $E(X_t)$ does not depend on 't', i.e. the process is second-order stationary. The a.c.f. is

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{Cov(X_t, X_{t+k})}{V(X_t)} = \begin{cases} 1, & k=0 \\ 0.56, & k=1 \\ -0.2, & k=2 \\ 0, & k \geq 3 \end{cases} \quad 0.36$$



Correlogram of the process

⟨iii⟩ Simple MA Process:- In an MA(q) process $\{X_t\}$ where $X_t = \sum_{i=0}^{q} \beta_i Z_{t-i}$, if the weights $\{\beta_i\}$ are equal and $\sum_{i=0}^{q} \beta_i = 1$; i.e. if $\beta_i = \frac{1}{q+1}$, $i = 0(1)q$, then the process reduces to

$$X_t = \frac{1}{q+1} \sum_{i=0}^{q} Z_{t-i},$$ is known as simple MA(q) process.

DEFINITION:- Suppose $\{Z_t\}$ is a purely random process, i.e. $E(Z_t) = 0 \ \forall t$ and $Cov[Z_t, Z_{t+k}] = \begin{cases} \sigma_Z^2, & k=0 \\ 0, & k \neq 0 \end{cases}$

Then, a process $\{X_t\}$ is given by,

$$X_t = \sum_{i=0}^{q} \frac{Z_{t-i}}{q+1}$$ is known as a simple MA process of order q.

Clearly, $E(X_t) = 0$ and $Var(X_t) = V\left(\frac{1}{q+1} \sum_{i=0}^{q} Z_{t-i}\right)$

$$= \frac{1}{(q+1)^2}(q+1)\sigma_Z^2 \ \forall t$$

Also,
$$Cov(X_t, X_{t+k}) = Cov\left[\frac{1}{q+1}\sum_{i=0}^{q} Z_{t-i}, \frac{1}{q+1}\sum_{i=0}^{q} Z_{t+k-i}\right]$$

$$= Cov\left[\frac{1}{q+1}\sum_{i=0}^{q} Z_{t-i}, \frac{1}{q+1}\sum_{j=-k}^{q-k} Z_{t-j}\right], \text{ where } j = i-k.$$

$$= \frac{1}{(q+1)^2}\sum_{i=0}^{q-k} Cov(Z_{t-i}, Z_{t-i}), \text{ since } Z_i\text{'s are uncorrelated.}$$

$$= \frac{1}{(q+1)^2}\sum_{i=0}^{q-k} \sigma_Z^2$$

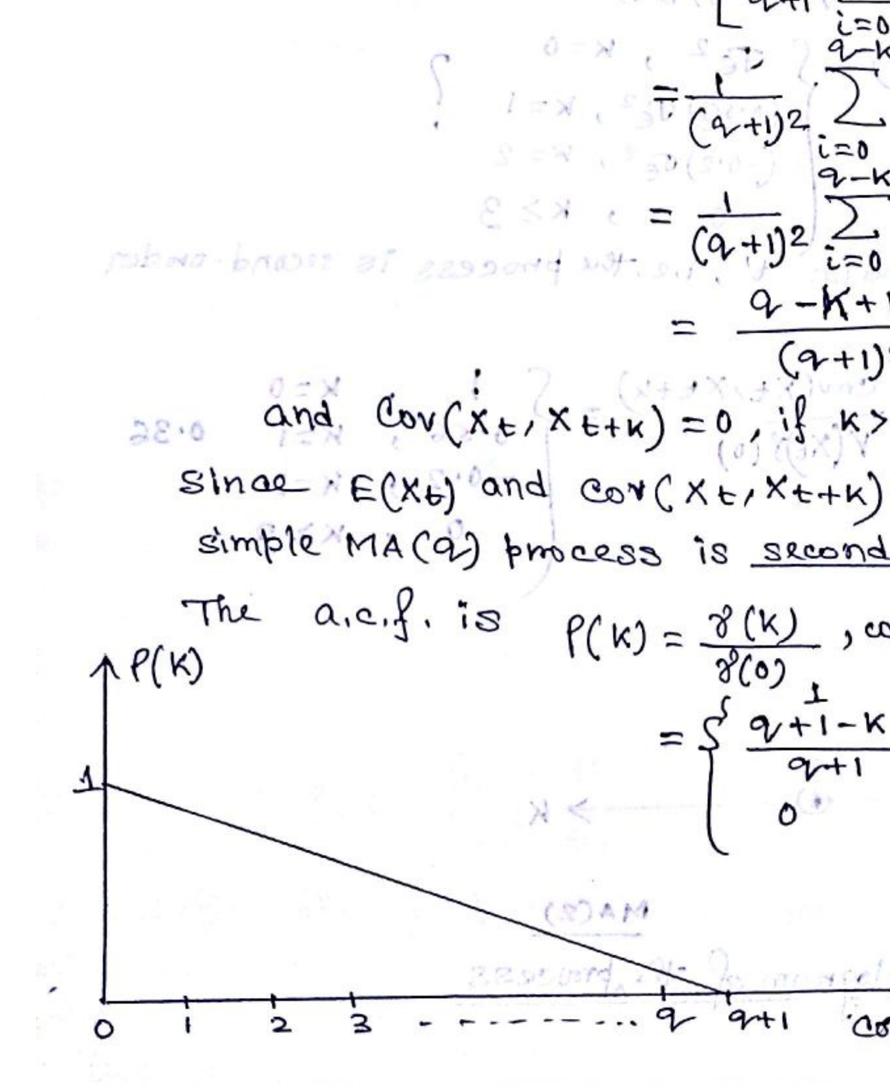
$$= \frac{q-k+1}{(q+1)^2} \text{ if } 0 \leq k \leq q.$$

and $Cov(X_t, X_{t+k}) = 0$, if $k > q$.

Since $E(X_t)$ and $Cov(X_t, X_{t+k})$ don't depend on 't', the simple MA(q) process is second order or weekly stationary.

The a.c.f. is $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$, where $\gamma(k) = Cov(X_t, X_{t+k})$

$$= \begin{cases} 1, & k=0 \\ \frac{q+1-k}{q+1}, & 1 \leq k \leq q \\ 0, & k > q. \end{cases}$$



[W'U20W] Correlogram of simple MA(q) process.

6) <u>Auto regressive (AR) process:</u>— Suppose that $\{Z_t\}$ is a purely random process with mean zero and variance σ_Z^2. Now, if we derive a process $\{X_t\}$ where the current value of the process X_t is expressed as a finite, linear aggregate of previous p values of the process and a random component Z_t. Then a process $\{X_t\}$ is said to be an <u>autoregressive process of order p</u> [abbreviated to an AR(p) process] if

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + Z_t .$$

The reason for this name is that a linear model $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + Z_t$, relating a 'dependent' variable X_t to a set of 'independent' variables $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$, plus an error term Z_t, is often referred to as a <u>multiple regression model</u>, but X_t is regressed on past values of X_t rather than on seperate predictor variables. Hence the model is called <u>'Auto-regressive'</u>

<u>Ques.</u> Show that an AR(p) process can be written as an MA process of infinite extend. Hence, comment on the stationarity of AR(p) process.

<u>ANS</u>:- Define, $BX_t = X_{t-1}$ and $B^j X_t = X_{t-j}$,

Then $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + Z_t$

$$= (\alpha_1 B + \alpha_2 B^2 + \cdots + \alpha_p B^p) X_t + Z_t$$

$$\Leftrightarrow (1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p) X_t = Z_t$$

$$\Leftrightarrow \alpha(B) X_t = Z_t$$

$$\Leftrightarrow X_t = \alpha^{-1}(B) Z_t = \theta(B) Z_t, \text{ say, where } \theta(B) = \alpha^{-1}(B) = \sum_{j=0}^{\infty} \theta_j B^j \text{ with } \theta_0 = 1$$

$$\Leftrightarrow X_t = \sum_{j=0}^{\infty} \theta_j B^j Z_t = \sum_{j=0}^{\infty} \theta_j Z_{t-j}$$

$$\Leftrightarrow X_t = \theta_0 Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots$$

$$= Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots,$$ provided the series $\sum_{j=0}^{\infty} \theta_j B^j$ converges. If $\sum_{j=0}^{\infty} |\theta_j| < \infty$, then $\sum_{j=0}^{\infty} \theta_j B^j$ converges for all $|B| \le 1$, then <u>AR process of a finite order can be expressed as an MA process of infinite order.</u>

If $\sum_{j=0}^{\infty} |\theta_j| < \infty$, then AR(p) can be written as MA(∞); that is,

$$X_t = \sum_{j=0}^{\infty} \theta_j Z_{t-j} .$$

As $E(Z_t)=0$ and $Cov(Z_t, Z_{t+k}) = \begin{cases}\sigma_Z^2, & k=0\\ 0, & k\neq 0,\end{cases}$

$E(X_t)=0$, $Var(X_t)=\sum_{j=0}^{\infty}\theta_j^2 < \infty$ and $Cov(X_t, X_{t+k})$

$$= Cov\left[\sum_{j=0}^{\infty}\theta_j Z_{t-j}, \sum_{j=0}^{\infty}\theta_j Z_{t+k-j}\right]$$

$$= Cov\left[\sum_{j=0}^{\infty}\theta_j Z_{t-j}, \sum_{m=-k}^{\infty}\theta_{m+k} Z_{t-m}\right]$$

where $m = -(k+j)$

$$= \sigma_Z^2 \sum_{j=0}^{\infty}\theta_j\theta_{j+k} < \infty$$

[Note that, $V(X_t)=\sum_{j=0}^{\infty}\theta_j^2<\infty$, is independent of t; implies $|Cov(X_t, X_{t+k})| \le \sqrt{V(X_t)V(X_{t+k})} = V(X_t)$, which is finite.]

Clearly, $E(X_t)=0$ and $Cov(X_t, X_{t+k}) = \sigma_Z^2\sum_{j=0}^{\infty}\theta_j\theta_{j+k}$, are independent of t. Therefore, AR(p) processes are stationary if $\sum_{j=0}^{\infty}|\theta_j|<\infty$.

Auto covariance function and a.c.f. of AR(p) Process:-

We have $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + Z_t$

Hence, $X_{t-k}X_t = \alpha_1 X_{t-1}X_{t-k} + \alpha_2 X_{t-2}X_{t-k} + \cdots + \alpha_p X_{t-k}X_{t-p} + Z_t X_{t-k}$, $k>0$

Taking expectations both sides

$\gamma_k = \alpha_1\gamma_{k-1} + \alpha_2\gamma_{k-2} + \cdots + \alpha_p\gamma_{k-p}$

since X_{t-k} depends on $Z_{t-k}, Z_{t-k+1}, \ldots$ but not on Z_{t-k+i}, $i>0$

Thus we have the recurrence relation

$$\gamma_k = \sum_{j=1}^{p}\alpha_j\gamma_{k-j}, \quad k=1,2,\ldots,$$

Dividing by $\gamma_0 = \sigma_X^2$,

$$\rho_k = \alpha_1\rho_{k-1} + \alpha_2\rho_{k-2} + \cdots + \alpha_p\rho_{k-p},$$

$$\text{i.e. } (1-\alpha_1 B - \cdots - \alpha_p B^p)\rho_k = 0 \quad \text{——— (1)}$$

writing $\rho_{k-j} = B^j\rho_k$. (1) is a difference equation in ρ_k of order p.

If $\Phi(B) = \prod_{j=1}^{p}(1-G_j B)$, then we have

$\rho_k = A_1 G_1^k + \cdots + A_p G_p^k$, for stationarity $|G_j|<1\ \forall j=1(1)p$

INVERTIBILITY & STATIONARITY :—

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots\cdots + \alpha_p X_{t-p} + Z_t .$$

This is called a autoregressive model of order p, AR(p), then,

$$Z_t = X_t - \alpha_1 X_{t-1} \cdots\cdots - \alpha_p X_{t-p}$$
$$= (1 - \alpha_1 B - \alpha_2 B^2 - \cdots\cdots - \alpha_p B^p) X_t$$
$$= \Phi(B) X_t \text{ , where } \Phi(B) \text{ is called AR operator.}$$
$$\Rightarrow X_t = \Phi^{-1}(B) Z_t \text{ as AR-model is a particular case of linear filter model.}$$
$$= \Psi(B) Z_t \text{ , taking } \Phi^{-1}(B) = \Psi(B)$$

Since there is a finite numbers of α values, $\alpha_1, \alpha_2, \cdots \alpha_p$, an AR(p) model is always INVERTIBLE.

AR model is stationary if weights $\alpha_1, \alpha_2, \cdots, \alpha_p$ are such that the weights $\psi_1, \psi_2, \cdots$, in $\Psi(B)$ form a convergent series. i.e. the function $\Psi(B) = \Phi^{-1}(B)$ must be convergent $\forall$ B with $|B| \le 1$ for stationarity.

