II INTERNAL

Time Series Data's Stationarity

- Stationarity ?
- first difference ?
- log linear or log transformation ?
- Normality assumptn (LLRM) (all data sets follow normal distn)

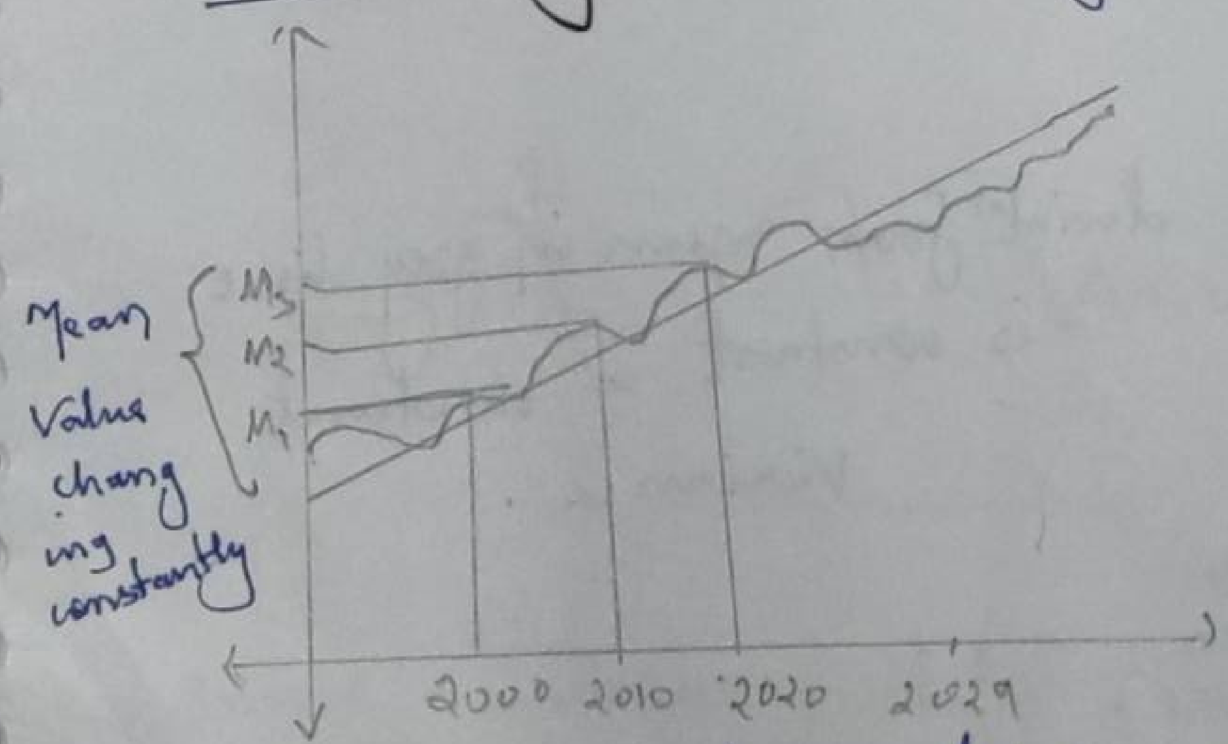
  (If not following normal dist, you cannot have an error term which is normal, if u is normal then only one can run OLS)

So to make a data set from non-normal data to normal data we do a log transformation. Especially in time series.

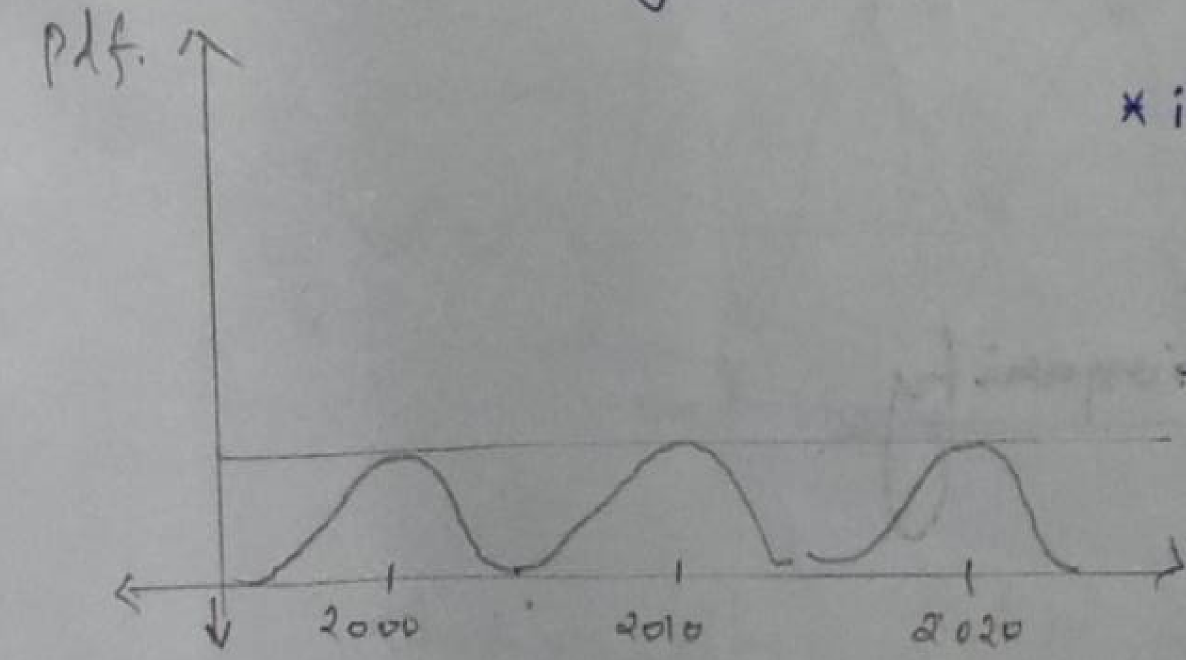
- log transformation makes the data set a unit free data.

- **Stationarity** (1st conditn of stationarity ; const. mean)

2 types of stationarity
- weak
- strong.



