Auto-Regression Series :- (General Series)

By studying the natural of oscillatory component in a time-series, we observe that the value of the series at any time 't' may depend upon its ~~on~~ own value at time $t-1, t-2, t-3, \ldots, t-k$. A time-series represented by the recurrence relation

$$y_t = f(y_{t-1}, y_{t-2} \ldots y_{t-k}) + \epsilon_t$$

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_k y_{t-k} + \epsilon_t \quad \text{---(1)}$$

where, f is a linear f^n and ϵ_i be the ~~dist~~ disturbance f^n such that $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. Then the series (1) is called the auto regression series of order k. Under certain conditions to be satisfied by $a_1, a_2, \ldots, a_k$. The generating series in (1) can be solve to represent the oscillatory time-series.

First Order Auto regression Series (Markoff Series) :-

The first order Auto regression series can be written as

$$y_{t+1} = a y_t + b + \epsilon_{t+1}, \quad |a| < 1, \quad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2). \quad \text{---(2)}$$

Here, it is assume that y_t's are measured from their means i.e. $E(y_t) = 0 = E(\epsilon_i)$ taking the expectation of (2), we get $b = 0$. $\therefore y_{t+1} = a y_t + \epsilon_{t+1}$ ---(3)

The eq^n (3) is a linear difference eq^n of order 1st.

To solve this series we need of complementary func^n (CF) and particular ~~func~~^{n} Integral (PI).

Let $y_t = \lambda^t$ be the trial solu^n of homogeneous eq^n of eq^n (3).

$$\lambda^{t+1} - a\lambda^t = 0$$

$$\lambda^t(\lambda - a) = 0$$

$$\Rightarrow \lambda = a, \; \lambda^t \neq 0$$

$\therefore$ The complementary f^n CF will be

$$CF = a^t$$

for particular Integral, from eq^n (3)

$$(E-a)\,y_t = \epsilon_{t+1}$$

$$\Rightarrow \boxed{(E-a)y_t = y_{t+1} - a\,y_t} \rightarrow \text{formula}$$

$$y_t = \frac{1}{(E-a)}\,\epsilon_{t+1}$$

$$= \frac{1}{E}\left(1-\frac{a}{E}\right)^{-1}\epsilon_{t+1}$$

$$= \left(1-\frac{a}{E}\right)^{-1}\epsilon_t$$

$$= \left(1+\frac{a}{E}+\frac{a^2}{E^2}+\cdots\right)\epsilon_t$$

$$= \epsilon_t + a\,\epsilon_{t-1} + a^2\epsilon_{t-2} + \cdots$$

$$= \sum_{j=0}^{\infty} a^j \epsilon_{t-j}$$

$$= \sum_{j=1}^{\infty} a^{t-j} e_j$$

Scanned with CamScanner

$PI = \sum_{j=1}^{t} a^{t-j} \epsilon_j$ (if $\epsilon_t = 0$, $t < 0$)

Hence, the general soln of linear diff. eqn (2) is given by $y_t = CF + PI = a^t + \sum_{j=1}^{t} a^{t-j} \epsilon_j$

If $|a| < 1$ and the series is very large or has be made so by assuming that then the series has started some times prior to the point $t = 0$, then we have $|a|^t \to 0$ and the effective soln becomes $\sum_{j=1}^{t} a^{t-j} \epsilon_j$. This is the moving average of random elements with weights $1, a, a^2, \ldots a^{t-1}$!

The Second Order Auto Regression Series (Yule's Series) :- The IInd order auto regressive series is given by

$$y_{t+2} + a y_{t+1} + b y_t = \epsilon_{t+2} \quad \text{—— (1)}$$

It is linear diff. eqn of order (2). for solving this we need CF and PI.

The CF is given for

$$y_{t+2} + a y_{t+1} + b y_t = 0 \quad \text{—— (2)}$$

Let the trial soln be,

$$y_t = \lambda^t$$

on putting eqn (2)

$$\lambda^{t+2} + a\lambda^{t+1} + b\lambda^t = 0$$

$$(\lambda^2 + a\lambda + b)\lambda^t = 0$$

$\Rightarrow \lambda^2 + a\lambda + B = 0 \; ; \; \lambda^t \neq 0$

$\Rightarrow \lambda_1 = \frac{-a + \sqrt{a^2 - 4B}}{2}$,

$\lambda_2 = \frac{-a - \sqrt{a^2 - 4B}}{2}$

$\therefore$ The complementary f^n (CF) is obtained as

$CF = A_1 \lambda_1^t + A_2 \lambda_2^t$

where, A_1 and A_2 are some arbitrary constant. Since eq^n (1) represents an oscillatory series, $a^2 - 4B < 0$.

Thus, λ_1 & λ_2 are complex and y_t can be expressed as sin and cosin f^n of t.

Let $\lambda_1 = -\frac{a}{2} + \frac{i}{2}\sqrt{4b - a^2} = p(\cos\theta + i\sin\theta)$

where, $p\cos\theta = -\frac{a}{2}$, $p\sin\theta = \frac{1}{2}\sqrt{4B - a^2}$

& $\lambda_2 = p(\cos\theta - i\sin\theta)$

Now, the CF becomes

$$y_t = A_1 p^t (\cos\theta t + i\sin\theta t) + A_2 p^t (\cos\theta t - i\sin\theta t)$$

$$= p^t [(A_1 + A_2)\cos\theta t + i(A_1 - A_2)\sin\theta t]$$

$$= p^t [A\cos\theta t + B\sin\theta t]$$

where, $A = A_1 + A_2$ and $B = i(A_1 - A_2)$

Since 't' can take very large values so it is necessary to assume that

$$|p| = \sqrt{B} < 1$$

Scanned with CamScanner

Since if $|p|>1$ then as $t \to \infty$, $y_t \to \infty$

Particular Integral :- For PI of eqn (1), we have

$$(E^2 + aE + b) y_t = \epsilon_{t+2}$$

$$y_t = \frac{1}{(E-\lambda_1)(E-\lambda_2)} \epsilon_{t+2}$$

$$= \frac{1}{E^2\left[1-\frac{\lambda_1}{E}\right]\left[1-\frac{\lambda_2}{E}\right]} \epsilon_{t+2}$$

$$= \left(1-\frac{\lambda_1}{E}\right)^{-1}\left(1-\frac{\lambda_2}{E}\right)^{-1} \epsilon_t$$

$$= \left[1+\frac{\lambda_1}{E}+\frac{\lambda_1^2}{E^2}+\cdots\right]\left[1+\frac{\lambda_2}{E}+\frac{\lambda_2^2}{E^2}+\cdots\right] \epsilon_t$$

$$= \left[1+\left(\frac{\lambda_1+\lambda_2}{E}\right)+\frac{\lambda_1^2+\lambda_1\lambda_2+\lambda_2^2}{E^2}+\cdots\right]\epsilon_t$$

$$= \left[\frac{\lambda_1-\lambda_2}{\lambda_1-\lambda_2}+\frac{\lambda_1^2-\lambda_2^2}{\lambda_1-\lambda_2}\cdot\frac{1}{E}+\frac{\lambda_1^3-\lambda_2^3}{\lambda_1-\lambda_2}\frac{1}{E^2}+\cdots\right]\epsilon_t$$

$$= \sum_{j=1}^{\infty} \frac{\lambda_1^j-\lambda_2^j}{\lambda_1-\lambda_2} \epsilon_{t-j+1}$$

Now,

$$\frac{\lambda_1^j-\lambda_2^j}{\lambda_1-\lambda_2} = \frac{p^j(\cos j\theta + i\sin j\theta) - p^j(\cos j\theta - i\sin j\theta)}{p(\cos\theta + i\sin\theta) - p(\cos\theta - i\sin\theta)}$$

$$= \frac{p^j \cdot 2i\sin j\theta}{p\, i\, 2\sin\theta}$$

putting the value of $p\sin\theta$, we get

$$= \frac{2p^j \sin j\theta}{\sqrt{4b-a^2}} = \frac{2}{\sqrt{4b-a^2}} p^j \sin j\theta$$

$$= \varepsilon_{ij} \text{ (let)}$$

$$\therefore PI = \sum_{j=1}^{\infty} \varepsilon_{ij} \, \epsilon_{t-j+1}$$

Hence, the complete solⁿ of linear diff. eqⁿ (1) as

$$y_t = CF + PI$$

$$= p^t (A\cos\theta t + B\sin\theta t) + \sum_{j=1}^{\infty} \varepsilon_{ij} \, \epsilon_{t-j+1}$$

If the series is very large $p^t \to 0$ as $|p| < 1$ and the resulting series will be

$$y_t = \sum_{j=1}^{\infty} \varepsilon_{ij} \, \epsilon_{t-j+1}$$

The above series is generated by moving avg. of random element with harmonic weight ε_j's.

Auto Correlation and correlogram :- Suppose from the original series y_t ($t = 1, 2, \ldots n$), we obtain $(n-k)$ pair of obs^n.s (y_t, y_{t+k}), ($t = 1, 2, \ldots n-k$) with a long period k the ordinary product moment corr. b/w the two series y_t and y_{t+k} is called auto corr or serial corr of order k. Thus,

$$\therefore r_k = \frac{E[y_t, y_{t+k}]}{\sqrt{Var(y_t)Var(y_{t+k})}} \quad t = 1, 2, \ldots n-k$$

Obviously we have, $r_0 = 1$ and $r_{-k} = r_k$.