Problem:— Draw the correlogram of a series $\{X_t\}$ with

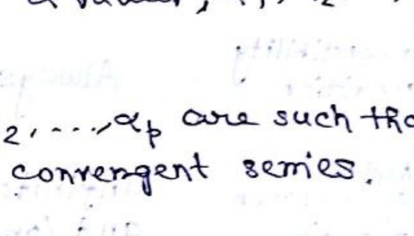
$$X_t = \frac{1}{3}\epsilon_t + \frac{1}{3}\epsilon_{t-1} + \frac{1}{3}\epsilon_{t-2} \text{ , where } E(\epsilon_t) = 0,$$
$$\operatorname{cov}(\epsilon_t, \epsilon_s) = \begin{cases} \sigma^2 & \text{if } t = s \\ 0 & \text{if } t \neq s . \end{cases}$$

Solution:—

$$X_t = \frac{1}{3}(\epsilon_t + \epsilon_{t-1} + \epsilon_{t-2})$$

simple MA(2) process.

$$\rho(k) = \frac{q+1-k}{q+1} \qquad k = 1, 2.$$
$$= 1 \qquad k = 0$$
$$= 0 \qquad k > 2$$



$$\rho(k) = \begin{cases} \frac{2}{3} & k = 1 \\ \frac{1}{3} & k = 2 \end{cases}$$

Distinguish, or, Comparison between AR and MA Process:—

Items	AR Process	MA Process
Model in terms of previous X_t's	$\Phi(B)X_t = Z_t$	$\theta^{-1}(B)X_t = Z_t$
Model in terms of previous Z_t's	$X_t = \Phi^{-1}(B)Z_t$	$X_t = \theta(B)Z_t$
Stationary condition	If roots of $\phi(B)=0$ lie outside the unit circle.	Always stationary
Invertibility condition	Always invertible.	If roots of $\theta(B)=0$ lie outside the unit circle.
Auto-correlation function	Infinite (damped exponential and/or damped sine waves).	Finite.

(Alternative Approach)

Duality Between Autoregressive and Moving Average Process:—

The various characteristics, e.g. mean, variance, auto correlation function, etc. of an AR process and an MA process illustrate a duality between two processes which may be represented or stated as follows:—

1. In a stationary AR(p) process, Z_t can be represented as a finite weighted sum of previous X's or X_t as an infinite weighted sum

$$X_t = \Phi^{-1}(B)Z_t$$

of previous Z's.

Also in an invertible MA(q) process, X_t can be represented as a finite weighted sum of previous Z's or Z_t's as an infinite sum

$$Z_t = \theta^{-1}(B)X_t\,.$$

of previous X's.

2. The finite MA process has an a.c.f. that is zero beyond a certain point, but since it is equivalent to an infinite AR process, its partial auto correlation function is infinite in extent and is dominated by damped exponentials and (or) damped sine waves. Conversely, an AR process has a partial auto correlation function that is zero beyond a certain point, but it's a.c.f. is infinite in extent and consists of a mixture of damped exponentials and (or) damped sine waves.

3. For an AR process of finite order p, the parameters are not required to satisfy any condition to ensure invertibility. However, for stationarity, the roots of $\Phi(B)=0$ must lie outside the unit circle. Conversely, the parameters of the MA process are not required to satisfy any condition to stationarity. However, for invertibility of the MA process, the roots of $\theta(B)=0$ must lie outside the unit circle.

4. The correlogram of an MA(q) process is easy to recognize as it "cuts off" at lag q; whereas the correlogram of AR(p) process is a mixture of damped exponential and sinusoids and die out slowly.

(a) First-order AR Process [Markov Process] :

The first order autoregressive process $\{X_t\}$ is defined as $X_t = \alpha X_{t-1} + Z_t$, where $\{Z_t\}$ is a purely random process.

Now, we may write $(1-\alpha B) X_t = Z_t$

so that
$$X_t = (1-\alpha B)^{-1} Z_t$$
$$= (1+\alpha B + \alpha^2 B^2 + \cdots\cdots) Z_t$$
$$= Z_t + \alpha Z_{t-1} + \alpha^2 Z_{t-2} + \cdots\cdots,$$
provided that the series $\sum_{j=0}^{\infty} \alpha^j B^j$ converges $\forall\, B$ with $|B|<1$

This requires $|\alpha|<1$ to ensure stationarity.

Now, $E(X_t) = 0$ and $Var(X_t) = (1+\alpha^2+\alpha^4+\cdots\cdots)\sigma_Z^2$

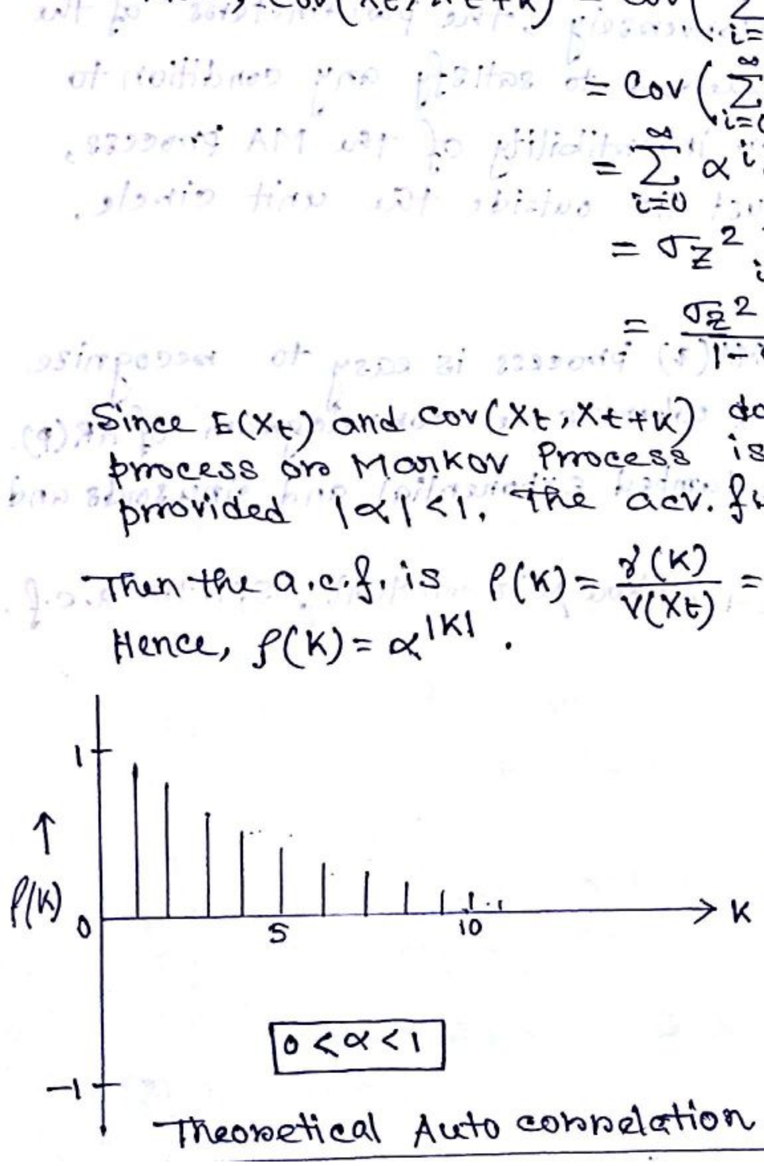
If $|\alpha|<1$, then the variance is finite and $Var(X_t) = \frac{\sigma_Z^2}{1-\alpha^2}\ \forall t$.

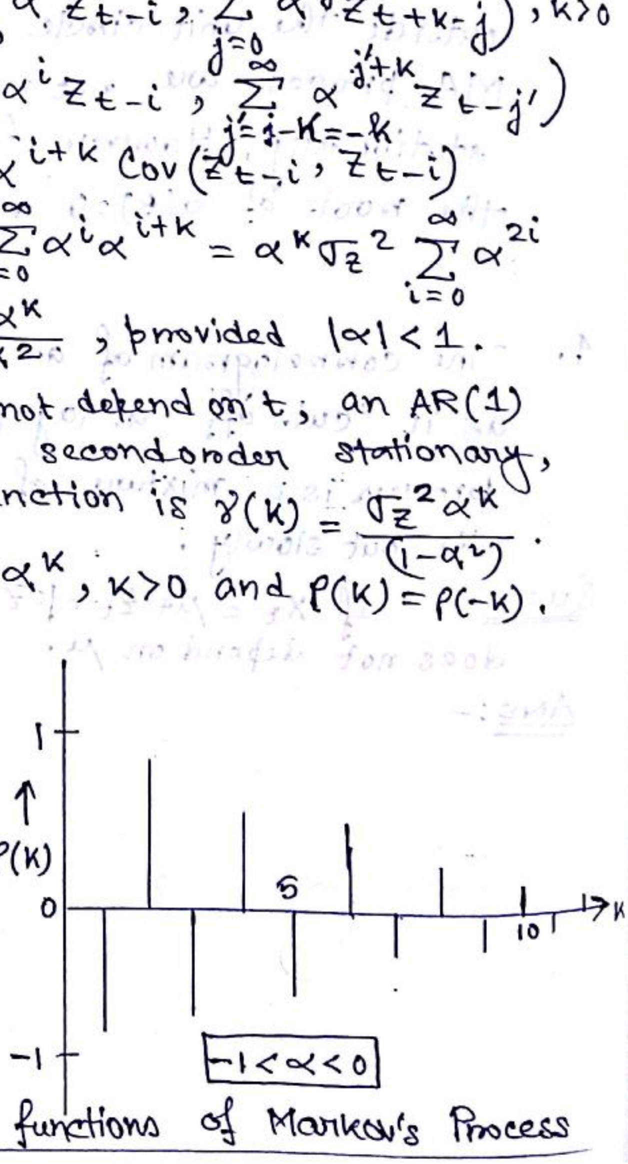
Now,
$$Cov(X_t, X_{t+k}) = Cov\left(\sum_{i=0}^{\infty} \alpha^i Z_{t-i}, \sum_{j=0}^{\infty} \alpha^j Z_{t+k-j}\right), k>0$$
$$= Cov\left(\sum_{i=0}^{\infty} \alpha^i Z_{t-i}, \sum_{j'=j-k=-k}^{\infty} \alpha^{j'+k} Z_{t-j'}\right)$$
$$= \sum_{i=0}^{\infty} \alpha^i \alpha^{i+k} Cov(Z_{t-i}, Z_{t-i})$$
$$= \sigma_Z^2 \sum_{i=0}^{\infty} \alpha^i \alpha^{i+k} = \alpha^k \sigma_Z^2 \sum_{i=0}^{\infty} \alpha^{2i}$$
$$= \frac{\sigma_Z^2 \alpha^k}{1-\alpha^2}, \text{ provided } |\alpha|<1.$$

Since $E(X_t)$ and $cov(X_t, X_{t+k})$ do not depend on t, an AR(1) process or Markov Process is second order stationary, provided $|\alpha|<1$. The acv. function is $\gamma(k) = \frac{\sigma_Z^2 \alpha^k}{(1-\alpha^2)}$.

Then the a.c.f. is $\rho(k) = \frac{\gamma(k)}{V(X_t)} = \alpha^k$, $k>0$ and $\rho(k) = \rho(-k)$.

Hence, $\rho(k) = \alpha^{|k|}$.





Theoretical Autocorrelation functions of Markov's Process

As shown in the figure, the a.c.f. decays exponentially to zero when $0<\alpha<1$, but decays exponentially to zero and oscillates in sign when $-1<\alpha<0$.

(b) Second Order Autoregressive Process [Yule Process] :–

The second order auto-regressive process is defined as
$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t$, where $\{Z_t\}$ is a purely random process,

Stationary Condition :–

We may write $X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} = Z_t$

$$\Rightarrow (1-\alpha_1 B - \alpha_2 B^2) X_t = Z_t .$$

$$\Rightarrow (1-\lambda_1 B)(1-\lambda_2 B) X_t = Z_t ,$$ where $\lambda_1^{-1}, \lambda_2^{-1}$ are the roots of $1-\alpha_1 B - \alpha_2 B^2 = 0$

$$\Rightarrow (1-\lambda_1 B)^{-1}(1-\lambda_2 B)^{-1} Z_t = X_t$$

$$\Rightarrow X_t = \left(\sum_{j=0}^{\infty} \lambda_1^j B^j\right)\left(\sum_{k=0}^{\infty} \lambda_2^k B^k\right) Z_t ,$$ provided both the series $\sum \lambda_1^j B^j$ and $\sum \lambda_2^k B^k$ converge ; i.e. if $|\lambda_1|<1$ and $|\lambda_2|<1$.

As, $(1-\alpha_1 B - \alpha_2 B^2) = (1-\lambda_1 B)(1-\lambda_2 B)$,

$$\lambda_1 + \lambda_2 = \alpha_1 , \quad \lambda_1 \lambda_2 = -\alpha_2$$

Now, $$\lambda_1, \lambda_2 = \frac{\alpha_1 \pm \sqrt{\alpha_1^2 + 4\alpha_2}}{2}$$

Therefore, $$|\lambda_i|<1 \Rightarrow -2 < \alpha_1 \pm \sqrt{\alpha_1^2+4\alpha_2} < 2$$

$$\Rightarrow -2-\alpha_1 < \pm\sqrt{\alpha_1^2+4\alpha_2} < 2-\alpha_1$$

$$\Rightarrow \alpha_1^2 + 4\alpha_2 < (2+\alpha_1)^2$$

and

$$\alpha_1^2 + 4\alpha_2 < (2-\alpha_1)^2$$

$$\Rightarrow \alpha_2 - \alpha_1 < 1 \text{ and } \alpha_1 + \alpha_2 < 1$$

Again, $|-\alpha_2| = |\lambda_1 \lambda_2| < 1$.,

Therefore, $\alpha_1 + \alpha_2 < 1$, $\alpha_2 - \alpha_1 < 1$, i.e. $|\alpha_2|<1$ are the stationarity conditions for AR(2) process.