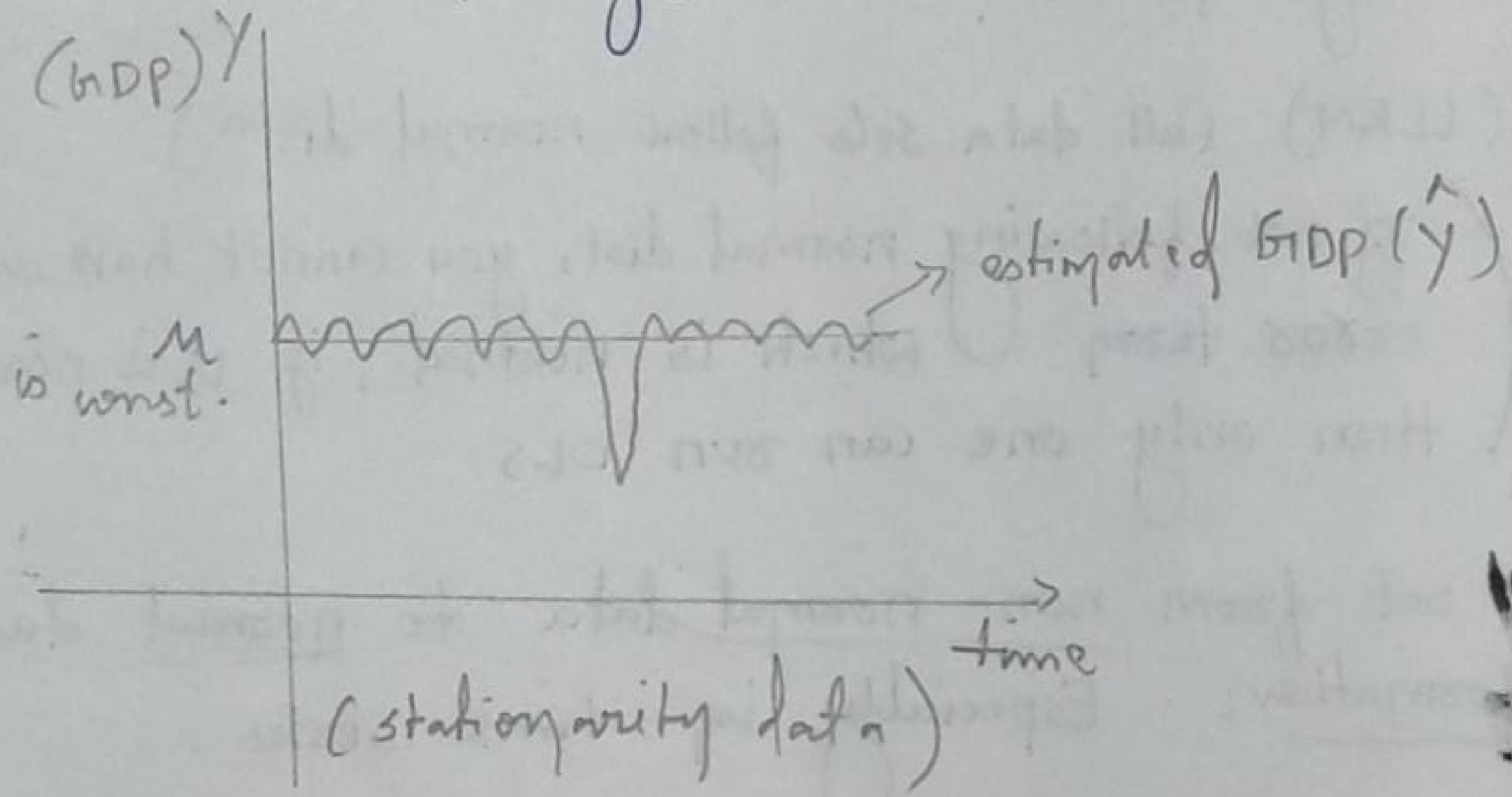
(non stationary data)



* if Pdf is constant, our data is in strong stationary state.
* We never find strong stationary state in time series

A random variable is a variable having a probability distributn attached to it.

My objective i.e, to see can I fit a trend line or forecast the GDP or not. To do that stationarity is important. (Weak stationarity)