Correlogram of Moving Average :- The correlogram is a diagram b/w r_k (corr. b/w y_t, y_{t+k}) and period k. Correlogram provides a criterion for explaining the structure of a time-series. For a moving avg of extent m with weights $a_1, a_2, \ldots a_m$ of random component ϵ_i's where, $i = 1, 2, \ldots$ The generated series can be written as

$$y_i = a_1 \epsilon_{i+1} + a_2 \epsilon_{i+2} + \cdots + a_{k+1} \epsilon_{i+k+1} + \cdots + a_m \epsilon_{i+m}$$

$$y_{i+k} = a_1 \epsilon_{i+k+1} + a_2 \epsilon_{i+k+2} + \cdots + a_{m-k} \epsilon_{i+m-k} + \cdots + a_m \epsilon_{i+k+m}$$

where, ϵ_i are $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. Thus,

$$E(y_i) = 0 = E(y_{i+k})$$

$$\text{cov}(\epsilon_i \epsilon_j) = 0 \quad \forall\, i \neq j$$

$$\text{Var}(y_i) = E(y_i^2) - [E(y_i)]^2$$

$$= (a_1^2 + a_2^2 + \cdots + a_m^2)\sigma^2 - 0$$

$$= \sigma^2 \sum_{j=1}^{m} a_j^2 \quad \forall\, i = 1, 2, \ldots$$

$$\text{Var}(y_{i+k}) = E(y_{i+k}^2) = \sigma^2 \sum_{j=1}^{m} a_j^2$$

$$E[y_i, y_{i+k}] = (a_1 a_{k+1} + a_2 a_{k+2} + \cdots + a_{m-k} a_m)\sigma^2$$

$$= \sigma^2 \sum_{j=1}^{m-k} a_j a_{j+k} \quad \text{where, } k < m$$

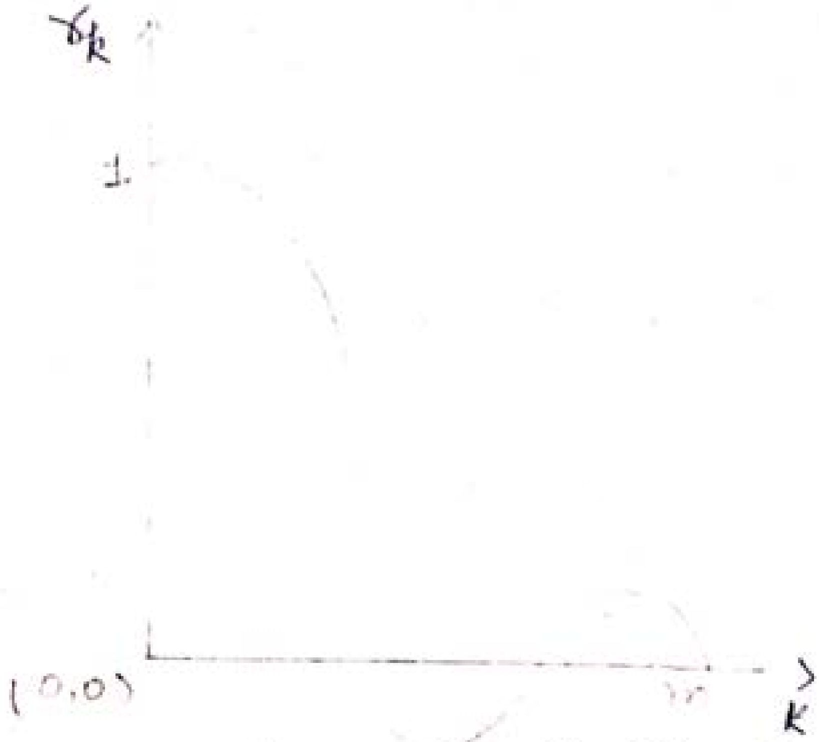
$$\therefore \; r_k = \frac{E(y_i, y_{i+k})}{\sqrt{\text{Var}(y_i)\text{Var}(y_{i+k})}}$$

$$\boxed{= \frac{\sum_{j=1}^{m-k} a_j a_{j+k}}{\sum_{j=1}^{m} a_j^2}} \; ; \quad k < m$$

and $r_k = 0$ if $k \geq m$.

Thus, correlogram will oscillate b/w the points ~~0 and 1~~ $(0, 1)$ and $(m, 0)$ and then coincide with k-axis when $k > m$.

as shown in the diagram.



In particular case if $a_i = \frac{1}{m}$, $i = 1, 2, \dots m$

Then, $\gamma_k = \dfrac{(m-k)\cdot\frac{1}{m^2}}{m\cdot\frac{1}{m^2}} = 1 - \frac{k}{m}$; $k \le m$

$= 0$ if $k > m$

$\gamma_k + \frac{k}{m} = 1$ which is the eq^n of straight line in intersect form $\left[\frac{x}{a} + \frac{y}{b} = 1\right]$

Thus, for a series generated by an m point simple moving avg of random component the correlogram consist of a straight line joining the points $(m, 0)$ and $(0, 1)$. Together with k-axis from the points $(m, 0)$.

Correlogram of Harmonic Series :- Let we have a harmonic series

$y_t = a \sin \theta t + \epsilon_t$, where 'a' is the amplitude of the time-series of 'sin' terms and the period of oscillation is $2\pi/\theta$, and ϵ_i be the random component independant of $a \sin \theta t$, and ϵ's are iid normal with mean '0' and var σ^2 i.e.

$\epsilon \overset{iid}{\sim} N(0, \sigma^2)$. $\therefore$ $\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0$

Now, $\mathrm{Cov}(y_t, y_{t+k}) = E[y_t, y_{t+k}]$

$\because$ $E(y_t) = 0 = E(y_{t+k})$

Now,

$$\mathrm{Cov}(y_t, y_{t+k}) = a^2 E[\sin \theta t, \sin \theta (t+k)]$$

$$= \frac{a^2}{2} E[2 \sin \theta t \cdot \sin \theta (t+k)]$$

$$= \frac{a^2}{2} E[\cos \theta k - \cos \theta (2t+k)]$$

$$= \frac{a^2}{2} \frac{1}{n} \sum_{i=1}^{n} [\cos \theta k - \cos \theta (2t+k)]$$

$$= \frac{a^2}{2n} \sum_{i=1}^{n} [\cos \theta k - \cos \theta (2t+k)]$$

Let, $\sum_{i=1}^{n} \cos (2t+k)\theta = \frac{\cos \theta (n+k+1) \sin n\theta}{\sin \theta}$

Now,

$$\mathrm{Cov}(y_t, y_{t+k}) = \frac{a^2}{2n} \sum_{i=1}^{n} [\cos \theta k] - \frac{a^2}{2n} \frac{\cos \theta (n+k+1) \sin n\theta}{\sin \theta}$$

$$= \frac{a^2}{2\cancel{n}} \cdot \cancel{n} \cos\theta k - \frac{a^2}{2n} \frac{\cos\theta(n+k+1)\sin n\theta}{\sin\theta}$$

Since $\lim n \to \infty$

$$\therefore \lim_{n\to\infty} \frac{a^2}{2n} \frac{\cos\theta(n+k+1)\sin n\theta}{\sin\theta} \to 0$$

$$\therefore \quad \operatorname{Cov}(y_t, y_{t+k}) = \frac{a^2}{2} \cos\theta k$$

Now,

$$\operatorname{Var}(y_t) = E[y_t^2]$$

$$= E\left[\left[a\sin\theta t + \epsilon_t\right]^2\right]$$

$$= \frac{a^2}{n} \sum_{t=1}^{n} \sin^2\theta t + \sigma^2$$

$$= \frac{a^2}{n} \sum_{t=1}^{n} \left(\frac{1 - \cos 2\theta t}{2}\right) + \sigma^2$$

$$= \frac{a^2}{2} - \frac{a^2}{n} \sum_{t=1}^{n} \frac{\cos 2\theta t}{2} + \sigma^2$$

Let $\frac{\cos 2\theta t}{2}$

$: \frac{\cos\theta(n+1)\sin n\theta}{\sin\theta}$

$$= \frac{a^2}{2} + \sigma^2 - \frac{a^2}{n} \frac{\cos\theta(n+1)\sin n\theta}{\sin\theta}$$

For large n i.e. $\lim n \to \infty$

$$\Rightarrow \lim_{n\to\infty} \frac{a^2}{n} \frac{\cos\theta(n+1) \cdot \sin n\theta}{\sin\theta} \to 0$$

$$\therefore \quad \operatorname{Var}(y_t) = \frac{a^2}{2} + \sigma^2$$

Hence, $$r_k = \frac{\operatorname{Cov}(y_t, y_{t+k})}{\sqrt{\operatorname{Var}(y_t)\operatorname{Var}(y_{t+k})}}$$

$$\gamma_K = \frac{\frac{a^2}{2}\cos\theta K}{\frac{a^2}{2}+\sigma^2} = C_1 \cos\theta K \text{ (let)}$$

where $C_1 = \dfrac{\frac{a^2}{2}}{\frac{a^2}{2}+\sigma^2}$

Since, γ_K is ind. of t ∴ The correlogram of a harmonic series also a cyclic curve with the same period of oscillation as that original series. But changed amplitude C_1 which is ind of K.