A.c.f. of a stationary AR(2) Process :

$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t$ is multiplied by X_{t-k} and we get,

$$X_{t-k}X_t = \alpha_1 X_{t-k}X_{t-1} + \alpha_2 X_{t-k}X_{t-2} + X_{t-k}Z_t, \; k>0$$

$$\Rightarrow E(X_{t-k}X_t) = \alpha_1 E(X_{t-k}X_{t-1}) + \alpha_2 E(X_{t-k}X_{t-2}) + \underline{E(X_{t-k}Z_t)} \quad (*)$$

$\Rightarrow \gamma(k) = \alpha_1\gamma(k-1) + \alpha_2\gamma(k-2)$, since $E(X_{t-k}Z_t) = 0$, as X_{t-k} can only involve the error Z_{t-k} which is uncorrelated with Z_t. On dividing throughout in (*) by $\gamma(0)$, we get

$$\rho(k) = \alpha_1\rho(k-1) + \alpha_2\rho(k-2), \; k>0 \quad \text{———} (**)$$

The general solution of the above difference equation is given by $\rho(k) = A_1 u_1^k + A_2 u_2^k$, where A_1, A_2 being found from the initial conditions and u_1, u_2 are the roots of the equation $u^2 = \alpha_1 u + \alpha_2$, i.e. $u^2 - \alpha_1 u - \alpha_2 = 0$, the characteristic equation of the process.

Case I :- if $\underline{(\alpha_1^2 + 4\alpha_2) \geq 0}$, then the roots u_1, u_2 of the characteristic equation are real.

Note that $1 = \rho(0) = A_1 + A_2$

$$\rho(1) = A_1u_1 + A_2u_2$$
$$= A_1u_1 + (1-A_1)u_2$$
$$\Rightarrow A_1 = \frac{\rho(1) - u_2}{u_1 - u_2}$$

Again, from (**), $\rho(1) = \alpha_1\rho(0) + \alpha_2\rho(-1)$

$$\Rightarrow \rho(1) = \frac{\alpha_1}{1-\alpha_2} = \frac{u_1+u_2}{1+u_1u_2}$$

Therefore, $A_1 = \dfrac{u_1(1-u_2^2)}{(u_1-u_2)(1+u_1u_2)}$ and

$$A_2 = 1 - A_1 = \frac{-u_2(1-u_1^2)}{(u_1-u_2)(1+u_1u_2)}$$

$$\therefore \rho(k) = \frac{u_1(1-u_2^2)u_1^k - u_2(1-u_1^2)u_2^k}{(u_1-u_2)(1+u_1u_2)}$$

and it consists of a mixture of damped exponentials.

Case II : — If $(\alpha_1^2 + 4\alpha_2) < 0$, the roots u_1, u_2 of the characteristic equations are imaginary.

Let $u_1 = p(\cos\theta + i\sin\theta)$, $u_2 = p(\cos\theta - i\sin\theta)$, where, $0 < p < 1$.

$$\therefore \rho(k) = p^k \{A_1(\cos\theta + i\sin\theta)^k + A_2(\cos\theta - i\sin\theta)^k\}$$

$$= p^k \{A_1(\cos k\theta + i\sin k\theta) + A_2(\cos k\theta - i\sin k\theta)\}$$

$$= p^k \{A_1'\cos k\theta + A_2'\sin k\theta\}, \text{ where } A_1' = A_1 + A_2,\ A_2' = i(A_1 - A_2).$$

For $k=0$, $\rho(0) = 1 = A_1'$.

For $k=1$, $\rho(1) = p(A_1'\cos\theta + A_2'\sin\theta) \Leftrightarrow \rho(1) = p(\cos\theta + A_2'\sin\theta)$

For $k=-1$, $\rho(-1) = p^{-1}(A_1'\cos\theta - A_2'\sin\theta) \Leftrightarrow \rho(-1) = \frac{1}{p}(\cos\theta - A_2'\sin\theta)$.

solving, we get $A_1' = 1$, $A_2' = \frac{1-p^2}{1+p^2}\cot\theta = \cot\psi$, say.

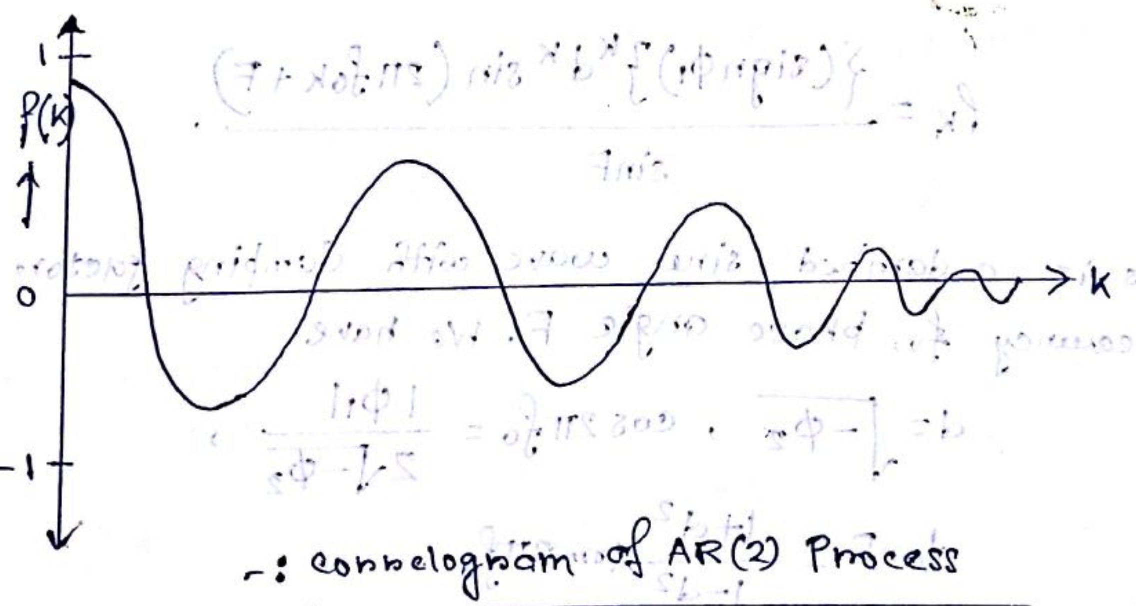
Then we can express $\rho(k)$ as $\rho(k) = p^k\{\cos\theta k + \cot\psi \cdot \sin\theta k\}$

$$= p^k \cdot \frac{\sin(\theta k + \psi)}{\sin\psi},$$

where, p^k is called the 'damping' factor.

since, $0 < p < 1$; and $|\sin(\theta k + \psi)| \leq 1$.

Hence, the correlogram of the process is a damped sine wave and will be damped out of existance if the lag k is quite large.



-: correlogram of AR(2) Process

Result:- The correlogram of an AR(2) $[y_t = a_t + \phi_1 y_{t-1} + \phi_2 y_{t-2}]$ process is a damped sine wave with damping factor $d = \sqrt{-\phi_2}$, frequency $f_0 = \frac{1}{2\pi} \cos^{-1}\left(\frac{|\phi_1|}{2d}\right)$ and the phase angle $F = \tan^{-1} \frac{1+d^2}{1-d^2} \tan(2\pi f_0)$, if the root of the characteristic equation $1 - \phi_1(B) - \phi_2 B^2 = 0$ are complex.

Proof:-

For an AR(2) process, the a.c.f. is given by,

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

Starting values are: $\rho_0 = 1$, $\rho_1 = \phi_1 + \phi_2 \rho_1$, i.e. $\rho_1 = \frac{\phi_1}{(1-\phi_2)}$.

If the roots of the characteristic equation

$$1 - \phi_1 B - \phi_2 B^2 = 0$$ are G_1^{-1}, G_2^{-1}, then

we can write

$$\rho_k = A_1 G_1^k + A_2 G_2^k, \quad |G_j| < 1 \ \forall\ j = 1, 2.$$

When roots are all real, i.e., $\phi_1^2 + 4\phi_2 \geq 0$, ρ_k consists a mixture of damped oscillations. If roots are complex, ρ_k displays pseudo-periodic behaviour. Substituting $G_1 = de^{i2\pi f_0}$, $G_2 = de^{-i2\pi f_0}$,

$$\rho_k = \frac{\{(\text{sign}\,\phi_1)\}^k d^k \sin(2\pi f_0 k + F)}{\sin F}.$$

This is a damped sine wave with damping factor d, frequency f_0, phase angle F. We have

$$d = \sqrt{-\phi_2}, \quad \cos 2\pi f_0 = \frac{|\phi_1|}{2\sqrt{-\phi_2}};$$

$$\tan F = \frac{1+d^2}{1-d^2} \tan 2\pi f_0.$$

Note:- Consider: $y_t = 0.75\, y_{t-1} - 0.50\, y_{t-2} + a_t$.

Here, $\rho_1 = \frac{0.75}{1+0.50} = 0.5$, the roots of the ch. equation $1 - 0.75B + 0.5B^2 = 0$ are complex. the damping factor $d = \sqrt{0.50} = 0.71$; frequency $f_0 = \frac{\cos^{-1}(0.5303)}{2\pi} = \frac{1}{6.2}$; fundamental period of oscillation is 6.2.

Example 1. Show that the a.c.f. of the stationary second-order AR process $X_t = \frac{1}{12}X_{t-1} + \frac{1}{12}X_{t-2} + Z_t$ is given by

$$\rho(k) = \left(\frac{45}{77}\right)\left(\frac{1}{3}\right)^{|k|} + \left(\frac{32}{77}\right)\left(-\frac{1}{4}\right)^{|k|}, \quad k = 0, \pm 1, \pm 2, \ldots$$

Solution:

In order to find the a.c.f. of the process, we use the Yule-Walker equation for $k > 0$,

$$\rho(k) = \frac{1}{12}\rho(k-1) + \frac{1}{12}\rho(k-2) \quad \text{------} (*)$$

We could use these equations to find $\rho(1), \rho(2), \rho(3)$ and so on by successive substitution, but it is easier to find the general solution by solving the set of Yule-Walker equation as a set of difference equations, The general form of the above Yule-walker equation has the auxilary equation

$$y^2 = \frac{1}{12}y + \frac{1}{12}$$

$$\text{i.e. } 12y^2 - y - 1 = 0$$

$$\Rightarrow (4y+1)(3y-1) = 0$$

$$\text{i.e. } y = -\frac{1}{4}, \frac{1}{3}$$

The general solution of the difference equation $(*)$ is given by,

$\rho(k) = A_1 y_1^k + A_2 y_2^k = A_1\left(\frac{1}{3}\right)^k + A_2\left(-\frac{1}{4}\right)^k$, where A_1, A_2 being found from the initial conditions.

We have, $\rho(0) = 1$; $\rho(1) = \frac{1}{12}\rho(0) + \frac{1}{12}\rho(-1)$, from $(*)$

$$\Rightarrow \rho(1) = \frac{1}{11} \quad [\because \rho(1) = \rho(-1)]$$

Hence, $1 = \rho(0) = A_1 + A_2$; $\frac{1}{11} = \rho(1) = A_1\left(\frac{1}{3}\right) + A_2\left(-\frac{1}{4}\right)$

$$\Rightarrow A_1 = 1 - A_2; \quad \Rightarrow A_1 = \frac{45}{77}; \; A_2 = \frac{32}{77}.$$

Hence, $\rho(k) = \frac{45}{77}\left(\frac{1}{3}\right)^k + \frac{32}{77}\left(-\frac{1}{4}\right)^k$, $k > 0$ and also $\rho(k) = \rho(-k)$.

Therefore, $\rho(k) = \frac{45}{77}\left(\frac{1}{3}\right)^{|k|} + \frac{32}{77}\left(-\frac{1}{4}\right)^{|k|}$, $k = 0, \pm 1, \pm 2, \ldots$

Example: 2. Consider the AR(2) process is given by

$$X_t = X_{t-1} - \frac{1}{2}X_{t-2} + Z_t .$$

Is this process stationary? If so, what is its a.c.f.?

Solution:- In order to answer the first equation we find the roots of equation

$$\phi(B) = 1 - \alpha_1 B - \cdots - \alpha_p B^p = 0$$

in this case, $\phi(B) = 1 - B + \frac{1}{2}B^2 = 0$

The roots of this equation (regarding B as a variable) are complex, namely, $1 \pm i$. As the modulus of both roots exceeds one, the roots are both outside the unit circle and so the process is stationary.