1st conditn of stationarity

constant mean

i.e $E(Y_t) = \mu$

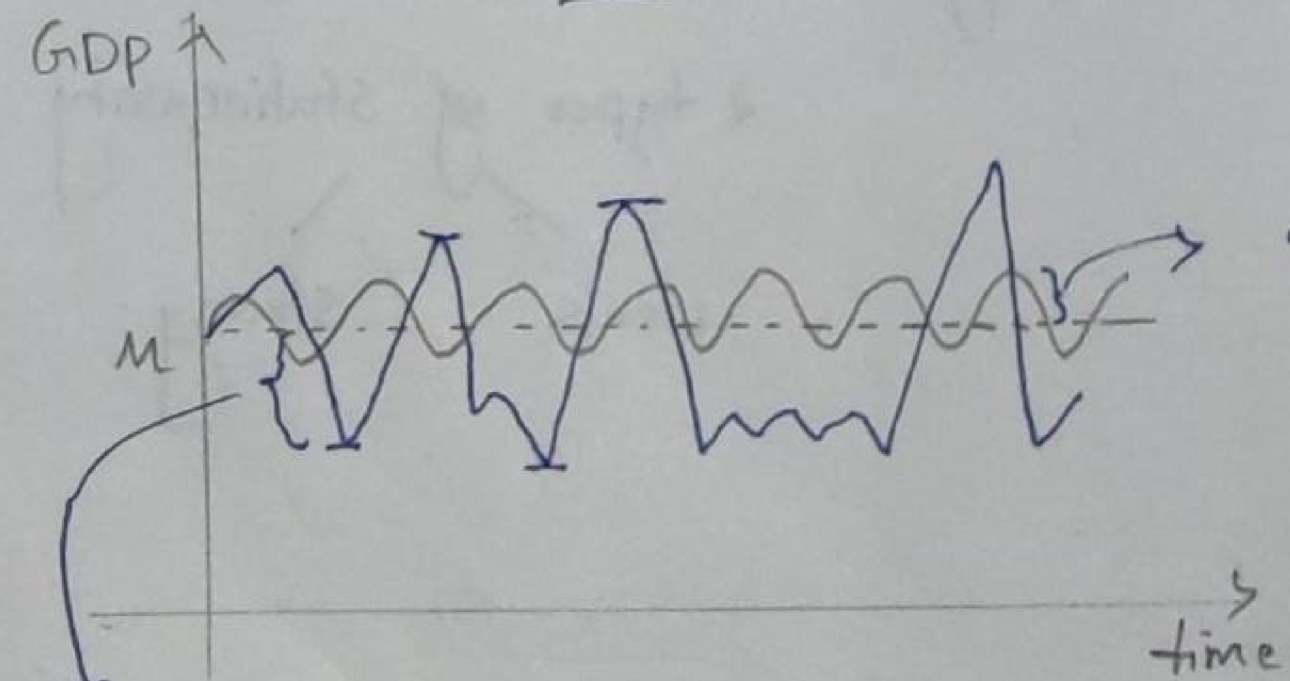
2nd conditn of stationarity

constant variance

$E(Y_t - \mu)^2 = \sigma^2 < \sigma^{2A}$

should also be minimum.

* **constant & non constant variance.**



deviatn from mean in grey line is constant = constant variance.

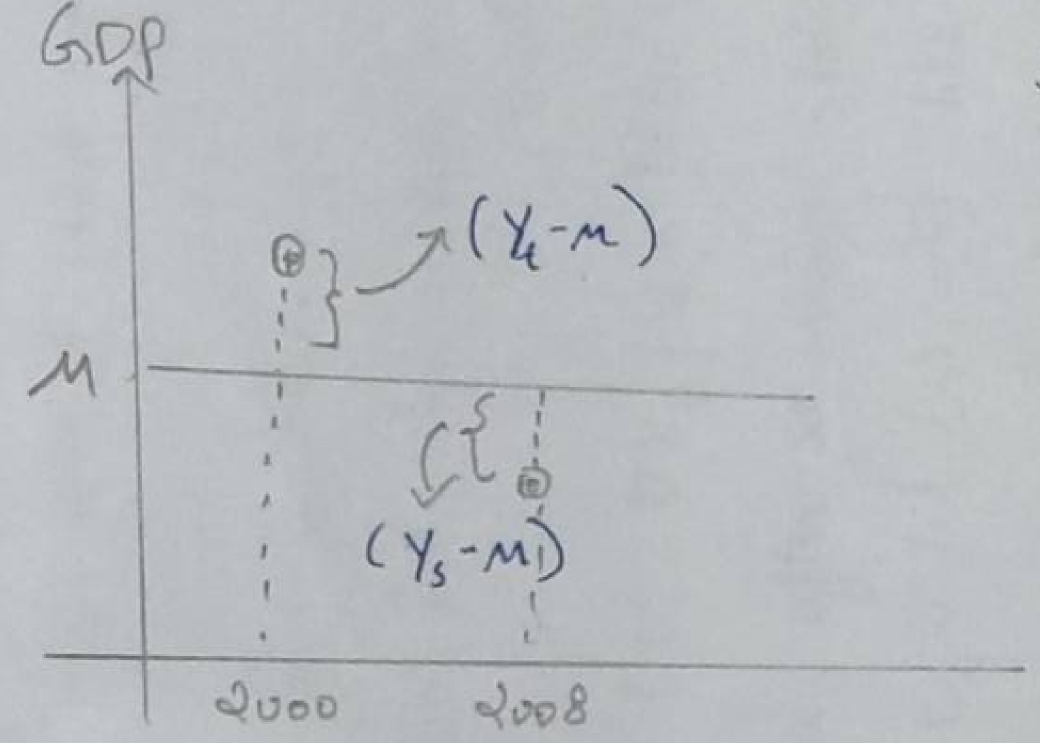
deviation from mean is purple line is always unequal = non constant variance.

* **3rd condition of stationarity**

3) constant - Autocovariance.

$$E\{(Y_t-\mu)(Y_s-\mu)\} \rightarrow \gamma \quad \forall\, t,s$$

$t = \underbrace{0, 1, 2, \ldots \text{'}T\text{'}}_{s}$



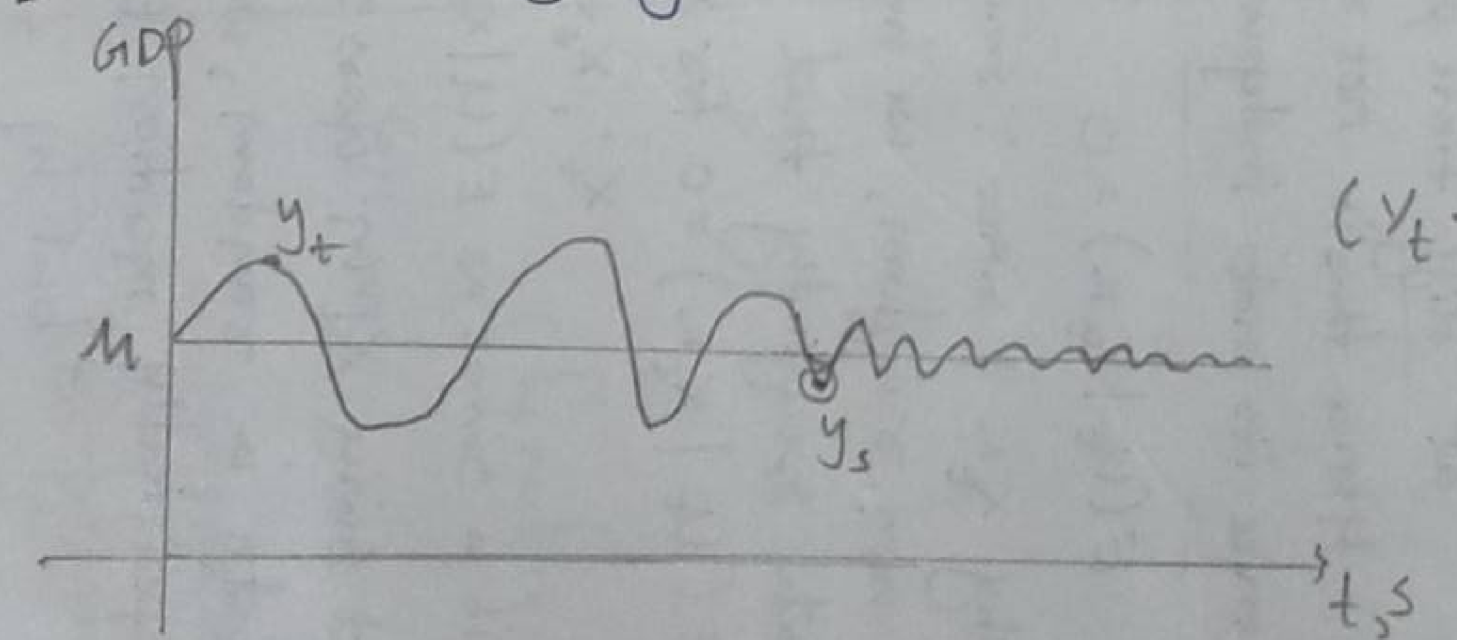
$(Y_t-\mu) = (Y_s-\mu)$ for autocov. to be const, to hold