Hence, the correlogram of a harmonic series of harmonic terms will oscillates but not tend to 'zero.'

Correlogram of Auto-Regressive Series :-

(i) Correlogram of first Order Auto-Regressive Series :- The first order auto-regressive series can be written as

$$y_{t+1} = a y_t + \epsilon_t \;;\; |a|<1 \quad \text{——(1)}$$

To get the correlogram of the time-series we multiplied both sides by y_{t-K} and take expectation and divided by $V(y_t)$.

$$y_{t+1} \cdot y_{t-K} = a y_t y_{t-K} + \epsilon_t y_{t-K}$$

taking Expectation we get

$$\frac{E[y_{t+1} y_{t-K}]}{Var(y_t)} = \frac{aE[y_t y_{t-K}]}{Var(y_t)} + \frac{\epsilon_t E(y_{t-K})}{Var(y_t)}$$

$\Rightarrow \quad r_{k+1} = a r_k$

$\Rightarrow \quad r_{k+1} - a r_k = 0 \quad \text{—— (2)}$

The eq^n (2) is the linear difference eq^n and soln^n of this eq^n is given by

$$r_k = a^k \quad \text{—— (3)}$$

Correlogram of Second Order Regressive Series :-

The second order auto-regressive series can be written as

$$y_{t+2} + a y_{t+1} + b y_t = \epsilon_{t+2} \; ; \; |b| < 1 \quad \text{—— (4)}$$

multiply both sides of eq^n (4) by y_{t-k}, and devided by $Var(y_t)$, taking expectation we get

$$\frac{E[y_{t-k} \cdot y_{t+2}]}{Var(y_t)} + a\frac{E[y_{t-k} \cdot y_{t+1}]}{Var(y_t)} + b\frac{E[y_{t-k} \, y_t]}{Var(y_t)}$$

$$= \frac{\epsilon_{t+2} E(y_{t-k})}{Var(y_t)}$$

$\Rightarrow r_{k+2} + a r_{k+1} + b r_k = 0$

which is homogeneous differential eq^n of order (2) in r_k, where

$$r_k = p^k \left[A\cos\theta k + B\sin\theta k \right] \quad \text{—— (5)}$$

where, $\cos\theta = -\frac{a}{2p}$, $p = \sqrt{b}$,

$a^2 - 4b < 0$

The arbitrary constant involved i.e. A and B can be obtained by using condition $r_0 = 1$ and $r_K = +r_{-K}$.

Using the above condition, we get

$$A = 1, \quad B = \frac{1-p^2}{1+p^2} \cot\theta$$

putting these values of A and B in eqn (5),
we get

$$r_K = p^K\left[\cos\theta K + \frac{1-p^2}{1+p^2}\cot\theta \sin\theta K\right]$$

$$= p^K\left[\cos\theta K + \cot\phi \sin\theta K\right]$$

$$= p^K\left[\cos\theta K + \frac{\cos\phi}{\sin\phi} \cdot \sin\theta K\right]$$

Hence,

$$\boxed{r_K = p^K\left[\frac{\sin(K\theta + \phi)}{\sin\phi}\right]} \; ; \; |p| < 1$$

Here, p^K is called damping factor

Thus the correlogram of any auto-regressive series will oscillate and will be damp due to damping factor p^K.

<u>Note</u> :- Suppose we have a series y_t generated by a moving avg with weight $a_1, a_2, \ldots a_m$, weights of the random element ϵ_t with $E(\epsilon_t) = 0$ and $Var(\epsilon_t) = \sigma^2$.

Then, Auto correlation or Serial Correlation of order s b/w y_t and y_{t+s} is given by,

$$\gamma_s = \frac{Cov[y_t, y_{t+s}]}{\sqrt{Var(y_t) \cdot Var(y_{t+s})}}$$

$$= \frac{\sigma^2 \sum_{j=1}^{m-s} a_j a_{j+s}}{\sigma^2 \sum_{j=1}^{m} a_j^2} \quad , j < m$$

$$= 0 \quad ; j > m$$

This shows that the series generated by the effect of moving avg will not be uncorrelated as the period of moving avg $m > s$ b/w y_t, y_{t+s}.

Hence, moving avg when operated on the random element will provide us a much smoother series as compared to original series y_t.

Random Component of a Time-Series :-

As the def. suggest no formula or mathematical func^n can be suggested to represent the random component of a time-series, in view of this we first estimate the components of time-series and the residual is used as a measure of random variation which are left unaccounted by the known random components. The variate difference method unable us to estimate the var. of the random component in a time-series.

Variate Difference Method :- Suppose that the time-series can be represented as the sum of functional part (trend) and random component as given below.

$$y_t = a_0 + a_1 t + a_2 t^2 + \cdots + a_{k-1} t^{k-1} + \epsilon_t \quad \text{---(1)}$$

where, ϵ's are iid $N(0, \nu^2)$ and covariance b/w ϵ_i and ϵ_j is equal to zero. i.e. $cov(\epsilon_i, \epsilon_j) = 0$

$E(\epsilon_i) = 0$, $V(\epsilon_i) = \nu$, we know that for a polynomial y_t of m^{th} degree in t, then

$$\Delta^r y_t = 0$$

where, Δy_t is such that

$$\Delta y_t = y_{t+h} - y_t$$

$\therefore$ from eq^n (1), we have

$$\Delta^k y_t = 0 + \Delta^k \epsilon_t$$

Now, $\Delta^k y_t = (E-1)^k \epsilon_t \quad (\because E = 1+\Delta)$

$$= \left[E^k - \binom{k}{1} E^{k-1} + \binom{k}{2} E^{k-2} + \cdots + (-1)^k\right] \epsilon_t$$

$$= \epsilon_{t+k} - \binom{k}{1}\epsilon_{t+k-1} + \binom{k}{2}\epsilon_{t+k-2} + \cdots + (-1)^k \epsilon_t$$

Now, taking the var, both sides we get

$$Var(\Delta^k y_t) = Var\left(\epsilon_{t+k} - \binom{k}{1}\epsilon_{t+k-1} + \binom{k}{2}\epsilon_{t+k-2} - \cdots + (-1)^k \epsilon_t\right)$$

Scanned with CamScanner

$$= E\left(\epsilon_{t+k} - \binom{k}{1}\epsilon_{t+k-1} + \cdots + (-1)^k \epsilon_t\right)^2$$

$$- \left[E\left(\epsilon_{t+k} - \binom{k}{1}\epsilon_{t+k-1} + \cdots + (-1)^k \epsilon_t\right)\right]^2$$

$$= E\left(\epsilon_{t+k} - \binom{k}{1}\epsilon_{t+k-1} + \cdots + (-1)^k \epsilon_t\right)^2 - 0$$

Now,

$$= E\left(\epsilon_{t+k}^2 + \binom{k}{1}^2 \epsilon_{t+k-1}^2 + \cdots + \epsilon_t^2\right)$$

$$Var(\Delta^k y_t) = \left[\binom{k}{0}^2 + \binom{k}{1}^2 + \binom{k}{2}^2 + \cdots + \binom{k}{k}^2\right] V$$

Now,

$$V = \frac{Var(\Delta^k y_t)}{\binom{2k}{k}}$$

$$V = \frac{\mu_2'(\Delta^k y_t)}{\binom{2k}{k}} \quad \text{——— (2)}$$

Here, it is noted that we calculate $\mu_2'(\Delta^k y_t)$ and note the observed var of $(\Delta^k y_t)$. Since, $E(\Delta^k y_t) = 0$. Hence, Once the value of 'k' known, the var V can be estimated.

Now, the question arise how to determine the value of k such that estimate of the random component can be obtained for this purpose, we use the following steps:

Step I :- Prepare difference table such that

t	y_t	Δy_t	$\Delta^2 y_t$

Step II :- Calculate $V_1 = \frac{\mu_2'(\Delta y_t)}{2}$

$$V_2 = \frac{\mu_2'(\Delta^2 y_t)}{6}$$

Step III :- If V_1 and V_2 do not differ significantly, then either of them can be regard as an estimate of V.

If they differ significantly then calculate V_3

$$V_3 = \frac{\mu_2'(\Delta^3 y_t)}{20}$$

If V_2 and V_3 do not differ significantly, any one of them can be taken as an estimate of V.

seasonal indices. [Sardar Patel U. B.Sc. 1975]

2·7. Auto-regression Series. The value of a series at any time 't' may depend upon its own value at times $t-1$, $t-2$, ..., $t-k$ (say) the relationship being linear. A series represented by the recurrence relation

$$U_t = f(U_{t-1}, U_{t-2}, \cdots, U_{t-k}) + \epsilon_t \qquad \ldots(2\cdot48)$$

$$= a_1 U_{t-1} + a_2 U_{t-2} + \ldots + a_k U_{t-k} + \epsilon_t$$

where f is a linear function and ϵ_t is the "disturbance" function such that ϵ_t's are identically and independently distributed (*i.i.d.*) random variables, $\epsilon_t \sim N(0, \sigma^2)$, is known as "auto-regressive" series of order k.

Under certain conditions to be satisfied by $a_1, a_2, \ldots, a_k$ the generating series in (2·48) can be shown to represent an oscillatory time series.

In the following sequences we shall assume (which we can do without loss of generality) that u_t's are measured from their means so that $E(U_t)=0$.