The difference equation for $\rho(k)$ of the process is

$$\rho(k) = \rho(k-1) - \frac{1}{2}\rho(k-2) \quad \text{———} (*)$$

The characteristic or auxiliary equation of the difference equation or Yule Walker equation (*) is

$$u^2 - u + \frac{1}{2} = 0, \text{ with roots,}$$

$$u = \frac{1 \pm i}{2} = \frac{1}{\sqrt{2}}\left(\cos\frac{\pi}{4} \pm i\sin\frac{\pi}{4}\right)$$

The general solution of the difference equation (*) is

$$\rho(k) = A_1 u_1^k + A_2 u_2^k$$

$$= A_1\left\{\frac{1}{\sqrt{2}}\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right)\right\}^k + A_2\left\{\frac{1}{\sqrt{2}}\left(\cos\frac{\pi}{4} - i\sin\frac{\pi}{4}\right)\right\}^k$$

$$= \left(\frac{1}{\sqrt{2}}\right)^k \left\{\left(\cos\frac{k\pi}{4}\right)A_1^* + \left(\sin\frac{k\pi}{4}\right)A_2^*\right\}, \text{ say.}$$

We have, $\rho(0) = 1$, and $\rho(1) = \rho(0) - \frac{1}{2}\rho(-1)$

$$\Rightarrow \rho(1) = \frac{2}{3} = \rho(-1)$$

$$\therefore 1 = \rho(0) = A_1^*, \quad \frac{2}{3} = \rho(-1) = \left(\frac{1}{\sqrt{2}}\right)^{-1}\left\{\frac{1}{\sqrt{2}}A_1^* - \frac{1}{\sqrt{2}}A_2^*\right\}$$

$$\Rightarrow \frac{2}{3} = A_1^* - A_2^*,$$

$$\Rightarrow A_2^* = \frac{1}{3}.$$

Hence, $\rho(k) = \left(\frac{1}{\sqrt{2}}\right)^k\left\{\cos\frac{\pi k}{4} + \frac{1}{3}\sin\frac{\pi k}{4}\right\}, \; k = 0, 1, 2, \ldots$

Alternative way to check stationarity for AR(1) and AR(2) Process:-

- Suppose for AR(1) process : $X_t = \alpha X_{t-1} + Z_t$

  or, $Z_t = (1-\alpha B) X_t$

Now, consider, $\Phi(B) = (1-\alpha B)$

$$\psi(B) = (1-\alpha B)^{-1} = \sum_{j=0}^{\infty} (\alpha B)^j$$

which has to be $< \infty \ \forall B$ with $|B| \le 1$.

This requires $|\alpha| \le 1$ to ensure stationarity.

The roots of the characteristic equation $\Phi(B)=0$ is $B = 1/\alpha$. Since $|\alpha| \le 1$, the root should lie in the region $|B|>1$ for stationarity.

- For AR(2) Process :

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t$$

$$\text{or, } Z_t = (1-\alpha_1 B - \alpha_2 B^2) X_t$$

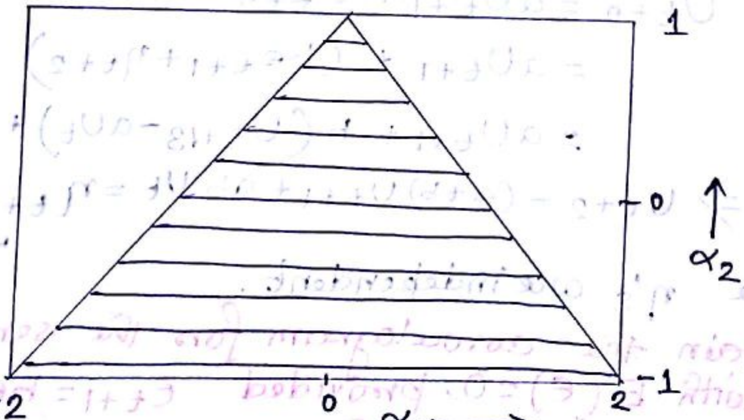
$$\therefore \psi(B) = (1-\alpha_1 B - \alpha_2 B^2)^{-1}.$$

For stationarity roots of the characteristic equation

$\Phi(B) = 0$, i.e.

$(1-\alpha_1 B - \alpha_2 B^2) = 0$

should lie outside the unit circle, which implies that the parameters α_1, α_2 must lie inside the triangular region.



$\alpha_1 + \alpha_2 < 1$, $\alpha_2 - \alpha_1 < 1$, i.e. $|\alpha_2| < 1$ are the stationarity conditions for AR(2) Process.

IMPORTANT QUESTIONS:—

1. For the series determined by

$$U_{t+1} = aU_t + \epsilon_{t+1}, \quad |a|<1 \qquad (*)$$

where ϵ has zero mean, find the correlogram if

(i) successive values of ϵ are independent,

(ii) If ϵ itself obey's a relation of the form

$$\epsilon_{t+1} = b\epsilon_t + \eta_{t+1}, \quad |b|<1, \qquad (**)$$

where successive values of η are independent. [ISS EXAM'12] [10 Marks]

Solution: (i) Here correlogram means auto-correlation function.

$$U_{t+1} = aU_t + \epsilon_{t+1}, \quad |a|<1$$

It's an AR Process

$$\therefore U_t = 1\epsilon_t + a\epsilon_{t-1} + a^2\epsilon_{t-2} + \cdots$$

a moving average of random elements of infinite extent with weights

$$\xi_j = a^{j-1} \quad \forall\, j = 1,2,3,\ldots$$

$$\therefore r_K = \frac{\sum_{j=1}^{\infty} \xi_j \xi_{j+k}}{\sum_{j=1}^{\infty} \xi_j^2} = \frac{1.a^k + a.a^{k+1} + a^2 k^{k+2} + \cdots}{1+a^2+a^4+\cdots}$$

$$\therefore r_K = a^k.$$

(ii) Here we shall first estimate ϵ's in (*) and (**). We have from (*)

$$U_{t+2} = aU_{t+1} + \epsilon_{t+2}$$

$$= aU_{t+1} + (b\epsilon_{t+1} + \eta_{t+2}) \qquad [\text{From}(**)]$$

$$= aU_{t+1} + b(U_{t+1} - aU_t) + \eta_{t+2} \; [\text{From}(*)]$$

$$\Rightarrow U_{t+2} - (a+b)U_{t+1} + ab\,U_t = \eta_{t+2}, \; |a|<1, \; |b|<1 \qquad (1)$$

where η's are independent.

[ISS] Ques:- Obtain the correlogram for the series $U_{t+1} = aU_t + \epsilon_{t+1}$, $|a|<1$ with $E(\epsilon)=0$, provided $\epsilon_{t+1} = b\epsilon_t + \eta_{t+1}, |b|<1$, and successive values of η are independent.

Multiplying both sides of ① by U_{t-k}, taking expectations and dividing by $Var(U_t)$, we get for a long series,

$$r_{k+2} - (a+b)r_{k+1} + ab\,r_k = 0 \quad \text{———— ②}$$

which is a (homogeneous) linear difference equation of order 2 in r_k. The trial solution $r_k = {}^k$ gives

$$-(a+b)\lambda + ab = 0$$

$$\Rightarrow \lambda = a, b$$

$$\therefore r_k = A.a^k + B.b^k \quad \text{———— ③}$$

is the general solution of ②, A and B being arbitrary constants which are determined from the fact that

$r_0 = 1$ and $r_{-1} = r_1$

$$r_0 = 1 \Rightarrow A + B = 1 \quad \text{———— ④}$$

and $r_1 = r_{-1} \Rightarrow \frac{A}{a} + \frac{B}{b} = A.a + B.b$

$$\Rightarrow A\left(a - \frac{1}{a}\right) + B\left(b - \frac{1}{b}\right) = 0 \quad \text{———— ⑤}$$

Solving ④ and ⑤, we get

$$A = \frac{a(b^2-1)}{(b-a)(1+ab)}, \quad B = \frac{b(1-a^2)}{(b-a)(1+ab)}$$

Hence from ③, we get,

$$r_k = \frac{1}{(b-a)(1+ab)}\left[a^{k+1}(b^2-1) + b^{k+1}(1-a^2)\right]$$

2. Explain how a correlogram can be used in determining among different schemes of oscillatory movement in a stationary TS.

Ans:-

3. For the auto-regressive scheme,

$$U_{t+2} + aU_{t+1} + bU_t = \epsilon_{t+2} \quad \text{———} (*)$$

show that if ϵ is random and the series is long

$$\frac{Var(U)}{Var(\epsilon)} = \frac{1+b}{(1-b)\{(1+b)^2 - a^2\}}$$

and hence that the variance of the generated series may be much greater than that of ϵ itself.

[ISS EXAM'11 10 Marks]

Solution :-

Since the series is long and $E(\epsilon_t) = 0$, we have

$$E(U_t) = 0 \quad \text{and} \quad Var(U_t) = E(U_t^2) = E(U_{t+2}^2)$$

Squaring both sides of (*) and taking expectations, we get

$$Var(U)[1 + a^2 + b^2 + 2ar_1 + 2br_2 + 2abr_1] = Var(\epsilon)$$

$$\Rightarrow \frac{Var(U)}{Var(\epsilon)} = \frac{1}{1 + a^2 + b^2 + 2a(1+b)r_1 + 2br_2} \quad \text{———} (1)$$

Multiplying both sides of (*) by U_{t+1} and U_t, taking expectations and dividing by $Var(U_t)$, we get respectively,

$$r_1 + a + br_1 = 0 \quad \Rightarrow \quad r_1 = -a/(1+b)$$

$$\text{and} \quad r_2 + ar_1 + b = 0 \quad \text{and} \quad r_2 = -b - ar_1 = \frac{(a^2 - b - b^2)}{(1+b)}$$

Substituting in (1), we get, the denominator of R.H.S. as

$$\text{Denominator} = 1 + a^2 + b^2 - 2a^2 + \frac{2b}{(1+b)}(a^2 - b - b^2)$$

$$= \frac{1}{(1+b)}\left[(1+b)(1 - a^2 + b^2) + 2b\{a^2 - b(1+b)\}\right]$$

$$= \frac{1}{(1+b)}\left[1 - a^2 - b^2 + b - b^3 + a^2 b\right]$$

$$= \frac{1}{(1+b)}\left[(1-b^2) + b(1-b^2) - a^2(1-b)\right]$$

$$= \frac{1}{(1+b)}\left[(1-b)\{(1+b)^2 - a^2\}\right]$$

substituting in (1), we get the required result.

variance of the generated series will be much greater than that of ϵ_t if $\dfrac{Var(U)}{Var(\epsilon)} > 1$

or if $(1+b) > (1-b)\left[(1+b)^2 - a^2\right]$

or if $(a^2+b^2) - b(a^2-b^2) > 0$

or if $\dfrac{a^2+b^2}{a^2-b^2} > b$

which is always true since $|b| < 1$.

4. Let $\epsilon_t, \epsilon_{t+1}, \dots\dots$ and ξ be independent variables with zero mean and unit variance. Making

$$U_t = a\xi + \epsilon_t, \quad -\infty < t < \infty$$

[ISS EXAM '11 8 Marks]

Show that the process is stationary with correlation

$$\rho_1 = \rho_2 = \dots\dots = \frac{a^2}{1+a^2}.$$

Solution: — We are given

$$\left.\begin{array}{l} E(\epsilon_t) = 0 = E(\xi); \qquad E(\epsilon_i \epsilon_j) = 0, \; i \neq j \\ \qquad\qquad\qquad\qquad\qquad\qquad\; = 1, \; i = j \\ Var(\epsilon_t) = 1 = Var(\xi) \text{ and } E(\xi.\epsilon_t) = 0, \end{array}\right\} \dots\dots (*)$$

$$\begin{aligned} Var(U_t) &= Var(a\xi + \epsilon_t) \\ &= a^2 V(\xi) + V(\epsilon_t) + 2a\,Cov(\xi, \epsilon_t) \\ &= a^2 + 1. \end{aligned}$$

Similarly, we get $Var(U_{t+k}) = a^2+1$

$$Cov(U_t, U_{t+k}) = E(U_t U_{t+k}) - E(U_t)E(U_{t+k})$$

But $E(U_t) = E(a\xi + \epsilon_t) = 0 = E(U_{t+k})$

$$\begin{aligned} \therefore Cov(U_t, U_{t+k}) &= E(U_t U_{t+k}) \\ &= E\left[(a\xi + \epsilon_t)(a\xi + \epsilon_{t+k})\right] \\ &= a^2 E(\xi^2) + aE(\xi.\epsilon_{t+k}) + aE(\xi\epsilon_t) \\ &\quad + E(\epsilon_t \epsilon_{t+k}) = a^2 \quad \left[\text{Using } (*)\right] \end{aligned}$$

$\therefore$ Auto correlation of order k is given by,

$$\rho_k = \frac{Cov(U_t, U_{t+k})}{\sqrt{Var(U_t)\,Var(U_{t+k})}} = \frac{a^2}{1+a^2}, \; (k = 1, 2, 3, \dots\dots)$$

which is independent of k. Hence the process is stationary.