$\forall\, t, s$ are time periods

& $t \neq s$

★ Violation of const. Auto covariance (2 types of Assimilation)

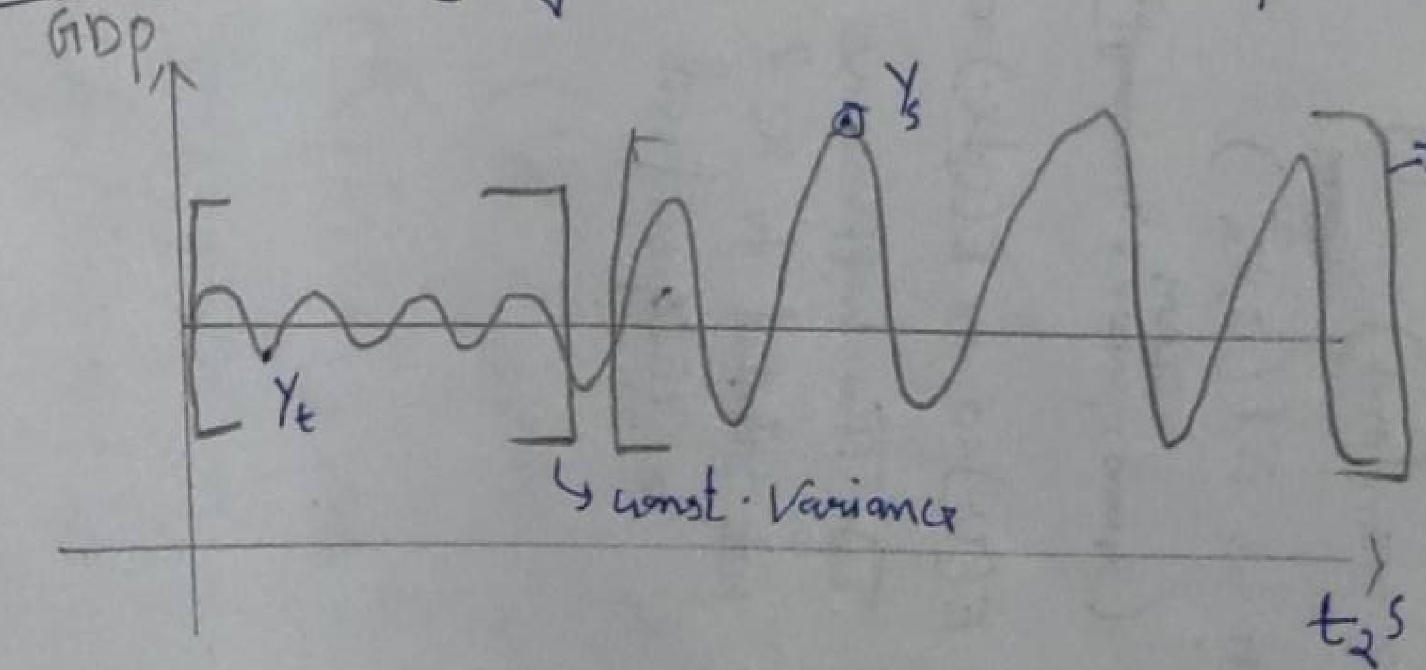
case 1 - converging data (stable equilibrium)



$(Y_t-\mu) \neq (Y_s-\mu)$

Violates Autocovariance (const)

case 2 - Diverging data (Unstable equilibrium)



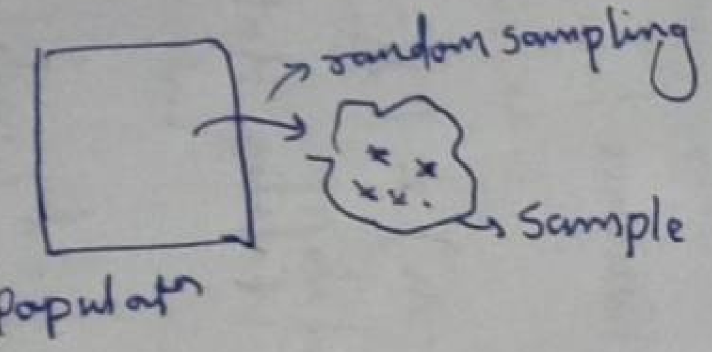
→ const. Variance

$Y_t-\mu) \neq (Y_s-\mu)$

but violates Autocov (constant)

Weak stationarity conditions - 1st, 2nd, 3rd condition

Cross sect (weak assumpt) | (Strict assumptn) Time Series

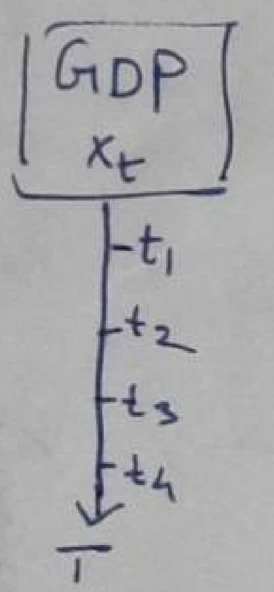


- The elements of Sample have eql probability (random)

- $U_i \rightsquigarrow iid(0, \sigma_u^2)$
  (Gauss markov 6th assumptn) CLRM

ii) $E(U_i)=0$ or $E(U_i|X_i)=0$

~~Tr~~ This happens as U_i not related to X_i & X_i is independent



As event like GDP is happening we are selecting the element of that event i.e., X_t which is happening at diff. time period

Hence this is not random hence its not independent

ii) $E(U_t | X_t k) = 0$

Why $X_t k$? ~~since~~ $\because$ since $X_t k$ is non random, we have to state specifically that $E(U_t | X_t k) = 0$ for $X_t k = X_t^1, X_t^2, \ldots X_t^k$

it is same as $E(U_i|X_i) = 0$

But since X_i in cross sectn data is random, we don't specifically mention $E(U_i|X_i) = 0$ but in time series we need to state $E(U_t | X_t^k) = 0$ $\forall X_t^k = X_t^1, X_t^2 \ldots X_t^k$.