2·7·1. First Order Auto-regression (Markoff's Series)

$$U_{t+1}=aU_t+b+\varepsilon_{t+1},\ |a|<1$$

where ε_t's are *i.i.d.* $N(0, \sigma^2)$.

Since $E(U_t)=0=E(\varepsilon_t)$, by taking expectations of both sides, we get $b=0$. Thus the first order auto-regression equation reduces to

$$U_{t+1}=aU_t+\varepsilon_{t+1} \qquad \ldots(2\cdot49)$$

This is a linear differance equation of order 1. Its complementary function (C.F.) is the solution of

$$U_{t+1}=aU_t,$$

a homogeneous linear difference equation of order 1. If $U_t=\lambda^t$ is the trial solution, then

$$\lambda^{t+1}=a\lambda^t \Rightarrow \lambda=a,$$

ignoring the trivial solution $\lambda=0$.

$$\therefore \quad \text{C.F. is } U_t=a^t \qquad \ldots(2\cdot50)$$

Also (2·49) $\Rightarrow$ $(\mathsf{E}-a)\,U_t=\varepsilon_{t+1}$, E being the difference operator given by $\mathsf{E}U_t=U_{t+1}$, the interval of differencing being unity. Therefore particular integral (P.I.) is given by

$$U_t=\frac{1}{\mathsf{E}-a}\,\varepsilon_{t+1}=\frac{1}{\mathsf{E}}\left(1-\frac{a}{\mathsf{E}}\right)^{-1}\varepsilon_{t+1}$$

$$=\frac{1}{\mathsf{E}}\left(1+\frac{a}{\mathsf{E}}+\frac{a^2}{\mathsf{E}^2}+\ldots\right)\varepsilon_{t+1}$$

$$=\varepsilon_t+a\varepsilon_{t-1}+a^2\varepsilon_{t-2}+\ldots$$

$$=\sum_{j=0}^{\infty} a^j\,\varepsilon_{t-j}=\sum_{j=1}^{\infty} a^{t-j}\,\varepsilon_j \qquad \ldots(2\cdot50a)$$

$$=\sum_{j=0}^{t-1} a^j\,\varepsilon_{t-j}=\sum_{j=1}^{t} a^{t-j}\,\varepsilon_j, \text{ (if we assume that } \varepsilon_t=0, t<0)$$

Thus the general solution of (2·49) is

$$U_t=\text{C.F.}+\text{P.I.}$$

$$=a^t+\sum_{j=1}^{\infty} a^{t-j}\,\varepsilon_j \qquad \ldots(2\cdot50b)$$

If $|a|<1$ and the series is very large or has been made so by assuming that the series has started sometimes prior to the point $t=0$, then we have $|a|^t \to 0$ and the effective solution becomes

$$U_t=\sum_{j=1}^{\infty} a^{t-j}\,\epsilon_j \qquad \ldots(2\cdot50c)$$

which is a moving average of random elements with weights,

$$1,\ a,\ a^2,\ldots\ldots,\ a^{t-1}$$

2·7·2. Second Order Auto-regressive Series. (Yule's Series)

$$U_{t+2}+aU_{t+1}+bU_t=\epsilon_{t+2} \qquad \ldots(2\cdot51)$$

This is a second order linear difference equation. Its C.F. is the solution of

$$U_{t+2}+aU_{t+1}+bU_t=0$$

The trial solution $U_t=\lambda^t$ gives

$$\lambda^2+a\lambda+b=0$$

$$\Rightarrow \quad \lambda_1=\frac{-a+\sqrt{a^2-4b}}{2}, \quad \lambda_2=\frac{-a-\sqrt{a^2-4b}}{2}.$$

$\therefore$ C.F. is $\qquad U_t=A_1\lambda_1{}^t+A_2\lambda_2{}^t$

where A_1 and A_2 are arbitrary constants.

In order that (2·51) represents an oscillatory series we must have $a^2-4b<0$, since in that case λ_1 and λ_2 are complex and U_t can be expressed as a sine and cosine function of 't'. Hence we assume that $a^2-4b<0$ and then

$$\lambda_1=-\frac{a}{2}+\frac{i}{2}\sqrt{4b-a^2}=p(\cos\theta+i\sin\theta)$$

$$\text{where}\quad \left.\begin{aligned} -\frac{a}{2}&=p\cos\theta \\ \tfrac{1}{2}\sqrt{4b-a^2}&=p\sin\theta \end{aligned}\right\} \Rightarrow \begin{aligned} &p^2=b \Rightarrow p=\sqrt{b} \\ &\text{and } \cos\theta=-\frac{a}{2\sqrt{b}}, \\ &\sin\theta=\frac{\sqrt{4b-a^2}}{2p}=\sqrt{1-\frac{a^2}{4b}} \end{aligned}$$

and $$\lambda_2=p(\cos\theta-i\sin\theta)$$

Then the C F. becomes

$$\begin{aligned} U_t&=A_1p^t(\cos\theta+i\sin\theta)^t+A_2p^t(\cos\theta-i\sin\theta)^t \\ &=p^t(A\cos t\theta+B\sin t\theta) \qquad \text{(using De-Moivre's theorem)} \end{aligned}$$

where $A=A_1+A_2$, $B=i\,(A_1-A_2)$.

Since 't' can take very large values, it is necessary to assume that $|p|=\sqrt{b}<1$, since otherwise U_t can be infinite.

P.I. of (2·51). We have

$$(E^2+aE+b)U_t=\epsilon_{t+2}$$

Thus P.I. of (2·51) is given by

$$U_t=\frac{1}{(E^2+aE+b)}\,\epsilon_{t+2}$$

$$= \frac{1}{(E-\alpha)(E-\beta)} \varepsilon_{t+2}, \qquad \left.\begin{array}{l}\alpha = p(\cos\theta + i\sin\theta) = \lambda_1 \\ \beta = p(\cos\theta - i\sin\theta) = \lambda_2\end{array}\right\}$$

$$= \frac{1}{E^2}\left[\left(1-\frac{\alpha}{E}\right)^{-1}\left(1-\frac{\beta}{E}\right)^{-1}\right]\varepsilon_{t+2}$$

$$= \left\{1 + \frac{\alpha}{E} + \left(\frac{\alpha}{E}\right)^2 + \ldots\right\}\left\{1 + \frac{\beta}{E} + \left(\frac{\beta}{E}\right)^2 + \ldots\right\}\varepsilon_t$$

$$= \left(1 + \frac{\alpha+\beta}{E} + \frac{\alpha^2+\alpha\beta+\beta^2}{E^2} + \ldots\right)\varepsilon_t$$

$$= \left[\frac{\alpha-\beta}{\alpha-\beta} + \frac{\alpha^2-\beta^2}{\alpha-\beta}\cdot\frac{1}{E} + \frac{\alpha^3-\beta^3}{\alpha-\beta}\cdot\frac{1}{E^2} + \ldots\right]\varepsilon_t$$

$$= \sum_{j=0}^{\infty} \frac{\alpha^j - \beta^j}{\alpha - \beta}\varepsilon_{t-j+1}$$

Now $$\frac{\alpha^j - \beta^j}{\alpha - \beta} = \frac{p^j\, 2i \sin j\theta}{p.\, 2i \sin\theta} = \frac{2}{\sqrt{4b-a^2}} \cdot p^j \sin j\theta = \xi_j, \text{ (say).}$$

$$\therefore \qquad U_t = \sum_{j=0}^{\infty} \xi_j \; \varepsilon_{t-j+1}$$

Hence the complete solution of (2·51) is

$$U_t = p^t (A\cos t\theta + B\sin t\theta) + \sum_{j=0}^{\infty} \xi_j . \varepsilon_{t-j+1} \qquad \ldots(2·52)$$

If the series is very large or it has started up sometimes prior to the point $t=0$, then $p^t \to 0$ (since $|p|<1$) and $(A\cos t\theta + B\sin t\theta)$ being bounded functions for all t and θ, the effective solution of (2·51) reduces to

$$U_t = \sum_{j=0}^{\infty} \xi_j \; \varepsilon_{t-j+1}, \qquad \ldots(2·52a)$$

which is a moving average of random elements with damped harmonic weights $\xi_t = 2p^t \sin t\theta/\sqrt{4b-a^2}$ and infinite extent. Hence under the conditions

$$a^2 - 4b < 0 \text{ and } |b| < 1, \qquad \ldots(2·53)$$

the second order auto-regression (2·51) represents an oscillatory scheme.

2·7·3. General Auto-regression. Auto-regressive scheme of order k is given by

$$U_{t+k} + a_1 U_{t+k-1} + a_2 U_{t+k-2} + \ldots\ldots + a_i U_{t+k-i} + \ldots\ldots + a_k U_t = \varepsilon_{t+k} \qquad \ldots(2·54)$$

The trial solution $u_t = \lambda^t$, gives the C.F. of (2·54) as

$$U_t = A_1\lambda_1^{\,t} + A_2\lambda_2^{\,t} + \ldots\ldots + A_k\lambda_k^{\,t} \qquad \ldots(2·55)$$

where $A_1, A_2, \ldots, A_k$ are arbitrary constant and λ_i's are the roots of the equation

$$\lambda^k + a_1\lambda^{k-1} + a_2\lambda^{k-2} + \ldots\ldots + a_k = 0$$

In order that (2·54) represents an oscillatory series without increasing indefinitely, none of the λ_i's should be real and $|\lambda_i|$ must be less than unity.