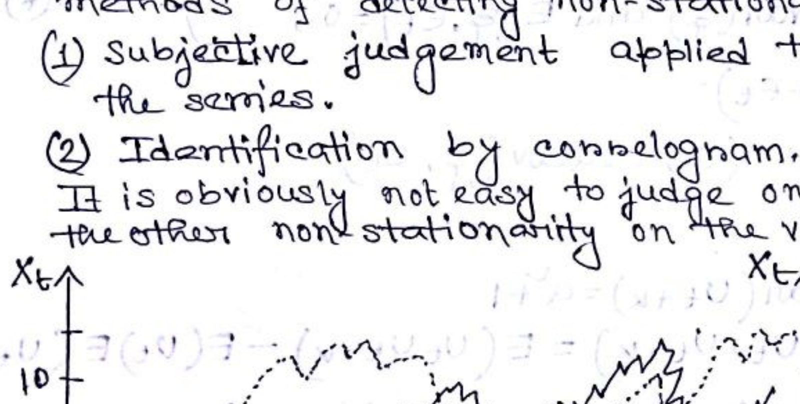
A stationary Process: — If in a sample correlogram; the values of r_K do not come down to zero reasonably quickly, indicates non-stationarity of the process. For stationary series, the sample correlogram is with the theoretical or population correlogram of different stationary processes in order to choose the one which is most appropriate.
The correlogram of an MA(q) process is easy to recognize as it "cuts off" at lag q, where as the correlogram of an AR(p) process is a mixture of damped exponentials and sinusoids and dies out slowly. For example, suppose we find that r_1 is significantly different from zero but that subsequent values of r_K are all close to zero, then an MA(1) model is indicated. Since its theoretical correlogram is of this form. Alternatively, if $r_1, r_2, \ldots, r_p$ appear to be decreasing exponentially, then an AR(1) model may be appropri

Graphical inspection for Stationarity: — There are two principal methods of detecting non-stationarity,

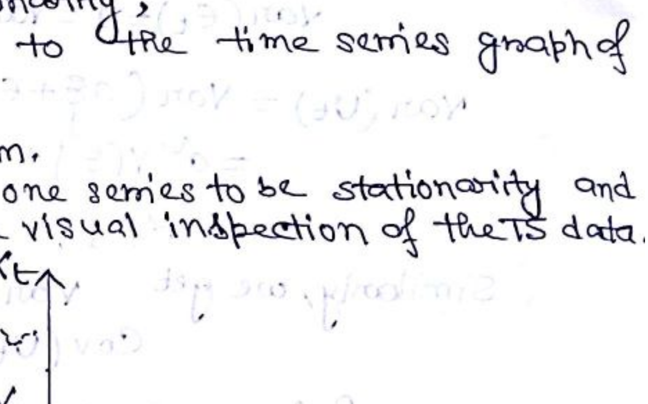
(1) Subjective judgement applied to the time series graph of the series.

(2) Identification by correlogram.

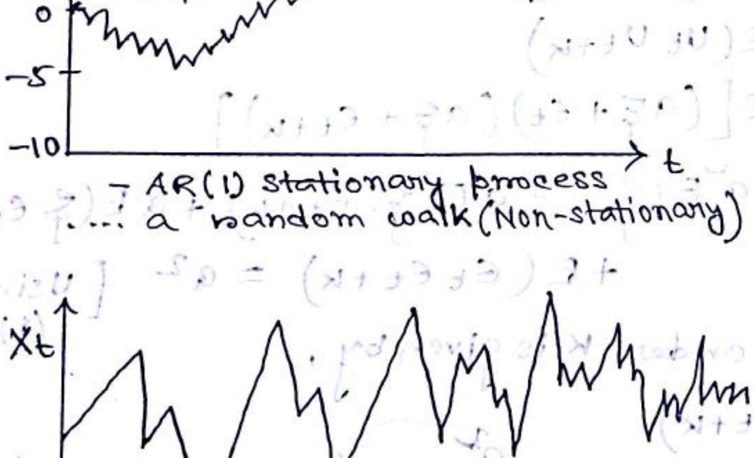
It is obviously not easy to judge one series to be stationarity and the other non-stationarity on the visual inspection of the TS data.



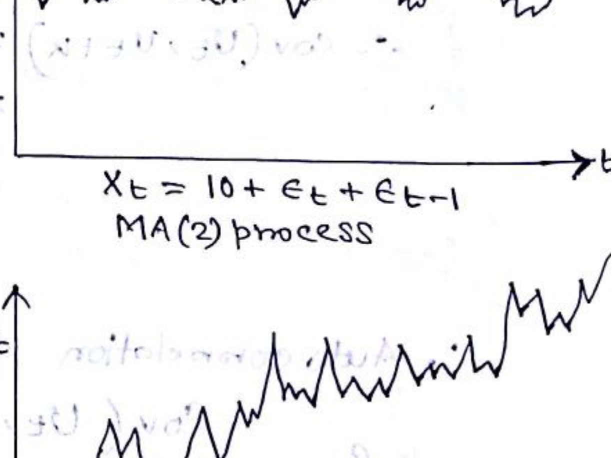
— AR(1) stationary process
.... a random walk (Non-stationary)



$X_t = 10 + \epsilon_t + \epsilon_{t-1}$
MA(2) process



$X_t = 0.75X_{t-1} - 0.50X_{t-2} + \epsilon_t$
Stationary Process



$X_t = 1.50X_{t-1} + \epsilon_t$: AR(1)
Non-stationary process

Autoregressive Parameters in terms of Autocorrelations of an AR(p) process (Yule-Walker Equation) :-

An important recurrence relation for the autocorrelation function of a stationary AR process is found by multiplying throughout in

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + \epsilon_t$$

by X_{t-k}, to obtain

$$X_{t-k}X_t = \alpha_1 X_{t-k}X_{t-1} + \alpha_2 X_{t-k}X_{t-2} + \cdots + \alpha_p X_{t-k}X_{t-p} + X_{t-k}\epsilon_t$$

Therefore,

$$E(X_{t-k}X_t) = \alpha_1 E(X_{t-k}X_{t-1}) + \alpha_2 E(X_{t-k}X_{t-2}) + \cdots + \alpha_p E(X_{t-k}X_{t-p}) + E(X_{t-k}\epsilon_t)$$

$$\Rightarrow \gamma(k) = \alpha_1\gamma(k-1) + \alpha_2\gamma(k-2) + \cdots + \alpha_p\gamma(k-p) \quad (*)$$

Note that $E(X_{t-k}\epsilon_t) = 0$ for $k > 0$, since X_{t-k} can only involve the errors ϵ_j up to time $t-k$ which are uncorrelated with ϵ_t.
On dividing throughout in (*) by $\gamma_0 = V(X_t)$, we have the following difference equation:

$$\rho(k) = \alpha_1\rho(k-1) + \alpha_2\rho(k-2) + \cdots + \alpha_p\rho(k-p), \quad k = 1(1)p.$$

Note that if we write $\underset{\sim}{\alpha} = (\alpha_1, \ldots, \alpha_p)'$ and $\underset{\sim}{\rho} = (\rho(1), \ldots, \rho(p))'$

and $R_p = \begin{bmatrix} 1 & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & 1 & \cdots & \rho(p-2) \\ \vdots & \vdots & & \vdots \\ \rho(p-1) & \rho(p-2) & \cdots & 1 \end{bmatrix}$, R_p is <u>full rank</u> and square & symmetric, so that <u>invertibility</u> is guaranteed. Here $\rho(0) = 1$.

Hence, $\underset{\sim}{\alpha} = R^{-1}\underset{\sim}{\rho}$ or $\underset{\sim}{\rho} = R_p\underset{\sim}{\alpha}$

The above equations are called <u>Yule-Walker equations</u>. The parameters α_k are estimated by replacing ρ_k by r_k.

<u>Note</u>:-

$$\underbrace{\begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_p \end{pmatrix}}_{\underset{\sim}{\rho}} = \underbrace{\begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{pmatrix}}_{R_{p\times p}} \underbrace{\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix}}_{\underset{\sim}{\alpha}}$$

Note that this is a well-posed system (with a square coefficients matrix R_p), i.e., with the **same** numbers of constraints (equations, R's rows) as unknowns (the elements α_j of the unknown vector $\underset{\sim}{\alpha}$). $\hat{\underset{\sim}{\alpha}} = R^{-1}\underset{\sim}{\rho}$.

Fitting of an Autoregressive Process:—

If an AR process is thought to be appropriate, there are two related questions:
(a) How can we estimate the parameters of the process?
(b) What is the order of the process?

Solution:-

(a) Estimating the parameters of an autoregressive process:-

We obtain Yule-Walker estimates of the parameters by replacing the theoretical auto correlations $\rho(k)$ by the sample auto correlations r_K. In matrix form these equations are

$$\underset{\sim}{r}_p = R_p \underset{\sim}{\hat{\alpha}}, \text{ where } R = \begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & 1 & r_1 & \cdots & r_{p-2} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & 1 \end{bmatrix}$$

and $\underset{\sim}{\hat{\alpha}} = (\hat{\alpha}_1, \ldots\ldots, \hat{\alpha}_p)$, $\underset{\sim}{r}_p = (r_1, r_2, \ldots, r_p)$

Hence, $\underset{\sim}{\hat{\alpha}} = R_p^{-1} \underset{\sim}{r}_p$.

AR(1) Process:— $X_t = \alpha X_{t-1} + \epsilon_t$

The Yule-Walker equation for AR(1) process is

$$\rho(k) = \alpha \rho(k-1) \quad \left[\text{Hence, } \rho(k) = \alpha \rho(k-1) = \alpha^2 \rho(k-2) = \cdots \cdots = \alpha^{k-1}\rho(1) = \alpha^k \right]$$

Yule-walker equation for the process: $\rho(1) = \alpha$; —— (*)

therefore, the Yule Walker estimate of α is $\hat{\alpha} = r_1$, the sample auto-correlation of lag 1.

AR(2) Process:— $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \epsilon_t$

The Yule-Walker equations are

$$\rho(1) = \alpha_1 + \alpha_2 \rho(1)$$

$$\rho(2) = \alpha_1 \rho(1) + \alpha_2$$

Hence, $\alpha_1 = \dfrac{\rho_1(1-\rho_2)}{1-\rho_1^2}$, $\alpha_2 = \dfrac{\rho_2 - \rho_1^2}{1-\rho_1^2}$

$\rho_1 = \dfrac{\alpha_1}{1-\alpha_2}$, $\rho_2 = \alpha_2 + \dfrac{\alpha_1^2}{1-\alpha_2}$

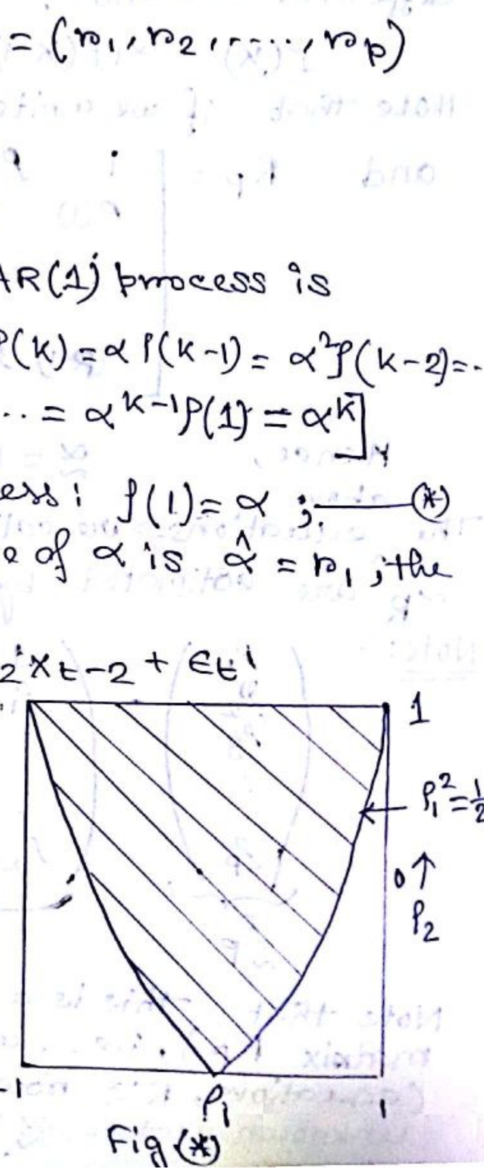


Fig (*)

For stationarity α_1, α_2 should lie in the region

$$\alpha_1+\alpha_2<1, \quad \alpha_2-\alpha_1<1 \quad, \text{i.e.,} |\alpha_2|<1.$$

This gives admissible region of ρ_1, ρ_2 as (Fig(*))

$$|\rho_i|<1, \; i=1,2, \quad \rho_1^2<\frac{1}{2}(\rho_2+1)$$

Alternative method :—

For AR(2) Process :—

Consider the AR(2) Process: $X_t=\alpha_1 X_{t-1}+\alpha_2 X_{t-2}+\epsilon_t$.

The Yule Walker equations are ——

$\rho(k)=\alpha_1\rho(k-1)+\alpha_2\rho(k-2) \;\; \forall \, k=1,2.$, which is in the matrix form

$$\begin{bmatrix}\rho(1)\\ \rho(1)\end{bmatrix}=\begin{bmatrix}1 & \rho(1)\\ \rho(1) & 1\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix} \quad \text{——————} \quad (**)$$

The Yule-Walker estimates of α_1, α_2 are given by

$\begin{bmatrix}\hat\alpha_1\\ \hat\alpha_2\end{bmatrix}=\begin{bmatrix}1 & r_1\\ r_1 & 1\end{bmatrix}^{-1}\begin{bmatrix}r_1\\ r_2\end{bmatrix}$, which is obtained from (**) by replacing $\rho(k)$ by r_k.

$$\therefore \begin{bmatrix}\hat\alpha_1\\ \hat\alpha_2\end{bmatrix}=\frac{1}{1-r_1^2}\begin{bmatrix}1 & -r_1\\ -r_1 & 1\end{bmatrix}\begin{bmatrix}r_1\\ r_2\end{bmatrix}$$

$$=\frac{1}{1-r_1^2}\begin{bmatrix}r_1(1-r_2)\\ r_2-r_1^2\end{bmatrix}$$

$$\Rightarrow \hat\alpha_1=\frac{r_1(1-r_2)}{1-r_1^2} \text{ and } \hat\alpha_2=\frac{r_2-r_1^2}{1-r_1^2}.$$

(b) Determining the order of an AR Process :—

When fitting an AR(p) model, the last coefficients, α_p will be a measure of excess correlation at lag p which is not accounted for by an AR(p-1) model. It is also called the p th partial auto correlation coefficient. For an AR(1) process, $\alpha = \rho(1)$ and for an AR(2) process

$$\alpha_2 = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}$$ and note that $\alpha_2 = 0$ for an AR(1) process where $\rho(2) = \alpha^2 = \rho(1)^2$. This means that if we fit an AR(2) model what is really an AR(1), then $\alpha_2 = 0$. The estimate of α_k is estimated by fitting AR processes of successively higher order. Values of $\hat{\alpha}_p$ which are outside the range $\pm \frac{2}{\sqrt{N}}$ are significantly different from '0' at 5% level.
If $H_0 : \alpha_p = 0$ is accepted, then the correct order of the AR process is $(p-1)$.