Strict Exogenity (only for time series)

* Exogenity : $cov(X_i, U_i) = 0$ weak exogenity (why?)

1) $Y_i = \alpha + \beta X_i + U_i \rightsquigarrow$ cross sectn

2) $Y_t = \alpha + \beta X_t + U_t \rightsquigarrow$ time series

$cov(U_t, X_t k) = 0 \rightsquigarrow$ exogenity → Strict exogenity

t is time k variable.

Cross Section (Weak assumptn)

iii) $cov(U_i, U_j) = 0$ (autocorrelatn)

we don't have to mention

$cov(U_i, U_j | X_i) = 0$ as X_i's are iid.

iv) $cov(U_i, X_i) = 0$ (exogenity)

Time Series (strict assumptn)

iii) $cov(U_t, U_{t-1} | X_t k) = 0$, where $X_t k$ is not iid, hence has to be mentioned (strict autocorr).

iv) $cov(U_t, X_t | X_t k) = 0$ (strict exogenity)

v) Hence, after so many strict assumptn $U_t \sim iid(0, \sigma^2)$ like $U_i \sim iid(0, \sigma^2)$

Note: earlier it didnt follow iid but after so many assumptn, $U_t \sim iid(0, \sigma^2)$

Asymtotic Assumptions (when observatn $n \to \infty$)

under time series, as $n \to \infty$, the strict assumptn get transformed to weak assumptn called assymptotic assumptn.

i) $E(U_t) = 0$

ii) $V(U_t) = \sigma^2$

iii) $cov(U_t, U_{t-1}) = 0$ — Asymptotic assumptn

iv) $cov(U_t, X_t) = 0$

v) linear in parameter

vi)** Stationarity + Weak dependency

Note (imp)

If you have $n > 30$, then for a time series data you can apply the asymptotic assumptions, making it easier.

6th Assumptn

Stationarity & weak dependency

Stationarity

i) $E(X_t) = \mu$

ii) $V(X_t) = \sigma^2$

iii) $cov(X_t, X_{t-1})$ = constant.

Weak dependency

as $t \to \infty$

$corr(X_t, X_{t+h}) = 0$

where h is +ive + some yrs.

* where there is too much dependency betn the independent Variable or observations, then we can't use assymptotic assumptn

- **Autoregressiveness**

if a function is a function of its own lagged Variable i.e

$X_t = f(X_{t-1})$ — denoted as AR 1 (autoregressiveness of order 1)

$X_t = f(X_{t-1}, X_{t-2}) \rightarrow AR_2$ (autoregressiveness of order 2)

$\vdots$

like this AR_3, AR_4, AR_5 is also possible.

3 different models of AR

$$\left.\begin{aligned} X_t &= \delta_0 + \delta_1 X_{t-1} + \epsilon_t \quad -(\text{I}) \\ Y_t &= \gamma_0 + \gamma_t Y_{t-1} + V_t \quad -(\text{II}) \end{aligned}\right\} \text{2 AR models}$$

then

$Y_t = \beta_0 + \beta X_t + U_t \quad - (\text{III})$ where nature of Y_t & X_t are in eq^n (I) & (II).

However all the models are independent, eq^n I & II aren't incorporated in (III), we are just saying there are 3 models & describing nature of each. Eq^n (I) & (II) may / may not be incorporated.

Note

Y_t & X_t is stationarity & weak dependency.

Suppose

$Y_t = f(Y_{t-1}, X_t)$ then $Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_t + U_t$

$X_t = f(X_{t-1})$

Now if

In $\quad Y_t = \beta_0 + \underbrace{\boxed{\beta_1 Y_{t-1}} + \beta_2 X_t}_{} + U_t$

if Y_{t-1} is in U_t

then autocorrelat^n arises i.e

$cov(Y_t, Y_{t-1}) \neq 0$ as $Y_t = f(Y_{t-1})$

Similarly, if X_t is in U_t, $cov(U_t, X_t) \neq 0$ {so endogenity}

hence, problem of Autocorrel^n & endogenity leads to problem of heteroskedasticity, hence leading inefficient estimators.

To solve this problem, we extract Y_t or X_t into the model to avoid such problems.

ρ - coeff. of the AR
↳ measures reln betn current & previous values of series

☆ RANDOM - WALK { Special case of AR-I }

Suppose we have an AR(1) such that

when $\rho = 1$
↳ non stationary
(series will drift or have random walk)