Particular solution of (2·54) involves terms containing ε's. If ξ_t is a particular solution of the reduced equation such that

$$\left.\begin{aligned} \xi_0&=0 \\ \xi_1+a_1\xi_0&=1 \\ \xi_2+a_1\xi_1+a_2\xi_0&=0 \\ &\vdots \\ \xi_{k-1}+a_1\xi_{k-2}+a_2\xi_{k-3}+\dots+a_{k-1}\xi_0&=0 \end{aligned}\right\} \qquad \dots(*)$$

then it can be verified on substitution that a particular solution of (2·54) is given by

$$U_t=\sum_{j=0}^{\infty} \xi_j\,\varepsilon_{t-j+1} \qquad \dots(2\text{·}55a)$$

We can always find ξ_t, since (*) amounts to imposing k conditions on the 'k' arbitrary constants $A_1, A_2, \dots, A_k$.

Hence complete solution of (2·54) is

$$U_t=A_1\lambda_1^t+A_2\lambda_2^t+\dots\dots+A_k\lambda_k^t+\sum_{j=0}^{\infty}\xi_j\,\varepsilon_{t-j+1} \qquad \dots(2\text{·}56)$$

If $|\lambda_i|<1\ \forall\ i=1, 2, \dots, k$ and the series is assumed to have started sometime prior to the point $t=0$, then the effective solution becomes

$$U_t=\sum_{j=0}^{\infty}\xi_j\,\varepsilon_{t-j+1}, \qquad \dots(2\text{·}56a)$$

which is a moving average of infinite extent.

The principle of least squares can be used to determine the constants $a_1, a_2, \dots, a_k$ of (2·54). Thus the normal equations for estimating a_i's (for a long time series) are

$$-r_i=a_1r_{i-1}+a_2r_{i-2}+\dots\dots+a_kr_{i-k},\ (i=1, 2, \dots, k) \qquad \dots(2\text{·}57)$$

where r_k is auto-correlation coefficient of order k (*c.f.* § 2·8), since $r_{-i}=r_i$ and ε_t and U_t are uncorrelated.

Alternately (2·57) can be obtained (for a long time series), on multiplying both sides of (2·54) by U_{t+k-1}, taking expectation and then dividing by Var (U_t).

2·8. Auto correlation and Correlogram. Suppose from the original series U_t $(t=1, 2, \dots, n)$ we obtain $(n-k)$ pairs of observations (U_t, U_{t+k}), $t=1, 2, \dots, (n-k)$ with a long period k. The ordinary product moment correlation between the two series U_t and U_{t+k}, $t=1, 2, \dots, (n-k)$ is called the *auto correlation* or *serial correlation* (r_k) of order k. Thus

$$r=\frac{E(U_i, U_{i+k})}{[\text{Var}(U_i)\ \text{Var}(U_{i+k})]^{1/2}}\ ; i=1, 2, \dots, n-k \qquad \dots(2\text{·}58)$$

Obviously, we have

$$r_0=1 \qquad \text{and} \qquad r_{-k}=r_k$$

The diagram obtained by plotting r_k against k is known as *correlogram*. Correlogram is an important tool which provides an objective criterion for exploring the nature of the internal structure of the time series. The oscillatory movements of an observed. time series may be attributed to any one of the following schemes :

(*i*) The moving average of random elements.

(*ii*) The series represented by a linear combination of harmonic elements.

(*iii*) An auto-regressive series.

Correlogram analysis enables us to decide in any particular case as to the cause of oscillation.

2·8·1. Correlogram of Moving Average. For a moving average of extent m, with weights $(a_1, a_2, \ldots, a_m)$ of random components $(\varepsilon_i\ ;\ i=1, 2, \ldots)$, the generated series is given by

$$U_i = a_1\varepsilon_{i+1} + a_2\varepsilon_{i+2} + \ldots + a_{k+1}\,\varepsilon_{i+k+1} + \ldots + a_m\varepsilon_{i+m}$$

$$U_{i+k} = a_1\varepsilon_{i+k+1} + a_2\varepsilon_{i+k+2} + \ldots + a_{m-k}\,\varepsilon_{i+m} + \ldots + a_m\varepsilon_{i+k+m}$$

where ε_i's are *i.i.d.* $N(0, \sigma^2)$. Thus

$$E(U_i) = 0 = E(U_{i+k})$$

and $$\text{Var}(U_i) = E(U_i^2) = (a_1^2 + a_2^2 + \ldots + a_m^2)\sigma^2 = \sigma^2 \sum_{j=1}^{m} a_j^2,\ \forall\ i = 1, 2, \ldots$$

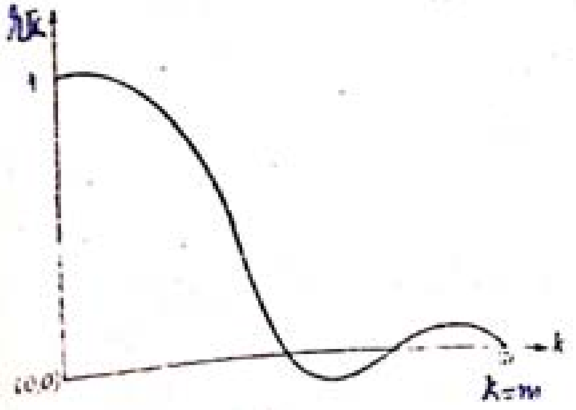
$$E(U_i U_{i+k}) = (a_1 a_{k+1} + a_2 a_{k+2} + \ldots + a_{m-k} a_m)\sigma^2 = \sigma^2 \sum_{j=1}^{m-k} a_j\, a_{j+k},\ k < m$$

$$r_k = \frac{E(U_i,\ U_{i+k})}{\sqrt{\text{Var}(U_i)\ \text{Var}(U_{i+k})}}$$

$$= \frac{\sum_{j=1}^{m-k} a_j a_{j+k}}{\sum_{j=1}^{m} a_j^2},\quad \text{if } k < m \qquad \ldots (2\cdot59)$$

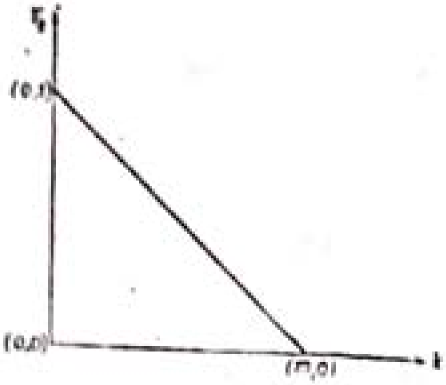
and $$r_k = 0,\ \text{if } k \geqslant m$$

Thus the corrlogram will oscillate between the points (0, 1) and $(m, 0)$ and then coincide with the k-axis when $k \geqslant m$, as shown in the figure below :



In particular if $a_i = \frac{1}{m}$, $(i=1, 2, \ldots, m)$ then

$$\left.\begin{aligned} r_k &= \frac{(m-k)\cdot\frac{1}{m^2}}{m\cdot\frac{1}{m^2}} = 1-\frac{k}{m}, \quad k<m \\ &= 0, \quad \text{if } k>m. \end{aligned}\right\} \quad \ldots(2{\cdot}59a)$$



Thus for a series generated by an m-point simple moving average of random components, the correlogram consists of a straight line joining the points $(m, 0)$ and $(0, 1)$ together with k-axis from the point $(m, 0)$ onwards.

2·8·2. **Correlogram of Harmonic Series.** For simplicity, let us consider the harmonic series

$$U_t = A \sin \theta t + \epsilon_t$$

where A is the amplitude of the sine term of period of oscillation $2\pi/\theta$ and ϵ_t is the random component independent of $A \sin \theta t$, and ϵ_t's are i.i.d $N(0, \sigma^2)$. Thus for $k>0$, we have

$$\text{Cov}(U_t, U_{t+k}) = E(U_t\, U_{t+k})\,; \qquad [\because\ E(U_t)=0]$$

$$= E[(A \sin \theta t + \epsilon_t)(A \sin \theta\, \overline{t+k} + \epsilon_{t+k})]$$

$$= A^2 E[\sin \theta t, \sin \theta\,(t+k)],$$

since ϵ_t is uncorrelated with $A \sin \theta t$.

$$\therefore\ \text{Cov}(U_t, U_{t+k}) = A^2 \cdot \frac{1}{n} \sum_{t=1}^{n} \sin \theta t \cdot \sin \theta(t+k),$$

if the series consists of n terms.

$$\text{Cov}(U_t, U_{t+k}) = \frac{A^2}{2n} \sum_{t=1}^{n} \Big[\cos \theta k - \cos (2t+k)\theta \Big]$$

$$= \frac{A^2}{2} \cos \theta k - \frac{A^2}{2n} \sum_{t=1}^{n} \cos \theta\,(2t+k) \qquad \ldots(*)$$

Let $\quad S = \sum_{t=1}^{n} \cos \theta(2t+k)$

$\sin C - \sin D = 2\cos\frac{C+D}{2}\cdot\sin\frac{C-D}{2}$

$$\therefore \quad 2\sin\theta\,.\,S=\sum_{t=1}^{n}[\sin\theta(2t+k+1)-\sin\theta\,(2t+k-1)]$$
$$=[\sin\theta\,(2n+k+1)-\sin\theta\,(k+1)]$$
$$=2\cos\theta\,(n+k+1)\sin n\theta$$
$$\Rightarrow \quad S=\sum_{t=1}^{n}\cos\theta\,(2t+k)=\frac{\cos\theta\,(n+k+1)\sin n\theta}{\sin\theta} \quad \ldots(**)$$