Another approach is to fit AR processes of successively higher order, to calculate the residual sum of squares for each 'p' and to plot this against p. It may then be possible to see the value of 'p' where the curve 'flattens out' and the addition of extra parameters gives little improvement in fit.

FORECASTING :-

Forecasting the future values of an observed time series is an important problem in many cases or areas, including economics, sales and stock control.

Suppose we have an observed time series $x_1, x_2, \dots, x_N$. Then the basic problem is to estimate future values such as x_{N+K}, where the integer K is called the lead time. The forecast of x_{N+K} made at time N for K steps ahead will be denoted by $\hat{x}(N, K)$.

1. Exponential Smoothing :-

This forecasting procedure, first suggested by C.C. Holt in about 1958, should only be used in its basic form for non-seasonal time series showing no systematic trend. Of course many time-series which arise in practice do contain a trend or seasonal pattern, but these effects can be measured and removed to produce a stationary series.

Given a non-seasonal time-series with no systematic trend $x_1, x_2, \dots, x_N$; it is natural to take as an estimate of x_{N+1}, a weighted sum of the past observations:

$$\hat{x}(N, 1) = c_0 x_N + c_1 x_{N-1} + c_2 x_{N-2} + \cdots \cdots \longrightarrow (*)$$

where $\{c_i\}$ are weights. It seems sensible to give more weight to recent observations and less weight to observations further in the past. An intuitively appealing set of weights are geometric weights, which decrease by a constant ratio. In order that the weights sum to one, we take

$$c_i = \alpha(1-\alpha)^i, \quad i = 0, 1, 2, \dots$$

where α is a constant such that $0 < \alpha < 1$. Then $(*)$ becomes

$$\hat{x}(N, 1) = \alpha x_N + (1-\alpha)\alpha x_{N-1} + \alpha(1-\alpha)^2 x_{N-2} + \cdots \longrightarrow (**)$$

Strictly speaking, equation $(**)$ implies an infinite number of past observations, but in practice there will only be a finite number. So equation $(**)$ is rewritten, in the form:

$$\begin{aligned}\hat{x}(N, 1) &= \alpha x_N + (1-\alpha)\{\alpha x_{N-1} + \alpha(1-\alpha) x_{N-2} + \cdots\} \\ &= \alpha x_N + (1-\alpha)\hat{x}(N-1, 1) \longrightarrow (***)\end{aligned}$$

If we set $\hat{x}(1,1) = x_1$, then equation (***) can be used recursively to compute forecasts. Equation (***) also reduces the amount of arithmatic involved since forecasts can easily be updated using only the latest observation and the previous forecast. The procedure defined by equation (***) is called <u>exponential smoothing.</u>

The value of the smoothing constant α depends on the properties of the given time series. Values between 0.1 and 0.3 are commonly used and produce a forecast which depends on a large number of past observations.

The value of α may be estimated from past data. The sum of squared prediction errors is computed for different values of α and the value is chosen which minimizes the sum of squares. With a given value of α, calculate

$$\hat{x}(1,1) = x_1$$
$$e_2 = x_2 - \hat{x}(1,1)$$
$$\hat{x}(2,1) = \alpha e_2 + \hat{x}(1,1)$$
$$e_3 = x_3 - \hat{x}(2,1)$$
$$\vdots$$
$$e_N = x_N - \hat{x}(N-1,1) \text{ and compute } \sum_{i=2}^{N} e_i^2 .$$

Repeat this procedure for other values of α between 0 and 1, say in steps of 0.1, and select the value which minimizes $\sum_{i=2}^{N} e_i^2$.

[2]. The Holt-Winters forecasting procedure : Exponential smoothing may be generalized to deal with time series containing trend and seasonal variation. The resulting procedure is referred to as "Holt-Winters procedure". Trend and seasonal-terms are introduced which are also updated by exponential smoothing.

Suppose the observations are monthly, and let L_t, T_t, I_t denote the local level, trend and seasonal index, respectively, at time t. Thus T_t is the expected increase or decrease per month in the current level. Let α, γ, δ denote three smoothing parameters for updating the level, trend and seasonal index, respectively. The smoothing parameters are usually chosen in the range (0,1). Then, when a new observation x_i becomes available, the values of L_t, T_t and I_t are all updated. If the seasonal variation is multiplicative, then the (recurrence form) updating equations are

$$L_t = \alpha\left(\frac{x_t}{I_{t-12}}\right) + (1-\alpha)\left(L_{t-1} + T_{t-1}\right)$$

$$T_t = \gamma\left(L_t - L_{t-1}\right) + (1-\alpha)T_{t-1}.$$

$$I_t = \delta\left(\frac{x_t}{L_t}\right) + (1-\delta)I_{t-12}$$ and the forecasts from time 't' are then

$$\hat{x}(t,K) = \left(L_t + KT_t\right)I_{t-12+K}, \text{ for } K = 1(1)12.$$

In order to apply the method, the user must carry out the following steps:

(a) Provide starting values for L_t, T_t and I_t at the beginning of the series.

(b) Estimate values for α, γ, δ by minimizing $\sum e_t^2$ over a suitable fitting period for which data are available.

(c) Decide whether or not to normalize the seasonal indices at regular intervals.

Write a short note on forecasting:-

Forecasting (or, prediction) refers to the process of generating future values of a particular event. The final goal of a forecast is to make decisions based on the future value(s) of some variable.

Suppose we have N observed time series values $x_1, x_2, \ldots, x_N$. Then it may be of common interest to estimate the time series values in the future time point on the basis of the existing N past observations $x_1, x_2, \ldots, x_N$, i.e. we use the data upto time N, to make forecasts $\hat{X}_N(1), \hat{X}_N(2), \ldots, \hat{X}_N(m)$ of future values of X, where,

$\hat{X}_N(h)$ = Estimate of the TS at the future time points (N+h) which means the h step ahead estimate of time series given the TS upto the time point N are available, $h = 1(1)m$.

Here, h is an integer, called the lead time or the forecasting horizon.

The methods to be considered here are conventionally regarded as being divided in two groups:

1. Averaging methods, and
2. Exponential smoothing methods.

Though it is convenient to follow this convention it is important to realize at the outset that this distinction is artificial in that all the methods discussed here are methods based on averages. They are thus all similar to the moving averages. The difference is that the averages are used for forecasting rather than for describing past data.

The point of potential confusion is made coorse by the use of the name 'exponential smoothing' for the second group. These methods are also based on weighted averages, where the weights decay in an exponential way from the most recent to the most distant data point. The term 'smoothing' is being used simply to indicate that this weighted average smoothes the data irregularities. Thus, though the term smoothing here is used in the same sense as previously, the smoothing is being carried out in a different context.

1. Averaging Method : The moving average forecast of order K, which we write as MA(K), is defined as

$$\hat{X}_N(1) = \frac{1}{K}\sum_{i=N-K+1}^{N} X_i$$

This forecast is only useful if the data does not contain a trend-cycle or a seasonal component. In other words, the data must be stationary. Data is said to be stationary if X_t, which is an r.v., has a probability distribution that does not depend on t.

A convenient way of implementing the forecast is to note that

$$\hat{X}_N(2) = \frac{1}{K}\sum_{i=N-K+2}^{N+1} X_i$$

$$= \hat{X}_N(1) + \frac{1}{K}\left(X_{N+1} - X_{N-K+1}\right)$$

This is known as an updating formula as it allows a forecast value to be obtained from the previous forecast value by a simpler calculation than using the defining expression.

The only point of note, is that moving average forecasts give a progressively smoother forecast as the order increases, but a moving average of large order will be slow to respond to real but rapid changes. Thus, in choosing K, a balance has to be drawn between smoothness and ensuring that this lag is not unacceptably large.

2. Simple Exponential Smoothing (SES) : This is nothing but a general form of forecasting method.

This method be applied to a non-seasonal time series having no systematic trend. But in practice, frequently a time series involve seasonality and or systematic trend. In such cases, removing the seasonality and or systematic trend from time series we should make it stationary and then we should apply SES to this time series.

Here $\hat{X}_N(1)$ can be expressed as a linear combination of the given observations $X_1, X_2, \ldots, X_N$ i.e.,

$$\hat{X}_N(1) = a_0 X_N + a_1 X_{N-1} + a_2 X_{N-2} + \cdots \quad \text{--- (1)}$$

where, the coefficients $a_0, a_1, a_2, \ldots$ denote the weightage of the associated time series value for forecasting $\hat{X}_N(1)$.

Now, it is a common fact that time series in the recent past of the time point 'N' should have the greater weightage than those in the remote past with this logic, X_N should have highest weight, X_{N-1} should have the second highest weight and so on, i.e. weights must be gradually decreasing as the lag of the time series increases. In this perspective an intuitively logical set of weights will be geometric weights of the forms

$$a_i = \alpha(1-\alpha)^i \ , \ i = 0,1,2,\ldots\ldots$$

where α is a constant such that $0<\alpha<1$.

Here, a_i is decreasing in i with common ratio $(1-\alpha)$.

Hence the forecasting model will be ——

$$\hat{X}_N(1) = \alpha X_N + \alpha(1-\alpha)X_{N-1} + \alpha(1-\alpha)^2 X_{N-2} + \cdots\cdots \text{———} \quad (2)$$

Equation (2) implies an infinite numbers of past observations, but in practice there will only be a finite numbers, thus eqn (2) is customarily rewritten in the recurrence form as

$$\hat{X}_N(1) = \alpha X_N + (1-\alpha)\left[\alpha X_{N-1} + \alpha(1-\alpha)X_{N-2} + \cdots\cdots\right]$$

$$= \alpha X_N + (1-\alpha)\hat{X}_{N-1}(1) \text{———} \quad (3)$$

This is the recursion relation in <u>one-step ahead</u> forecasting or prediction.

Repeated application of the formula (2) yields

$$\hat{X}_N(1) = (1-\alpha)^N \hat{X}_0(1) + \alpha \sum_{i=0}^{N-1} (1-\alpha)^i X_{N-i} \text{———} \quad (4)$$

Showing that the dependence of the current forecast on $X_N, X_{N-1}, X_{N-2}, \cdots\cdots$ fall away in an exponential way. The rate at which this dependence falls away in controlled by α. The larger the value of α the quicker does the dependence on previous values fall away.

Thus given the TS at the time point N and forecasts at the same point N we can update the forecast at the time point N+1 using the recursion relation (2). SES needs to be initialized. A sample choice is to use

$$\hat{X}_1(1) = X_1 \ ,$$

in the above recursion relation which reduces the amount of arithmatic in updating a time series value. Other values are possible, but we shall not agonise over this too much as we are more concerned with the behaviour of the forecast once it has been in use for a while.

since the chosen geometric weights decline geometrically, hence the weights are also called exponential weights and the corresponding forecasting procedure as exponential smoothing. In this respect the forecasting procedure can also be named geometric smoothing.

Let us define, $e_r = X_r - \hat{X}_{r-1}(1)$

= error in prediction or prediction error at the time point r.

Let, $e_1 = 0$.

$$e_2 = X_2 - \hat{X}_1(1) = X_2 - X_1$$

$$e_3 = X_3 - \hat{X}_2(1)$$

$$\vdots$$

$$e_N = X_N - \hat{X}_{N-1}(1)$$

In terms of prediction error, the recursion relation can be written as

$$\hat{X}_N(1) = \alpha X_N + (1-\alpha)\hat{X}_{N-1}(1)$$

$$= \alpha\left[X_N - \hat{X}_{N-1}(1)\right] + \hat{X}_{N-1}(1)$$

$$= \alpha e_N + \hat{X}_{N-1}(1).$$

Estimation of α in SES :- Generally, α is taken between 0.1 and 0.3 such that the predicted value can depend on the moderately large numbers of past observations. High value of α is seldomly chosen and in that case, predicted value depends predominantly on the very few recent past observations like $X_N, X_{N-1}, \ldots$, etc.

If $\alpha = 1$, then $\hat{X}_N = X_N$,

i.e. prediction completely depends on the most recent past observation.

In general, a good forecasting model must be encountered with small magnitude of all the prediction errors. Hence, we consider that choice of α as the best one which minimize $\sum_{i=2}^{N} e_i^2$.

In this respect, we study the value of $\sum_{i=2}^{N} e_i^2$ for several choices of α between $(0,1)$ (generally keeping 0.1 unit interval), and choose the best one among them for which $\sum_{i=2}^{N} e_i^2$ is minimum.

The Holt and Holt-Winter's forecasting procedure:

Single exponential smoothing can be applied to a non-seasonal time series having no systematic trend. Here we generalise the idea of SES to deal with TS containing trend and seasonal variation, i.e., to generalise the equations for SES by introducing trend and seasonal terms, which are also updated by exponential smoothing. The version for handling a trend with non-seasonal data is usually called Holt's (two parameters) exponential smoothing, while the version that also copes with seasonal variation is usually referred to as the Holt-Winter's procedure (three parameters).