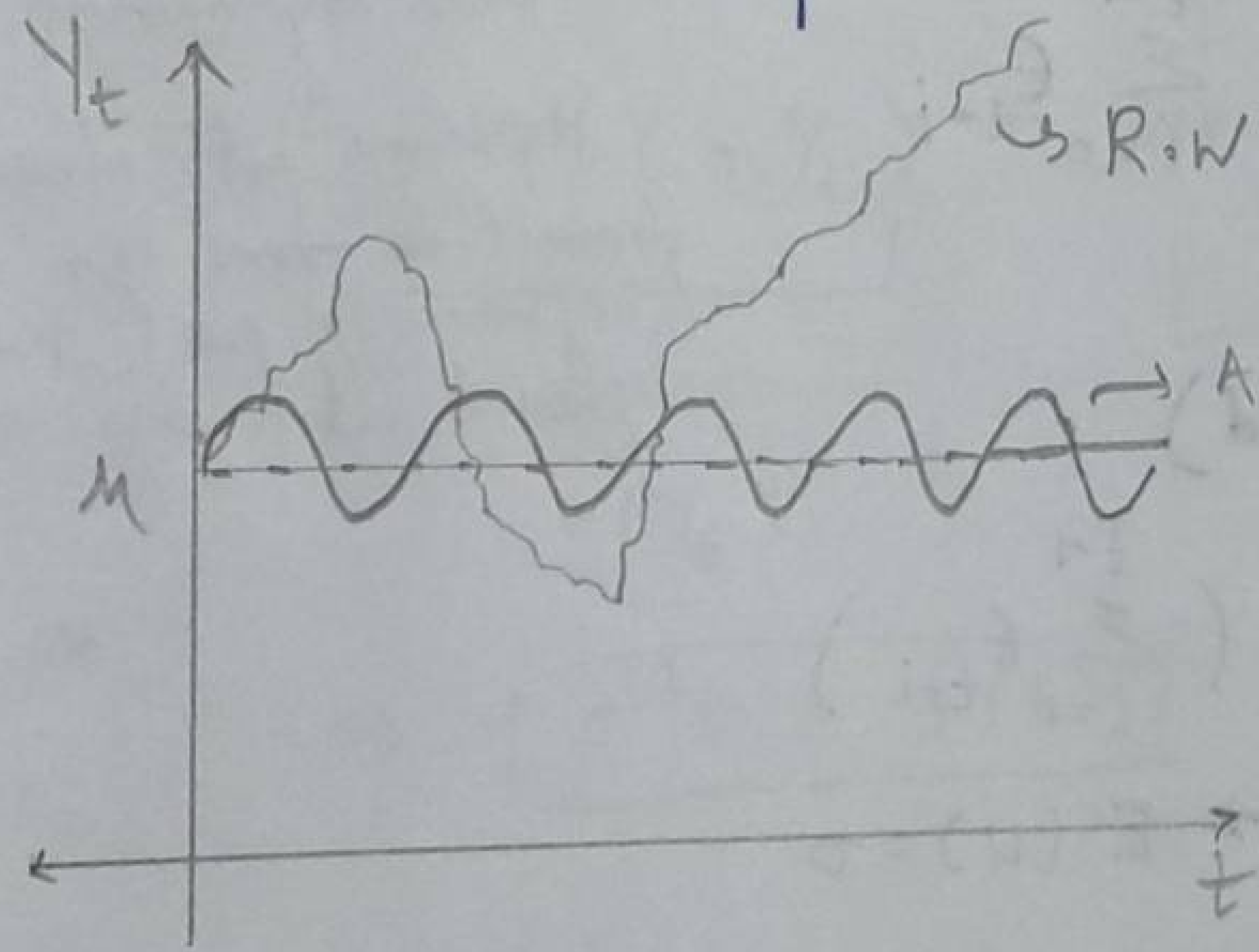
$$Y_t = \rho Y_{t-1} + \epsilon_t \quad \text{where} \quad \rho < 1$$

if $\rho = 1$ (unit root)

when $\rho > 1$ → series become explosive, small shock from mean will cause series diverge & grow without bound. Consecutive Model becomes unstable.

$$Y_t = Y_{t-1} + \epsilon_t \quad \text{—— RW}$$

Difference betn AR(1) - For $\rho < 1$ and RW is the downward force ρ on the independent variable Y_{t-1}



Now for more general form of Random walk

$$Y_t = Y_{t-1} + \epsilon_t$$

let $Y_{t-1} = f(Y_{t-2})$

and so on . . .

$$Y_t = Y_0 + \sum_{i=0}^{t-1} \epsilon_{t-i}$$

The problem with random walk is it don't follow Stationarity assumptn i.e → $E(Y_t) = \text{const}$

$V(Y_t) = \text{const}$

$\text{cov}(Y_t, Y_{t+h}) = \text{const.}$

No., if

and also violate weak dependency WD

i.e WD $\Rightarrow$ $corr(Y_t, Y_{t+h})$ as $h \to \infty$ = 0

Now for proving each of these assumptn for RW series we make 2 distinction.

R.W series i.e., $\rho = 1$ — RW with drift / RW without drift.

* Random walk (without) without drift violates 2 assumptions of stationarity & WD.

Weak dependency — dependence btn observatn decrease as time lag increasing (influence of past values over present, diminishes over time)

Given ; RW : $Y_t = Y_0 + \sum_{i=0}^{t-1} \epsilon_{t-i}$

Then

1) $\underline{E(Y_t)}$ (Not voilated)

$$E(Y_t) = E(Y_0) + E\left(\sum_{i=1}^{t-1} \epsilon_{t-i}\right)$$

$$\Rightarrow E(Y_0) = 0 + E(U) = 0$$

$$\Rightarrow 0$$

2) $\underline{V(Y_t)}$ (is voilated)

$$V(Y_t) = V(Y_0) + V\left(\sum_{i=0}^{t-1} \epsilon_{t-i}\right)$$

Now, $V(Y_0) = \{E(Y_0)^2\}$

$= E(Y_0^2) = 0$ {we assume}

$$\therefore V(Y_t) = 0 + V\left(\sum_{i=0}^{t-1} \epsilon_{t-i}\right)$$

Now $V(Y_t) = \sum_{i=0}^{t-1} V(\epsilon_{t-i})$

$$= \sum_{i=0}^{t-1} \sigma^2$$

to sim σ^2 for t times we can write

$$V(Y_t) = \sigma^2 t \quad = f(t)$$ { hence not const }

3) $cov(Y_t, Y_{t+h})$ (gets violated)

for $Y_{t+h} = Y_t + \sum_{i=1}^{h-1} \epsilon_{t+h-i}$

now

$$cov\left(Y_t, Y_t + \sum_{i=1}^{h-1} \epsilon_{t+h-i}\right)$$

$$\Rightarrow \underbrace{cov(Y_t, Y_t)}_{V(Y_t)} + \underbrace{cov\left(Y_t, \sum_{i=1}^{h-1} \epsilon_{t+h-i}\right)}_{0}$$ { no relation }

$$\Rightarrow \boxed{\sigma^2 t = f(t)}$$ { not constant }

- Weak Dependency (gets Violated)