Substituting in (*), we get

$$\text{Cov}\,(U_t, U_{t+k})=\frac{A^2}{2}\cos\theta k-\frac{A^2}{2n}\cdot\frac{\cos\theta\,(n+k+1)\sin n\theta}{\sin\theta}$$

Since $\quad |\cos\theta\,(n+k+1)|\leq 1$ and $|\sin n\theta|\leq 1$ for all θ and n,
$\{\cos\theta(n+k+1)\sin n\theta/\sin\theta\}$, is bounded for all θ and n.

$$\therefore \quad \lim_{n\to\infty}\frac{A^2}{2n}\cdot\frac{\sin\theta\,(n+k+1)\sin n\theta}{\sin\theta}=0$$

Hence for a large series (*i.e.* $n=\infty$),

$$\text{Cov}\,(U_t, U_{t+k})=\frac{A^2}{2}\cos\theta k$$

$$\text{Var}\,(U_t)=E(U_t^2)=E(A\sin\theta t+\epsilon_t)^2$$
$$=A^2.E\,(\sin^2\theta t)+\sigma^2$$
$$=\frac{A^2}{n}\sum_{t=1}^{n}\sin^2\theta t+\sigma^2$$
$$=\frac{A^2}{2n}\sum_{t=1}^{n}(1-\cos\theta.2t)+\sigma^2$$
$$=\frac{A^2}{2}+\sigma^2-\frac{A^2}{2n}\sum_{t=1}^{n}\cos\theta\,.\,2t$$
$$=\frac{A^2}{2}+\sigma^2-\frac{A^2}{2n}\left[\frac{\cos\theta\,(n+1)\,.\,\sin n\theta}{\sin\theta}\right]$$

[taking $k=0$ in (**)].

$\sin 2A = 2\sin A\cdot\cos A$

$\cos 2A = \cos^2 A-\sin^2 A$

$= 2\cos^2 A - 1$

$= 1 - 2\sin^2 A$

$\tan 2A = \dfrac{2\tan A}{1-\tan^2 A}$

$$\therefore \quad \text{Var}(U_t)=\frac{A^2}{2}+\sigma^2, \text{ for large values of } n.$$

Hence for large n, *i.e.*, for long series.

$$r_k=\frac{\text{Cov}\,(U_t, U_{t+k})}{\sqrt{\text{Cov}\,(U_t)\,\text{Var}\,(U_{t+k})}}$$
$$=\frac{\text{Cov}\,(U_t\,.\,U_{t+k})}{\text{Var}\,(U_t)} \qquad [\because \text{Var}\,(U_t)=\text{Var}\,(U_{t+k}) \text{ for long series}]$$
$$=\frac{\frac{A^2}{2}.\cos\theta k}{\sigma^2+\frac{A^2}{2}}=c\cos\theta k\,,\ k>0. \qquad \ldots(2\cdot60)$$

where $c = \dfrac{A^2/2}{\sigma^2+(A^2/2)} = \dfrac{A^2}{A^2+2\sigma^2}$, is independent of '$k$'.

Thus the corelogram is a cyclic curve with the same period as the original series and amplitude 'c', which is independent of 'k'. Hence the correlogram of a series of harmonic terms will oscillate but will not vanish or be damped.

For the more general series

$$U_t = A_1 \sin \theta_1 t + A_2 \sin \theta_2 t + \epsilon_t,$$

under the usual assumptions, *viz.*,

$$E(U_t)=0\ ;\ E(\epsilon_t)=0,\ \text{Var}\ (\epsilon_t)=\sigma^2=E(\epsilon_t)^2$$

and that $A_1 \sin \theta_1 t$ and $A_2 \sin \theta_2 t$ and ϵ_t are pairwise uncorrelated, we shall get for large n,

$$r_k = c_1 \cos \theta_1 k + c_2 \cos \theta_2 k$$

where

$$\left.\begin{aligned} c_i = \frac{\dfrac{A_i^2}{2}}{\sum\limits_{i=1}^{2}\left(\sigma^2+\dfrac{A_i^2}{2}\right)} = \frac{A_i^2}{\sum\limits_{i=1}^{2}(A_i^2+2\sigma^2)}\ ;\ i=1,2 \end{aligned}\right\} \quad \ldots(2{\cdot}60a)$$

2·8·3. Correlogram of Auto-Regressive Series. For the first order auto-regressive series,

$$U_{t+1} = aU_t + \epsilon_t\ ,\ |\,a\,| < 1 \quad \ldots(*)$$

the effective solution for long series is (*c.f.* 2·50*a*)

$$U_t = \sum_{j=1}^{\infty} a^{t-j}\,\epsilon_j$$

$$= 1.\,\epsilon_t + a\epsilon_{t-1} + a^2\epsilon_{t-2} + \ldots$$

a moving average of random elements of infinite extent with weights

$$\xi_j = a^{j-1},\ (j=1, 2, 3, \ldots)$$

Hence its correlogram is (*c.f.* 2·59)

$$r_k = \frac{\sum\limits_{j=1}^{\infty} \xi_j\,\xi_{j+k}}{\sum\limits_{j=1}^{\infty} \xi_j^2} = \frac{1.a^k + a.a^{k+1} + a^2.a^{k+2} + \ldots}{1 + a^2 + a^4 + \ldots}$$

$$\therefore \quad r_k = a^k \quad \ldots(2{\cdot}61)$$

Aliter. Multiplying both sides of (*) by U_{t-k}, taking expectation and dividing by Var (U_t), we get for large series

$$r_{k+1} - ar_k = 0,$$

which is a linear (homogeneous) difference equation of first order in r_k.

$$\Rightarrow \quad r_k = a^k \quad \ldots[c.f.\ (2{\cdot}50)]$$

Second Order Auto-regressive series is given by

$$U_{t+2} + aU_{t+1} + bU_t = \epsilon_{t+2}\ ,\ |\,b\,| < 1$$

Multiplying both sides by U_{t-k}, taking expectations and dividing by variance of U_t, we get for a long series, 2·62

$$r_{k+2}+ar_{k+1}+br_k=0\ ,\ k\geqslant -2$$

which is a homogeneous linear difference equation of second order in r_k

$$\therefore \qquad r_k=p^k\ (A\cos\theta k+B\sin\theta k)$$

where $p=\sqrt{b}$, $\cos\theta=-a/(2p)$, $|b|<1$, $a^2-4b<0$ (*c.f.* § 2·7·2)

The arbitrary constants A and B can be obtained by using the conditions $r_0=1$ and $r_{-1}=r_1$ giving $A=1$ and

$$\frac{1}{p}(A\cos\theta-B\sin\theta)=p\ (A\cos\theta+B\sin\theta)$$

$$\Rightarrow \qquad B=\frac{1-p^2}{1+p^2}\cot\theta=\cot\phi\ \text{(say)}$$

$$\therefore \qquad r_k=p^k\ (\cos k\theta+\cot\phi.\ \sin k\theta)$$

$$\Rightarrow \qquad r_k=p^k.\frac{\sin(k\theta+\phi)}{\sin\phi}\ .\ k\geqslant -2,\ |p|<1. \qquad \text{...(2·62)}$$

p^k is called the '*damping*' *factor*.

This elegant approach for finding 'r_k' enables us to generalise the result to a series of any order. Thus for auto-regressive series of order m, *viz.*,

$$U_{t+m}+a_1U_{t+m-1}+\ldots+a_mU_t=\varepsilon_{t+m}\ ,$$

multiplying both sides by U_{t-k}, taking expectations and dividing by Var (U_t), we get if the series is long one

$$r_{m+k}+a_1r_{m+k-1}+a_2r_{m+k-2}+\ldots+a_mr_k=0$$

Its solution is [*c.f.* (2·55)]

$$r_k=\sum_{i=1}^{m}A_i\,\lambda_i^k$$

where λ_i's are the roots of the equation

$$\lambda^m+a_1\lambda^{m-1}+a_2\lambda^{m-2}+\ldots+a_{m-1}\lambda+a_m=0$$

and A_i's are 'm' arbitrary constants which will be determined from the conditions :

$$r_0=1 \text{ and } r_{-i}=r_i\ ;\ i=1,2,\ldots,(m-1)$$

Important Remarks on Auto-correlation and Auto-regression. 1. It should be clearly noted that the 'disturbance factor' or 'error term' ε_t in an auto-regressive series plays an altogether different role. In ordinary regression ε_i is independent of U_j for all i and j. But in an auto-regressive series U_j and ε_i are independent if $i>j$ but if $i<j$, ε_i's are certainly not independent of U_j and become a part of the system itself.

2. The Cov (U_t, U_{t+k}), $k>0$ is not at all affected, whatever be the magnitude of ε_j but Var (U_t) will increase by an amount equal to the variance of ε_t. Hence greater the variance of ε_t, lesser the value of r_k, (this decrease in r_k is same for all k).

Scanned with CamScanner

$$(m+k)(m+k-1)\ldots(m+1)$$

2·9. 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.