Holt's Linear Exponential Smoothing (LES):

This is an extension of exponential smoothing (SES) to take into account a possible linear trend. Let,

L_t = estimated current mean in month t.

b_t = estimated trend term in month t (i.e. the expected increase or decrease per month in current mean).

As new observations become available, all the terms are updated. There are two smoothing parameters, viz.,

α, a smoothing constant for the data, and

β, a smoothing constant for the trend estimate.

Then the updating equations are, when a new observation X_t becomes available,

$$L_t = \alpha X_t + (1-\alpha)(L_{t-1} + b_{t-1})$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1}$$

Then the h-step-ahead forecast at time t will be of the form

$$\hat{X}_t(h) = L_t + b_t h, \quad h = 1, 2, 3, \ldots$$

Initial estimates are needed for L_1 and b_1. Simple choices are $L_1 = X_1$ and $b_1 = 0$.

If however zero is a typical of the initial slope than a more careful estimate of the slope may be needed to ensure that the initial forecasts are not badly out.

The parameters α and β are generally chosen to lie in the range $(0,1)$. It is natural to call this the two-parameter version of ES.

Holt-Winter's method :- This is an extension of Holt's LES to take into account seasonality. These are two versions, viz. multiplicative and additive, with the multiplicative model more widely used. Let,

L_t = estimated current mean in month t.
b_t = estimated trend term in month t (i.e. the expected increase or decrease per month in current mean).
S_t = estimated seasonal factors appropriate to month t.

As new observations become available, all the three terms are updated.

There are three smoothing parameters, viz.,
α, a smoothing constant for the data;
β, a smoothing constant for the trend estimate, and
γ, a smoothing constant for the seasonal estimate.

Holt-Winter's method, Multiplicative seasonality :- If the seasonal variation is mutliplicative, i.e. the model is of the form:

$$X_t = L_t \cdot S_t \cdot \epsilon_t \quad \text{or,} \quad X_t = L_t S_t + \epsilon_t,$$

then the updating equations, when a new observation X_t becomes available are:

$$L_t = \alpha \cdot \frac{X_t}{S_{t-s}} + (1-\alpha)(L_{t-1}+b_{t-1}).$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)\, b_{t-1}.$$

$$S_t = \gamma \cdot \frac{X_t}{L_t} + (1-\gamma) S_{t-s}.$$

Then the h-step-ahead forecast at time t will be of the form: $\hat{X}_t(h) = (L_t + b_t h)\, S_{t-s+h}$, $h = 1, 2, \ldots, s$.

where 's' is the no. of periods in one cycle of seasons, e.g., no. of months or quarters in a year.

To initialize we need one complete cycle of data, i.e., s values. Then set

$$L_s = \frac{1}{s}\sum_{t=1}^{s} X_t$$

To initialize trend we use $s+k$ time periods.

$$b_s = \frac{1}{k}\left[\frac{X_{s+1}-X_1}{s} + \frac{X_{s+2}-X_2}{s} + \cdots + \frac{X_{s+k}-X_k}{s}\right]$$

$$= \frac{1}{k}\sum_{t=1}^{k}\left(\frac{X_{s+t}-X_t}{s}\right).$$

If the series is long enough then a good choice is to make $k=s$ so that two complete cycles are used.

However, we can at a pinch, use $k=1$.

Initial seasonal indices can be taken as, $S_k = \dfrac{X_k}{L_s}$, $k = 1(1)s$.

The parameters α, β, γ should lie in the interval $(0,1)$.

Holt-Winter's method; Additive Seasonality: If the seasonal variation is additive, i.e., the model is of the form:

$$X_t = L_t + S_t + \epsilon_t$$

then the updating equations, when a new observation X_t become available, are:

$$L_t = \alpha\left(X_t - S_{t-s}\right) + (1-\alpha)\left(L_{t-1} + b_{t-1}\right)$$

$$b_t = \beta\left(L_t - L_{t-1}\right) + (1-\beta)\, b_{t-1}$$

$$S_t = \gamma\left(X_t - L_t\right) + (1-\gamma)\, S_{t-s}.$$

Then the h-step ahead forecast at time t will be of the form

$$\hat{X}_t(h) = L_t + b_t h + S_{t-s+h}, \quad h = 1(1)s.$$

where, 's' is the no. of periods in the cycle.

The initial values of L_s and b_s can be as in the multiplicative case. The initial seasonal indices can be taken as

$$S_k = X_k - L_s, \quad k = 1(1)s.$$

The parameters α, β, γ should lie in the interval $(0,1)$. A graph of the data should be examined to see if an additive or multiplicative seasonal effect is the more appropriate. The method may 'blow up' if the wrong model is used.

Choice of smoothing parameters α, β, γ: Evaluation of forecasts:

Forecasts are evaluated using different measures based on the difference between actual and the predicted value (the residual), which is defined as

$$e_t = X_t - \hat{X}_{t-1}(1), \quad t = 1(1)N.$$

Among these measures, the following are the most frequently used:

i. **MAPE:** Mean absolute percentage errors:

$$\text{MAPE} = \frac{\sum_{t=1}^{N} \left| \frac{e_t}{X_t} \times 100 \right|}{N}$$

This method is useful when the units of measure of X_t are relatively large.

ii. **RMSE:** Root mean squared errors:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N} e_t^2} = \sqrt{\text{MSE}}.$$

iii. **MAD:** Mean absolute deviation:

$$\text{MAD} = \frac{1}{N}\sum_{t=1}^{N} |e_t|$$

iv. MSE: Mean squared error:

$$MSE = \frac{1}{N}\sum_{t=1}^{N} e_t^2 .$$

This method is useful when the managers are interested in minimizing the occurance of a major error. This measure magnifies large errors. This method does not indicate whether the model is symmetrically underestimating or overestimating the actual values.

The parameters α, β, γ should lie in the interval $(0,1)$, and can be selected by minimizing MAD, MSE or MAPE.

• Validation: The goal of the forecast exercise is to obtain accurate predictions. In measuring the accuracy of the prediction the forecaster usually relies on the performance of the model using past information. This is basically the assumption adopted in the previous section.

A better approximation to measure how accurate a model predicts is to use only part of the sample and validate the model using the hold out sample. In this case the ratios based on forecast errors are calculated using the number of data points predicted for the hold out sample points. This provides a more reliable measure of the quality of the forecast for each model.

In order to apply Holt-Winter's smoothing to seasonal data, the analyst should carry out the following steps:

1. Examine a graph of the data to see whether an additive or a multiplicative seasonal effect is the more appropriate.
2. Provide starting values for L_1 and b_1 as well as seasonal values for the first year, say, $S_1, S_2, \ldots, S_s$, using the first few observations in the series in a fairly simple way; for e.g. the analyst could choose $L_1 = \frac{1}{s}\sum_{t=1}^{s} X_t$.
3. Estimate values for α, β, γ by minimizing $\sum e_t^2$ over a suitable fitting period for which historical data are available.
4. Choose between a fully automatic approach (for a large number of series) and a non-automatic approach. The latter allows subjective adjustments for particular series, for example, by allowing the removal of outliers and a careful selection of the appropriate form of seasonality.

5. Decide whether to normalize the seasonal indices at regular intervals by making them sum to zero in the additive case or have an average of one in the multiplicative case.

☒ The following table presents a guideline of the different forecasting methods based on different conditions :–

Forecasting Method	Data Pattern	Data Points	Forecast Horizon	Quantitative Skills
1. Moving Average	Stationary	At least the no. of periods in the MA.	Very short	Little
2. SES	Stationary	5 – 10	Short	Little
3. Holt-Winters Method	Trend & seasonality	4 – 5 per season	Short to Medium	Moderate
4. Time-series decomposition	Trend, seasonal and cyclical patterns.	Enough to see two peaks and throughs in the cycle	short, medium and long	Little.

Problem:– Examine whether the TS $X_t = (-1)^t e_t$ is stationary, where $\{e_t\}$ are mutually uncorrelated and identically distributed r.v.s with mean zero and variance σ^2.

Solution:–

— × ——— × ——— (END) ——— × ——— × ———

1. Introduction

Time series data is a collection of observations or data made sequentially in time. It has four components: Trend, Seasonality, Cyclical component & Irregular component. And Forecast is an estimate of the future value of some variable.

There are some forecasting techniques that usually used to forecast data time series with trend and seasonality, including additive and multiplicative methods. Those methods are Winter's exponential smoothing, Decomposition, Time series regression, and ARIMA models (see e.g. Bowerman and O'Connel (1993) or Hanke and Reitsch (1995)).

Many business and economic time series are non-stationary time series that contain trend and seasonal variations. The trend is the long-term component that represents the growth or decline in the time series over an extended period of time. Seasonality is a periodic and recurrent pattern caused by factors such as weather, holidays, or repeating promotions. Accurate forecasting of trend and seasonal time series is very important for effective decisions in retail, marketing, production, inventory control, personnel, and many other business sectors (Makridakis and Wheelwright, 1987). Thus, how to model and forecast trend and seasonal time series has long been a major research topic that has significant practical implications.

In this study we examine the forecasting of incoming calls to Call Center. The two different approaches used for forecasting the daily call volume include Box and Jenkins (ARIMA) methodology and Smoothing methodology. Both methods are smoothing methods. Our objective is to use past data to develop a forecasting model for the closest days to come. We will to this end use data from Call Center to

1. Develop different time series models for daily call volume.
2. Make comparison of different forecasting techniques to suggest the better one.

Our hope is that our findings will help to use better forecast model for Call Centre Data.

2. Review of Literature

In this section we summarize some different research articles concerning the method of forecasting of volume of calls to call centers.

2.1 Improving Forecasting For Telemarketing Centers by ARIMA Modeling With Intervention

The incoming calls to telemarketing centers was analyzed for the purposes of planning and budgeting by Lisa Bianchi, Jeffrey Jarrett and R. Choudary Hanumara (1998). In their publication, they used Box–Jenkins (ARIMA) modeling with intervention analysis (Intervention analysis in time series refers to the analysis of how the mean level of a series changes after an intervention, when it is assumed that the same ARIMA structure for the series holds both before and after the intervention) and additive and multiplicative versions of Holt–Winters (HW) exponentially weighted moving average models. With aid of these models they forecasted the daily call volumes. The data used for analysis was from March 1, 1991 to June 26, 1991.

Their first model was the ARIMA(p,d,q). Their second model was the multiplicative Holt-Winter model

$$Y(t) = (a(t) + b)\, s(t) + e(t)$$

When seasonal variation is constant over time, an additive seasonal factor model is appropriate. Hence the third additive model used was

$$Y(t) = (a(t) + b) + s(t) + e(t)$$

The Root Mean Square Error *(*RMSE) is used to compare different model forecasts performance. It was found that ARIMA models with intervention analysis provided better forecasts for planning and control.

A complete version of this study can be found from paper [1].

2.2 Wireless Traffic Modeling and Prediction

In this article Yantai Shu, Minfang Yu, and Jiakun Liu (2003) studied wireless traffic. In their study to predict traffic, seasonal ARIMA model with two periodicities was used. The hourly traffic data from 0:00 June 1 2001 (Friday) to 0:00 April 27 2002(Saturday) was measured. A total of 330 days from the dial-up access network of China net-Tianjin. To trace the daily traffic the model ARIMA(1,0,1) and ARIMA(1,1,0) were found and for the hourly traffic ARIMA(0,1,1). For estimating the model the first 300 daily data was used. The last 30 days to evaluate the model. An adjusted traffic prediction method is proposed using seasonal ARIMA

Models. The comparison is repeated with many prediction experiments on the actual measured GSM traces of China Mobile of Tianjin.

It is founded that the relative error between the actual values and forecasting values are all less than 0.02.Their study showed that the seasonal ARIMA model is a good traffic model capable of capturing the properties of real traffic.

A complete version of this study can be found from paper [2].

2.3 The application of forecasting to modeling emergency medical system calls

The emergency medical system calls of major Canadian city Alberta was analyzed by Nabil Channouf, Pierre L Ecuyer (2006). In their analysis two different methods was used, autoregressive model of data obtained after eliminating the trend, seasonality, special day effect and a double-seasonal ARIMA model with special day effect. Then the comparison of the both models is presented. For the purpose of analysis of the data for emergency medical calls was obtained from January 1, 2000 to March 16, 2004 including call priority, and the geographical zone where the call originated. The modeling is done on the first 1096 observations and the remaining 411 observation is used for evaluation.

The model found was an ARIMA decomposed model with two seasonal cycles.

$$Y_t = N_t + w_1 H_{t,1} + w_2 H_{t,2}$$

The ARIMA model with two seasonal cycles is suggested. They found that this model performed poorly when forecasting more than two weeks into the future.

A complete version of this study can be found from paper [3].

2.4 Forecasting Police Calls during Peak Times for the City of Cleveland USA

The police service calls during peak times for the city Cleveland, US was presented by the police department of the city. Professor John P.Holcomb Jr (2007) used autoregressive integrated moving average (ARIMA) modeling technique, Multiple Regression and different smoothing methods to analyze data. As a first step the data of call volume (per hour) is obtained and it was divided into 10 important categories. This provided 24,000 data points across all kinds of calls, further the calls are divided priority wise, priority 1 calls being the most important. Priority 1 calls are the calls where crime is in progress: such as robbery or domestic violence. The researcher used different methodologies for building models. For model evaluation, the mean absolute percent error (MAPE) is used.