$corr(Y_t, Y_{t+h})$ as $h \to \infty = ?$

$$\Rightarrow corr(Y_t, Y_{t+h}) = \frac{cov(Y_t, Y_{t+h})}{\sqrt{V(Y_t) V(Y_{t+h})}}$$

From (3), we know $cov(Y_t, Y_{t+h}) = \sigma^2 t$

$$corr(.) = \frac{\sigma^2 t}{\sqrt{V(Y_t) \cdot V(Y_{t+h})}}$$

$$V(Y_{t+h}) = V(Y_t) + V\left(\sum_{i=0}^{t-1} \epsilon_{t+h-i}\right)$$

$$= \sigma^2 t + \sigma^2 h$$

$$= \sigma^2 (t+h)$$

$$\therefore \text{corr}(\cdot) = \frac{\sigma^2 t}{\sqrt{\sigma^2 t \cdot \sigma^2 (t+h)}}$$

$$\Rightarrow \frac{t}{\sqrt{t(t+h)}} \quad (\sigma^2 \text{ cancel out})$$

as $h \to \infty$ $\quad$ corr $(\cdot) \to 0$

but ; here t is Variable (not fixed)

$\therefore$ if $h \to \infty$; t can $\to \infty$ as well

iff t is fixed ; as $h \to \infty$, corr $\to 0$

Note :- $\rho = 1$ is called Unit root

* Random Walk with drift Violates All 3 conditions of stationarity & WO.

if we add an intercept term to AR(1) for $\rho = 1$ then we get random walk with drift.

$$Y_t = \alpha + Y_{t-1} + \epsilon_t$$

as the same trend ; we expand Y_t over $t-1$ periods.

$\alpha \to$ is drift term, representing a systematic upward & downward trend
if $\alpha > 0$, upward drift, $\alpha < 0 \to$ downward drift

$$\Rightarrow Y_t = \alpha + \{\alpha + Y_{t-2} + \epsilon_{t-1}) + \epsilon_t \quad \{\text{where } Y_{t-1} = \alpha + Y_{t-2} + \epsilon_{t-1}\}$$

$$Y_t = \{\alpha + \alpha \ldots t \text{ times}\} + Y_0 + \sum_{i=0}^{t-1} \epsilon_{t-i}$$

$$Y_t = \alpha t + Y_0 + \sum_{i=0}^{t-1} \epsilon_{t-i}$$

1) $E(Y_t)$ (gets Violated)

$$E(Y_t) = E(\alpha t) + E(Y_0) + E\left(\sum_{i=0}^{t-1} \epsilon_{t-i}\right)$$

$$= \alpha t + 0 + 0 \qquad \{E(Y_0) = 0,\ E(\Sigma U) = 0\}$$

$$= \alpha t = f(t) \neq (\text{const}).$$

2) $V(Y_t)$ (gets Violated)

$$\Rightarrow V(Y_t) = \underbrace{V(\alpha t)}_{0} + \underbrace{V(Y_0)}_{0} + \underbrace{V\left(\sum_{i=0}^{t-1} \epsilon_{t-i}\right)}_{\sigma^2 t}$$

3) $cov(Y_t, Y_{t+h})$ (Violated)

$$Y_{t+h} = \alpha + Y_{t+h-1} + \epsilon_{t+h}$$

$$\vdots$$

$$Y_{t+h} = \alpha(t+h) + Y_t + \sum_{i=0}^{h-1} \epsilon_{t+h-i}$$

$$\Rightarrow cov\left(Y_t, \alpha(t+h) + Y_t + \sum_{i=0}^{h-1} \epsilon_{t+h-i}\right)$$

$$\Rightarrow cov(Y_t, Y_t) + \underbrace{cov(Y_t, \alpha(t+h))}_{0} + \underbrace{cov\left(Y_t, \sum_{i=0}^{h-1} \epsilon_{t+h-i}\right)}_{0}$$

$$\Rightarrow V(Y_t) + 0 + 0$$

$$= \sigma^2 t$$

Weak Dependency. (Voilated)

$$\text{corr}(Y_t, Y_{t+h}) = ??$$

as $h \to \infty$

$$\Rightarrow \text{corr}(Y_t, Y_{t+h}) = \frac{\text{cov}(Y_t, Y_{t+h})}{\sqrt{V(Y_t)\cdot V(Y_{t+h})}} = \frac{\sigma^2 t}{\sqrt{V(t)\cdot V(Y_{t+h})}}$$

$$V(Y_{t+h}) = V(\underset{0}{\underset{\shortparallel}{\alpha(t+h)}} + V(Y_t) + V(\sum_{i=0}^{h-1} \epsilon_{t+h-i})$$

$$= \sigma^2 t + \sigma^2 h$$

$$= \sigma^2 (t+h)$$

$$\text{corr}(Y_t, Y_{t+h}) = \frac{\sigma^2 t}{\sqrt{\sigma^2 + \sigma^2(t+h)}} = \frac{t}{\sqrt{t(t+h)}}$$

as $h \to \infty$ $\quad$ $\text{corr} \to 0$

but t is Variable; hence $\text{corr}(Y_t, Y_{t+h}) \neq 0$.

* Dicky fuller & Augmented Dicky fuller Test (Stationarity test)

Dicky fuller test (DFT) $\longrightarrow$ Used for AR (1) Series.

Augmented Dicky fuller test $\rightarrow$ used for AR $(2, 3, 4, \ldots \infty)$.
(ADFT)

DFT :

The AR(1) Series such as

$$Y_t = \rho Y_{t-1} + \epsilon_t$$

Now to test whether $\rho < 1$ or $\rho = 1$, we run it through a DFT.