2·9·1. Variate Difference Method. Let us suppose that the series can be represented as the sum of functional part and random component as given below

$$U_t = a_0 + a_1 t + a_2 t^2 + \ldots + a_{k-1} t^{k-1} + \varepsilon_t \qquad \ldots(2{\cdot}63)$$

where ε's are indepently and identically distributed with

$$\left.\begin{aligned} E(\varepsilon_t) &= 0 \\ \text{Cov}(\varepsilon_i, \varepsilon_j) &= 0, \quad i \neq j \\ \text{and} \quad \text{Var}(\varepsilon_t) &= V \text{ (say)} \end{aligned}\right\} \qquad \ldots(*)$$

We know that for a polynomial U_t of mth degree in t, $\Delta^r(U_t)=0$, $r \geq m$ where $\Delta U_t = U_{t+1} - U_t$, the interval of differencing being unity.

Hence as we proceed with differencing in (2·63), the random element becomes more and more predominant until finally the systematic component is completely eliminated. Thus from (2·63), we get

$$\Delta^k U_t = \Delta^k \varepsilon_t \qquad \ldots(2{\cdot}64)$$

We have

$$\Delta^k \varepsilon_t = (\mathrm{E}-1)^k \varepsilon_t, \quad \mathrm{E} = 1 + \Delta$$

$$= \left[\mathrm{E}^k - \binom{k}{1} \mathrm{E}^{k-1} + \binom{k}{2} \mathrm{E}^{k-2} + \ldots + (-1)^k \right] \varepsilon_t$$

$$= \varepsilon_{t+k} - \binom{k}{1} \varepsilon_{t+k-1} + \binom{k}{2} \varepsilon_{t+k-2} - \ldots + (-1)^k \varepsilon_t$$

$$\therefore \quad E(\Delta^k U_t) = E(\Delta^k \varepsilon_t) = 0, \text{ [using (*) and (2·64)]} \qquad \ldots(2{\cdot}65)$$

$$\therefore \quad \text{Var}(\Delta^k U_t) = \text{Var}\, \Delta^k(\varepsilon_t) \qquad \text{[from (2·64)]}$$

$$= E[\Delta^k \varepsilon_t]^2 \qquad [\because \; E(\Delta^k \varepsilon_t) = 0]$$

$$= E\left[\varepsilon_{t+k} - \binom{k}{1} \varepsilon_{t+k-1} + \binom{k}{2} \varepsilon_{t+k-2} - \ldots + (-1)^k \varepsilon_t \right]^2$$

$$= \left[\binom{k}{0}^2 + \binom{k}{1}^2 + \binom{k}{2}^2 + \ldots + \binom{k}{k}^2 \right] V$$

[using (*)]

$$= \binom{2k}{k} V,$$

the sum in the bracket is obtained by comparing the coefficient of x^k in both sides of the identity

$$(1+x)^k (x+1)^k = (1+x)^{2k}$$

Hence an estimate of V is given by

$$\hat{V} = \frac{\text{Var}(\Delta^k U_t)}{\binom{2k}{k}} = \frac{\mu_2'(\Delta^k U_t)}{\binom{2k}{k}} \qquad \ldots(2{\cdot}66)$$

It is to be noticed that we calculate $\mu_2'(\Delta^k U_t)$ and not the observed variance of $\Delta^k U_t$, since $E(\Delta^k U_t) = 0$ as proved in (2·65). Hence, once the value of k is known, the variance V can be estimated.

The variate difference method essentially consists of the following steps :

(i) Preparation of difference table as shown :

$$t \mid U_t \mid \Delta U_t \mid \Delta^2 U_t \mid \ldots$$

(ii) Calculate $V_1 = \mu_2'(\Delta U_t)/2$, ...[$k=1$ in (2·66)]

and $V_2 = \mu_2'(\Delta^2 U_t) \Big/ \binom{4}{2} = \dfrac{\mu_2'(\Delta^2 U_t)}{6}$, [$k=2$ in (2·66)]

(iii) If V_1 and V_2 do not differ significantly then either of them can be regarded as an estimate of V. If they differ significantly, calculate

$$V_3 = \mu_2'(\Delta^3 U_t \Big/ \binom{6}{3} = \mu_2'(\Delta^3 U_t)/20, \qquad [k=3 \text{ in } (2·66)]$$

If V_2 and V_3 do not differ significantly, then any one of them can be taken as estimate of V, otherwise, calculate V_4 etc. and proceed similarly till two successive estimates are homogeneous.

Significance of ($V_k - V_{k+1}$). Homogeneity of two successive estimates of V cannot be tested by Variance Ratio Test (F-test) since the consecutive terms are not independent. O. Andersen obtained the standard error of $(V_k - V_{k+1})$ and found that for large samples

$$R_k = \frac{V_k - V_{k+1}}{V_k} \cdot H_{kN} \sim N(0, 1) \qquad \ldots(2·67)$$

where V_k and V_{k+1} are consecutive estimates of V from the kth and $(k+1)$th differences of U_t and H_{kN}; a function of k and N, corresponds to the variance of the ratio and has been tabulated in the book "Variate Difference Method" by G. Tintner. Thus if $| R_k | > 1·96$, the difference is significant (otherwise not) at 5% level of significance.

Remarks 1. Incidentally the variate difference method gives the degree of the polynomial with which the trend component of a series can be represented, the only assumption regarding the functional form of the trend is that it must be smooth and not of zig-zag nature with small period, since in the latter case successive differencing will not eliminate the functional part.

2. Fisher's F-test cannot be applied here since the consecutive terms are not independent, *e.g.*,

$$\Delta\varepsilon_1 = \varepsilon_2 - \varepsilon_1, \quad \Delta\varepsilon_2 = \varepsilon_3 - \varepsilon_2, \quad \Delta\varepsilon_3 = \varepsilon_4 - \varepsilon_3$$

etc. Here $\Delta\varepsilon_1$ and $\Delta\varepsilon_2$ are not independet since ε_2 is common.

Example 2·18. *Find the variance of the random camponent in the following series by Variate-Difference method :*

t :	1	2	3	4	5	6	7	8	9	10
U_t :	106	118	124	94	82	88	87	88	88	68
t :	11	12	13	14	15	16	17	18	19	20
U_t :	98	115	135	104	96	110	107	97	75	86
t :	21	22	23	24	25					
U_t :	111	125	78	86	102					

Solution. First of all we prepare the difference table as given below :

$\triangle\epsilon_1=\epsilon_2-\epsilon_1$, $\triangle\epsilon_2=\epsilon_3-\epsilon_2$, $\triangle\epsilon_3=\epsilon_4-\epsilon_3$ ... consecutive

...re $\triangle\epsilon_1$ and $\triangle\epsilon_2$ are not independent since ϵ_2 is common.

Example 2·25. *Find the variance of the random camponent in the following series by Variate-Difference method :*

t :	1	2	3	4	5	6	7	8	9	10
U_t :	106	118	124	94	82	88	87	88	88	68
t :	11	12	13	14	15	16	17	18	19	20
U_t :	98	115	135	104	96	110	107	97	75	86
t :	21	22	23	24	25					
U_t :	111	125	78	86	102					

Solution. First of all we prepare the difference table as given on page 2·90.

$$\mu'_2(\triangle U_t)=\frac{\Sigma(\triangle U_t)^2}{24}=\frac{8536}{24}=355{\cdot}667$$

$$\mu'_2(\triangle^2 U_t)=\frac{\Sigma(\triangle^2 U_t)^2}{23}=\frac{17424}{23}=757{\cdot}565$$

$$\mu_2'(\triangle^3 U_t)=\frac{\Sigma(\triangle^3 U_t)^2}{22}=\frac{45{,}232}{22}=2{,}056$$

$$\mu'_2(\triangle^4 U_t)=\frac{\Sigma(\triangle^4 U_t)^2}{21}=\frac{1{,}36{,}365}{21}=6{,}493{\cdot}571$$

$$V_1=\frac{355{\cdot}667}{\binom{2}{1}}=177{\cdot}8335,\quad V_2=\frac{757{\cdot}565}{\binom{4}{2}}=126{\cdot}2608$$

$$V_3=\frac{2056}{\binom{6}{3}}=102{\cdot}8,\quad V_4=\frac{6{,}493{\cdot}571}{\binom{8}{4}}=92{\cdot}7653$$

From *Titner's book*, we note that

$$H_{1,25}=9{\cdot}065,\ H_{2,25}=11{\cdot}499,\ H_{3,25}=12{\cdot}9445$$

Now $R_1=\dfrac{51{\cdot}5727}{177{\cdot}8335}\times 9{\cdot}065=2{\cdot}62>1{\cdot}96$

$\therefore$ The difference between V_1 and V_2 is significant.

$$R_2=\frac{V_2-V_3}{V_2}\ H_{2,25}=\frac{23{\cdot}4605}{126{\cdot}2608}\times 11{\cdot}499=2{\cdot}13>1{\cdot}96$$

The difference between V_2 and V_3 is also significant.

$$R_3=\frac{V_3-V_4}{V_3}\ H_{3,25}=\frac{10{\cdot}0347}{102{\cdot}8}\times 12{\cdot}9445=1{\cdot}26$$

DIFFERENCE TABLE

t	U_t	ΔU_t	$\Delta^2 U_t$	$\Delta^3 U_t$	$\Delta^4 U_t$
1	106				
		12			
2	118		−6		
		6		−30	
3	124		−36		84
		−30		54	
4	94		18		−54
		−12		0	
5	82		18		−25
		6		−25	
6	88		−7		34
		−1		9	
7	87		2		−12
		1		−3	
8	88		−1		−16
		0		−19	
9	88		−20		89
		−20		70	
10	68		50		−133
		30		−63	
11	98		−13		79
		17		16	
12	115		3		−70
		20		−54	
13	135		−51		128
		−31		74	
14	104		23		−75
		−8		−1	
15	96		22		−38
		14		−39	
16	110		−17		49
		−3		10	
17	107		−7		−15
		−10		−5	
18	97		−12		50
		−22		45	
19	75		33		−64
		11		−19	
20	86		14		−6
		25		−25	
21	111		−11		−25
		14		−50	
22	125		−61		166
		−47		116	
23	78		55		−163
		8		−47	
24	86		8		
		16			
25	102				

Since $R_3 < 1{\cdot}96$, the difference between V_3 and V_4 is not significant and so any of these two can be taken as the variance of the random component.