He suggested that multiple regression approach have difficulty. The final ARIMA(1,0,0) and ARIMA(5,1,0) model is used. This model produced an improved MAPE over the Holt-Winters method approximately 12%. A complete version of this study can be found from paper [4].

2.5 Predicting call arrivals in call centre

The daily call volume of car damage insurance claims at Vrije University, Netherlands was analyzed by Koen Van Den Bergh (2006). In this publication he discussed four different methods:,ARIMA modeling, Dynamic Regression Modeling, Exponential smoothing and modeling by Regression. These four techniques are applied to the daily call center data to forecast the daily call volume. The models used for forecasting are given below.

The ARIMA model is

$$y_t = c + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

The dynamic regression model is

$$y_t = \alpha + v_0 x_t + v_1 x_{t-1} + \cdots + v_k x_{t-k} + \mu_t$$

The single exponential smoothing model is

$$\widehat{y_t} = w_0 y_{t-1} + w_1 y_{t-2} + w_2 y_{t-3} + \cdots$$

The regressions model is

$$Y_t = S_t + T_t + R_t + \sum_{i=1}^{n} b_i X_i$$

It is presented that all of this methodology can at least deal with Randomness. Single Exponential smoothing is not good enough to deal with seasonality and trend pattern but this methodology can handle the random part which is a least result that a forecasting technique can give. The Dynamic Regression model and Regression model can deal with the interventions as well, where the ARIMA models can't deal with intervention.

A complete version of this study can be found from paper [5]

3. Methodology

3.1 Necessity of Forecasting

Uncertainty means that no clarity about future may be achieved when uncertain decisions are made upon historical experiences. Historical data can be smoothed in different ways. But the scientific approach is essential to make decision. The forecasting is one of the major scientific approaches that help in process of making decision in condition of uncertainty. Forecasting is based on the assumption that the past patterns and behavior of a variable will continue into the future. The objective is to use past data to develop a forecasting model for the future periods. To reach our goal of forecasting daily call volume of call center, the sophisticated forecasting techniques known as ARIMA (Auto Regressive Integrated Moving Average) and Smoothing Methodology are applied.

3.2 Assumptions of Time Series Analysis

A major assumption in time series analysis is the stationarity of the series, this means that the average value and the variation of the series should be constant with respect to time. If the series is not stationary then we make it stationary by the different transformations the most commonly used transformations are log and first difference.

3.2.1 Stationarity Tests

There are different tests for checking the stationarity of the data, two important tests are: Augmented Dickey- Fuller test (ADF Test) and Kwiatkowski Philips Schmidt Shin Test (KPSS Test).

(i) Unit Root Test (ADF Test):-

ADF Test checks whether any specific pattern exists in the data. Here small p-value suggests that the data is stationary. The unit root presence can be illustrated as follows by using a first order autoregressive process: $y_t = \mu + \rho y_{t-1} + \epsilon_t$ --------- (1)
where, $\epsilon_t \sim N(0, {\sigma_\epsilon}^2)$

The basic Dickey- Fuller test examines whether $\rho < 1$
After subtracting y_{t-1} from both sides in equation above,

$$\Delta y_t = \mu + (\rho - 1) y_{t-1} + \epsilon_t$$

$$\Delta y_t = \mu + \theta y_{t-1} + \epsilon_t \text{ ----------- (2)}$$

H_0: θ = 0 (there is a unit root in y_t)

H_1 : θ < 0

Equation (1) and (2) are the simplest case where the residual is white noise. In general, there is serial correlation in the residuals and Δy_t can be represented as an autoregressive process:

$$\Delta y_t = \mu + \theta y_{t-1} + \sum_{i=1}^{p} \emptyset_i \, \Delta y_{t-i} \quad + \epsilon_t ----(3)$$

Corresponding to equation (3), Dickey-Fuller procedure becomes the Augmented Dickey-Fuller test. We can also include a deterministic trend in equation (2). Altogether, there are four test specification with regard to the combination of an intercept and a deterministic trend. [6]

(ii) KPSS Test:-

This is another test for stationary which check especially the existence of trend in the data set.

H_0: data is stationary

H_1 : data is not stationary

Larger p-value suggests data is stationary.

3.2.2 Differencing Method

A method for making series stationary. A differenced series is the series of difference between each observation Y_t and the previous observation Y_{t-1}

$$Yt' = Yt – Y_{t-1}$$

A series with trend can be made stationary with 1st differencing
A series with seasonality can be made stationary with seasonal differencing

3.2.3 White Noise & Lag

It describes the assumption that each element in a series is a random draw from a population with mean zero and constant variance.

Lag shift a series down by a specific number of rows in the worksheet.

3.3 Box-Jenkins modelling

The methodology introduced 1970 by Box and Jenkins assumes that the data is dependent on itself. And the very first thing to decide on is the number of lags. Then a number of parameters are estimated, the residuals are checked and finally a forecast is made.

The general ARIMA(p,q,d) model looks like.

$$y_t = c + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

Where

c: constant

$\phi_1, \phi_2, \theta_1, \theta_2$, - - - are model parameters

$e_{t-1} = y_{t-1} - s_{t-1}$, e_t are called errors or residuals

s_{t-1} : predicted value for the $(t-1)^{th}$ observation (y_{t-1})

p: number of auto regressive (AR) terms

q: number of moving average (MA) terms

d: level of differencing

3. 3.1 Auto-Regressive (AR) Model

In the pure AR (p) autoregressive with p lags model, we have

$$Y_t = U_t + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_{1p} Y_{t-p} + \epsilon_t$$

that is the series depend on itself up to p lags. The simplest and most widely used model with serial correlation is the first order autoregressive model of first order. The AR (1) model is specified by: $Y_t = U_t + \emptyset_1 Y_{t-1} + \epsilon_t$

where, $\emptyset_1, \emptyset_2, \dots, \emptyset_p$are the parameters of the model, μ_t is constant with respect to t and ϵ_t is white noise. Many authors omit the constant term.

3. 3.2 Moving Average (MA) model

The moving average model models the error terms, which are not observed. The moving average model is defined as:

$$Y_t = U_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

Where $\theta_1, \theta_2, \dots, \theta_q$ are the parameters of the model, μ_t is a constant with respect to t and ϵ_t is white noise. Many authors omit the constant term.

This model is useful when time series doesn’t exhibit a trend or a seasonal pattern.

3. 3.3 Auto-Regressive Integrated Moving Average (ARIMA) model

ARIMA (p, d, q) (P, D, Q) where (p is the order of AR process, q is the order of MA process and d is the order of differencing) is a regular model and (P, D, Q) are seasonal elements. The ARIMA models are generalization of the simple AR model that uses three tools for modeling series correlation in the disturbance. The first tool is the auto-regressive terms. The second tool is the integrated (difference) terms. A first order integrated component means that the forecasting model is designed for the first difference of the original series. A second order component difference of the original series and so on. The third tool is the moving average terms. A moving average forecasting model uses lagged values of the forecasted errors. A first order moving average term uses the forecasted errors from the two most recent periods, and so on [7].

3. 4 Exponential Smoothing

There are several exponential smoothing methods. The majors which we use, are Single Exponential Smoothing, Holt‟s Linear Model (1957) and Holt-Winters Trend and Seasonality Model.

3.4.1 Single Exponential Smoothing

The simplest form of exponential smoothing is single exponential smoothing, which may be used when data is without any systematic trend or seasonal components. Given such a time series, a logical approach is to take a weighted average of past values. So for a series $y_1, y_2, \dots, y_{t-1}$, the estimate of the value of Y_t, given the information available up to time t, is

$$\widehat{Y}_t = w_0 Y_{t-1} + w_1 Y_{t-2} + w_2 Y_{t-3} + \cdots$$

Where $w_i = \alpha (1 - \alpha)^i$ are the weights given to the past values of the series and they sum to 1.

Here the "α" lies between 0 and 1. Since the most recent observations of the series are also the most relevant, it is logical that these forecasting observations should be given more weight than the observations further in the past. This is done by giving declining weights to the series. These decrease by a constant ratio.

Single Exponential Smoothing gives more weight to recent values compared to the old values. More efficient for stationary data without any seasonality and trend.

3.4.2 Holt's Linear Model

Holt"s linear model is an extension of single exponential smoothing. This method allowed forecasting data with trends.

For a time series y_1, y_2, ... The estimate of the value of y_{t-k}, is given by the formula:

$\widehat{y_{t+k}} = m_t + b_t k$ where k = 1,2,3…

Where m_t denotes an estimate of the level of the series at time t and b_t denotes an estimate of the slope of the series at time t.

Where $m_t = \alpha_0 y_t + (1 - \alpha_0)(m_{t+1} + b_{t-1})$

$b_t = \alpha_1(m_t + m_{t-1}) + (1 - \alpha_1)b_{t-1}$ with $0 < \alpha_0 < 1$ and $0 < \alpha_1 < 1$

Holt's Linear Model is useful when we smooth the series that gives weights to older observations and provide short-term forecasts. Useful when the series exhibits a seasonal pattern, with or without a trend.

The following table presents a guideline of the different forecasting methods based on different conditions:

Forecasting Method	Data Pattern	Data Points	Forecast Horizon	Quantitative Skills
Moving Average	Stationary	At least the number of periods in MA	Very Short	Little
Single Exponential Smoothing	Stationary	5-10	Short	Little
Holt-Winter Method	Trend & Seasonality	4-5 per season	Short to Medium	Moderate
ARIMA Methodology	Stationary (Differencing/ Transformation)	4-5 per season	Medium	High

3.5 Model Selection Criteria

Here we discuss the few criteria we used in the study when selecting the best model among the competing models. Several criteria can be used for this purpose, here we discuss Akaike information criterion (AIC) and the Bayesian information criterion (BIC) or Schwarz information criterion (SIC). These criterions are used for measuring the goodness of fit of the model. These criterion are minimized over the choice of repressors, it will be minimum when the model is good fit and less complex. In comparing two or more models, the best model is the one having the least AIC and BIC values.

In a regression setting, the estimates of the β_I based on least squares and the maximum likelihood estimates are identical. The difference comes from estimating the common variance σ^2 of the normal distribution for the errors around the true means. We have been using the best unbiased estimator of σ^2, $\hat{\sigma}^2 = RSS/(n-p)$, where there are p parameters for the means (p different β_I parameters) and RSS is the residual sum of squares. This estimate does not tend to be too large or too small on average. The maximum likelihood estimate, on the other hand, is RSS/n. This estimate has a slight negative bias, but also has a smaller variance. Putting all of this together, we can write -2 times the log-likelihood to be

$$n + n\log(2\pi) + n\log(RSS/n).$$

In a regression setting. Now, AIC is defined to be -2 times the log-likelihood plus 2 times the number of parameters. If there are p different β_I parameters, there are a total of p+1 parameters if we also count σ^2. The correct formula for the AIC for a model with parameters $\beta_0, \beta_1, \dots, \beta_{p-1}$ and $\sigma^2$is

$$AIC = n + nlog2\pi + nlog\left(\frac{RSS}{n}\right) + 2(p+1)$$

and the correct formula for BIC is

$$BIC = n + nlog2\pi + nlog\left(\frac{RSS}{n}\right) + (logn)(p+1)$$

3.6 Measurements of Forecasting Accuracy

Before the forecasting results can be given, some measurements of forecasting accuracy must be determined. This section captures the equations of the most widely applied measurement methods. The following list of methods shall be utilized for assessing the accuracy of forecasts

3.6.1 Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{e_t}{y_t}\right| \times 100$$

3.6.2 Mean Square Error (MSE)

$$MSE = \sum_{i=1}^{n}\frac{{e_t}^2}{n}$$

3.6.3 Root Mean Square Error (RME)

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{{e_t}^2}{n}}$$

3.6.4 Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_t|$$

3.7 Residual Analysis

Residuals are the difference between the predicted output from the model and the original values (data). Residuals basically represent the portion of the data not explained by the model. Residual analysis may be regarded to consist of two tests: Whiteness Test and Normality Test.

3.7.1 Normality Test

A good model is the one for which the residuals fulfil the assumption of normality. The histogram of the residuals gives a good idea about the normality. The Normal probability graph is also used to assess that the data set is approximately normally distributed. In a normal probability graph the data is plotted against the theoretical normal distribution in such a way that it makes a straight line. If the points depart from straight line, we have a departure from normality. The Anderson Darling test is one of the three generally known tests for the normality. It is the modified form of Kolmogorov-Smirnov test and gives more weight to the tails as compared to the Kolmogorov-Smirnov test. In the Kolmogorov-Smirnov test the critical values do not depend on the specific distribution being tested but the Anderson Darling test use the specific distribution for calculating the critical value. The test statistic of the test is given below:

$$A^2 = -N - S$$

$$S = \sum_{i} \frac{(2i-1)}{N} [logF(Y_i) + \log\{1 - F(Y_{N=1-i})\}]$$

Where F is the cumulative distribution function of interest.

3.7.2 Whiteness Test

The purpose of this test is to analyze the correlation between the residuals at different lags. According to the whiteness test criteria all autocorrelation should be zero.

3.7.3 Ljung–Box test

This is an objective way to test the null hypothesis that there is no autocorrelation. The Q-statistic at lag k is a test statistic for the null hypothesis that there is no autocorrelation up to order k. It is computed as

$$Q = T(T+2) \sum_{j=1}^{k} \frac{{r_j}^2}{T-j}$$

Where r_j is the j-th autocorrelation and T is the number of observations. k is the number of lags being tested.