Hence variance of the random component $=$ **102·8 or 92·7653**.

"Variate Difference Method" by G. Tintner.

TABLE 20

COEFFICIENTS H_{kN}
(k = Order of Difference)

Number of Items in the Original Series	$k=0$		$k=1$		$k=2$		$k=3$	
N	H_{0N}	D.D.*	H_{1N}	D.D.*	H_{2N}	D.D.*	H_{3N}	D.D.*
10	2.875	0.1384	4.748	0.3140	5.287	0.4483	5.263	0.5486
20	4.259	0.1042	7.888	0.2354	9.770	0.3458	10.749	0.4401
30	5.301	0.0869	10.242	0.1948	13.228	0.2878	15.150	0.3710
40	6.170	0.0762	12.190	0.1694	16.106	0.2502	18.860	0.3244
50	6.932	0.0687	13.884	0.1517	18.608	0.2237	22.104	0.2907
60	7.619	0.0630	15.401	0.1386	20.845	0.2039	25.011	0.2650
70	8.249	0.0585	16.786	0.1282	22.884	0.1883	27.661	0.2449
80	8.834	0.0549	18.068	0.1198	24.767	0.1758	30.110	0.2284
90	9.383	0.0518	19.266	0.1130	26.525	0.1654	32.394	0.2147
100	9.901	0.04530	20.396	0.09818	28.179	0.14322	34.541	0.18562
150	12.166	0.03812	25.305	0.08212	35.340	0.11924	43.822	0.15402
200	14.072	0.03352	29.411	0.07420	41.302	0.10426	51.523	0.13436
250	15.748	0.03030	33.121	0.06276	46.515	0.09378	58.241	0.12062
300	17.263	0.02784	36.259	0.05958	51.204	0.08594	64.272	0.11038
350	18.655	0.02590	39.238	0.05536	55.501	0.07978	69.791	0.10236
400	19.950	0.02432	42.006	0.05194	59.490	0.07476	74.909	0.09582
450	21.166	0.02300	44.603	0.04910	63.228	0.07060	79.700	0.09044
500	22.316	0.02188	47.058	0.04664	66.758	0.06706	84.222	0.08586
550	23.410	0.02088	49.390	0.04454	70.111	0.06400	88.515	0.08188
600	24.454	0.02004	51.617	0.04270	73.311	0.06134	92.609	0.07842
650	25.456	0.01928	53.752	0.04108	76.378	0.05896	96.530	0.07538
700	26.420	0.01860	55.806	0.03962	79.326	0.05684	100.299	0.07266
750	27.350	0.01798	57.787	0.03828	82.168	0.05496	103.932	0.07020
800	28.249	0.01742	59.701	0.03712	84.916	0.05322	107.442	0.06798
850	29.120	0.01694	61.557	0.03602	87.577	0.05166	110.841	0.06596
900	29.967	0.01646	63.358	0.03502	90.160	0.05022	114.139	0.06414
950	30.790	0.01602	65.109	0.03412	92.671	0.04890	117.346	0.06240
1000	31.591	....	66.815	....	95.116	....	120.466	....

* Divided difference, positive.

TABLE 20 (continued)

COEFFICIENTS H_{kN}
(k = Order of Difference)

Number of Items in the Original Series	$k=4$		$k=5$		$k=6$		$k=7$	
N	H_{4N}	D.D*	H_{5N}	D.D*	H_{6N}	D.D*	H_{7N}	D.D*
10	4.985	0.6179	4.675	0.6539	4.893	0.6128	....	....
20	11.164	0.5188	11.214	0.5833	11.021	0.6357	10.662	0.6780
30	16.352	0.4443	17.047	0.5078	17.378	0.5624	17.442	0.6092
40	20.795	0.3916	22.125	0.4520	23.002	0.5054	23.534	0.5524
50	24.711	0.3526	26.645	0.4091	28.056	0.4604	29.058	0.5065
60	28.237	0.3223	30.736	0.3755	32.660	0.4244	34.123	0.4691
70	31.460	0.2952	34.491	0.3482	36.904	0.3947	38.814	0.4378
80	34.412	0.2814	37.973	0.3255	40.851	0.3699	43.192	0.4113
90	37.226	0.2617	41.228	0.3065	44.550	0.3488	47.305	0.3886
100	39.843	0.22620	44.293	0.26514	48.038	0.30244	51.191	0.33804
150	51.153	0.18738	57.550	0.21962	63.160	0.25084	68.093	0.28104
200	60.522	0.16316	68.531	0.19106	75.702	0.21818	82.145	0.24456
250	68.680	0.14628	78.084	0.17112	86.611	0.19528	94.373	0.21888
300	75.994	0.13370	86.640	0.15624	96.375	0.17820	105.317	0.19962
350	82.679	0.12384	94.452	0.14462	105.285	0.16482	115.298	0.18456
400	88.871	0.11588	101.683	0.13520	113.526	0.15400	124.526	0.17240
450	94.665	0.10920	108.443	0.12742	121.226	0.14506	133.146	0.16228
500	100.125	0.10372	114.814	0.12080	128.479	0.13748	141.260	0.15376
550	105.311	0.09884	120.854	0.11514	135.353	0.13094	148.948	0.14640
600	110.253	0.09462	126.611	0.11016	141.900	0.12528	156.268	0.14000
650	114.984	0.09088	132.119	0.10582	148.164	0.12026	163.268	0.13436
700	119.528	0.08760	137.410	0.10190	154.177	0.11554	169.986	0.12934
750	123.908	0.08460	142.505	0.09842	159.954	0.11206	176.453	0.12484
800	128.138	0.08190	147.426	0.09526	165.557	0.10818	182.695	0.12076
850	132.233	0.07946	152.189	0.09240	170.966	0.10490	188.733	0.11708
900	136.206	0.07722	156.809	0.08976	176.211	0.10188	194.587	0.11370
950	140.067	0.07514	161.297	0.08734	181.305	0.09912	200.272	0.11060
1000	143.824	....	165.664	....	186.261	....	205.802	....

* Divided difference, positive.

TABLE 20 (concluded)

COEFFICIENTS H_{kN}
(k = Order of Difference)

Number of Items in the Original Series	$k = 8$		$k = 9$		$k = 10$	
N	H_{8N}	D.D.*	H_{9N}	D.D.*	H_{10N}	D.D.*
10	....	....	....	....	....	....
20	10.191	0.7117	9.647	0.7378	9.061	0.7566
30	17.308	0.6491	17.025	0.6830	16.627	0.7120
40	23.799	0.5937	23.855	0.6300	23.747	0.6619
50	29.736	0.5479	30.155	0.5849	30.366	0.6178
60	35.215	0.5097	36.004	0.5465	36.544	0.5799
70	40.312	0.4774	41.469	0.5138	42.343	0.5469
80	45.086	0.4498	46.607	0.4853	47.812	0.5183
90	49.584	0.4259	51.460	0.4607	52.995	0.4930
100	53.843	0.37192	56.067	0.40402	57.925	0.43436
150	72.439	0.31022	76.268	0.33832	79.643	0.36532
200	87.950	0.27022	93.184	0.29518	97.909	0.31938
250	101.461	0.24192	107.943	0.26440	113.878	0.28634
300	113.557	0.22062	121.163	0.24118	128.195	0.26128
350	124.588	0.20392	133.222	0.22290	141.259	0.24152
400	134.784	0.19040	144.367	0.20808	153.335	0.22548
450	144.304	0.17920	154.771	0.19580	164.609	0.21210
500	153.264	0.16968	164.561	0.18538	175.214	0.20080
550	161.748	0.16156	173.830	0.17642	185.254	0.19108
600	169.826	0.15444	182.651	0.16864	194.808	0.18258
650	177.548	0.14818	191.083	0.16174	203.937	0.17510
700	184.957	0.14260	199.170	0.15564	212.692	0.16846
750	192.087	0.13762	206.952	0.15014	221.115	0.16250
800	198.968	0.13310	214.459	0.14520	229.240	0.15710
850	205.623	0.12900	221.719	0.14072	237.095	0.15222
900	212.073	0.12526	228.755	0.13658	244.706	0.14774
950	218.336	0.12180	235.584	0.13282	252.093	0.14364
1000	224.426	....	242.225	....	259.275	....

* Divided difference, positive.