To do DFT, we need to make an algebric adjustment for running t-test.

i.e., $Y_t = \rho Y_{t-1} + \epsilon_t$

$\Rightarrow Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + \epsilon_t$

$\Rightarrow \Delta Y_t = (\rho - 1) Y_{t-1} + \epsilon_t$

Now,

$\Delta Y_t = (\rho - 1) Y_{t-1} + \epsilon_t$ — AR (1)

To test; we use t-test

i.e let $B = \rho - 1$

$\therefore$ if $B = 0 \Rightarrow \rho - 1 = 0 \Rightarrow \rho = 1$

$B < 0 \Rightarrow \rho - 1 < 0 \Rightarrow \rho < 1$

i.e., $H_0 : B = 0$ → has unit root hence non stat

$H_1 : B < 0$

if $t_B >$ t table value reject null hypothesis

Hence $B < 0$, stationary.

$$t_\beta = \frac{\hat{\beta}}{S.E(\hat{\beta})} \quad \text{where} \quad SE(\hat{\beta}) = \sqrt{V(\hat{\beta})} = \sqrt{\frac{\sigma^2}{\Sigma x_i}}$$

std error ← $S.E(\hat{\beta})$

ADFT :

Suppose we have an AR (2) series

i.e., $Y_t = f(Y_{t-1}, Y_{t-2})$ → AR(2) or AR(∞); we can't dicky fuller test here in this.

Suppose

$$\Delta Y_t = (\rho - 1) Y_{t-1} + \delta \Delta Y_{t-1} + \epsilon_t$$

→ $(Y_{t-1} - Y_{t-2})$

Now suppose we expand $\delta \Delta Y_{t-i}$ further to incorporate other independent variables to make AR (3) or (4)... Series

$$\Rightarrow \Delta Y_t = (\rho - 1) Y_{t-1} + \sum_{i=1}^{h} \delta^i \Delta Y_{t-i} + \epsilon_t$$

Now,

To do dicky fuller we run the adf test

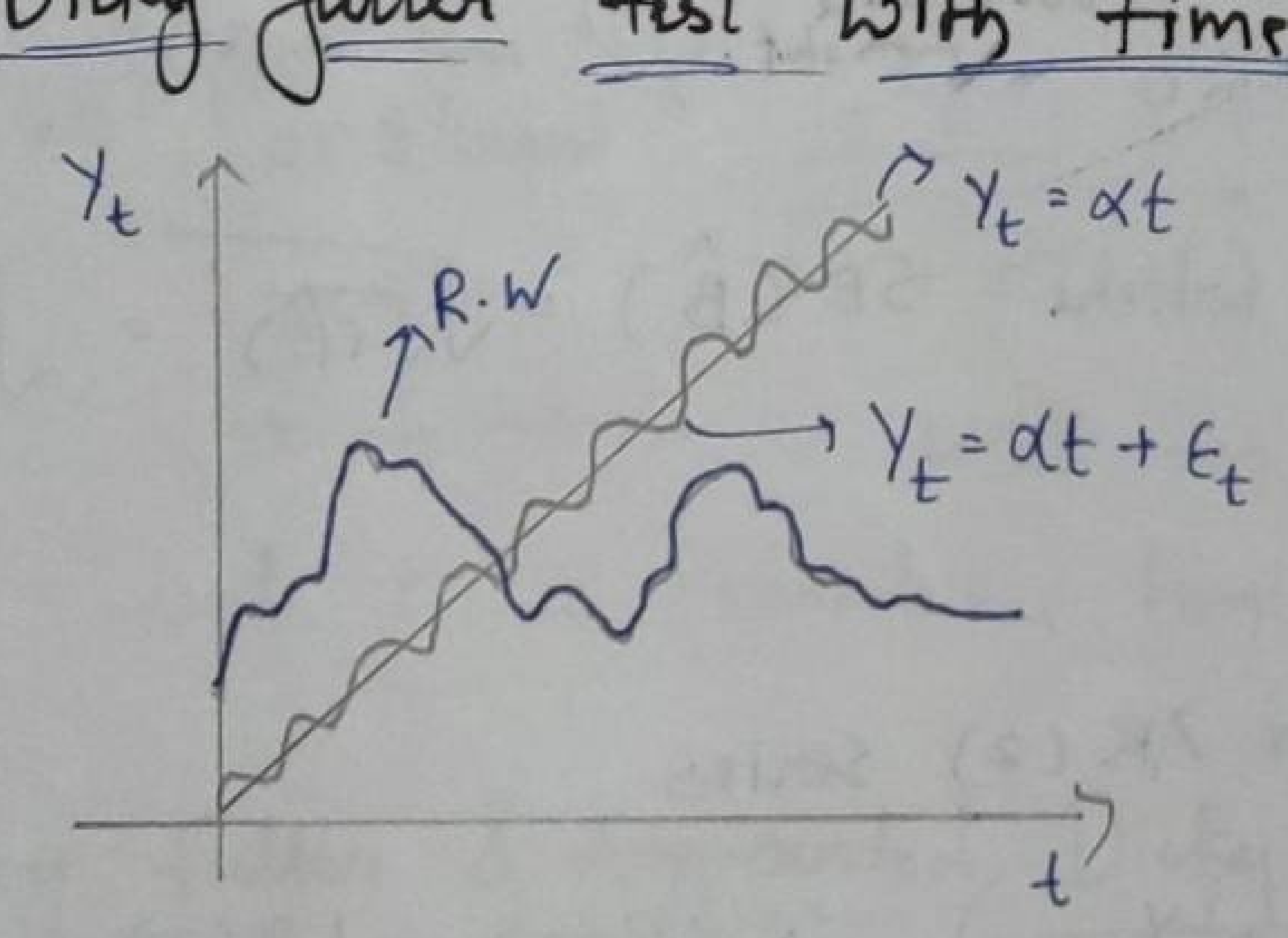
or $H_0 : \beta = 0$ and also $H_0 : \delta = 0$

$H_2 : \beta < 0$

$H_0 : \delta < 0$

Thus the adjustment under ADF is to include $\underline{\sum_{i=0}^{h} \delta^i \Delta Y_i}$ into the AR series.

* <u>Dicky fuller test with time trend</u>. (important) (not complete)



$Y_t = \alpha t$ (mathematical model)

$Y_t = \alpha t + \epsilon_t$ (econometric model)

Now,
$Y_t = \alpha t + \epsilon_t$ is deterministic model even though ϵ_t is stochastic.

To be deterministic, it has to satisfy the 3 stationarity condition. & W.D.

i) $E(Y_t) = E(\alpha t) + E(\epsilon_t)$

$= \alpha t + 0$

$\hookrightarrow$ not const. as t is Variable.

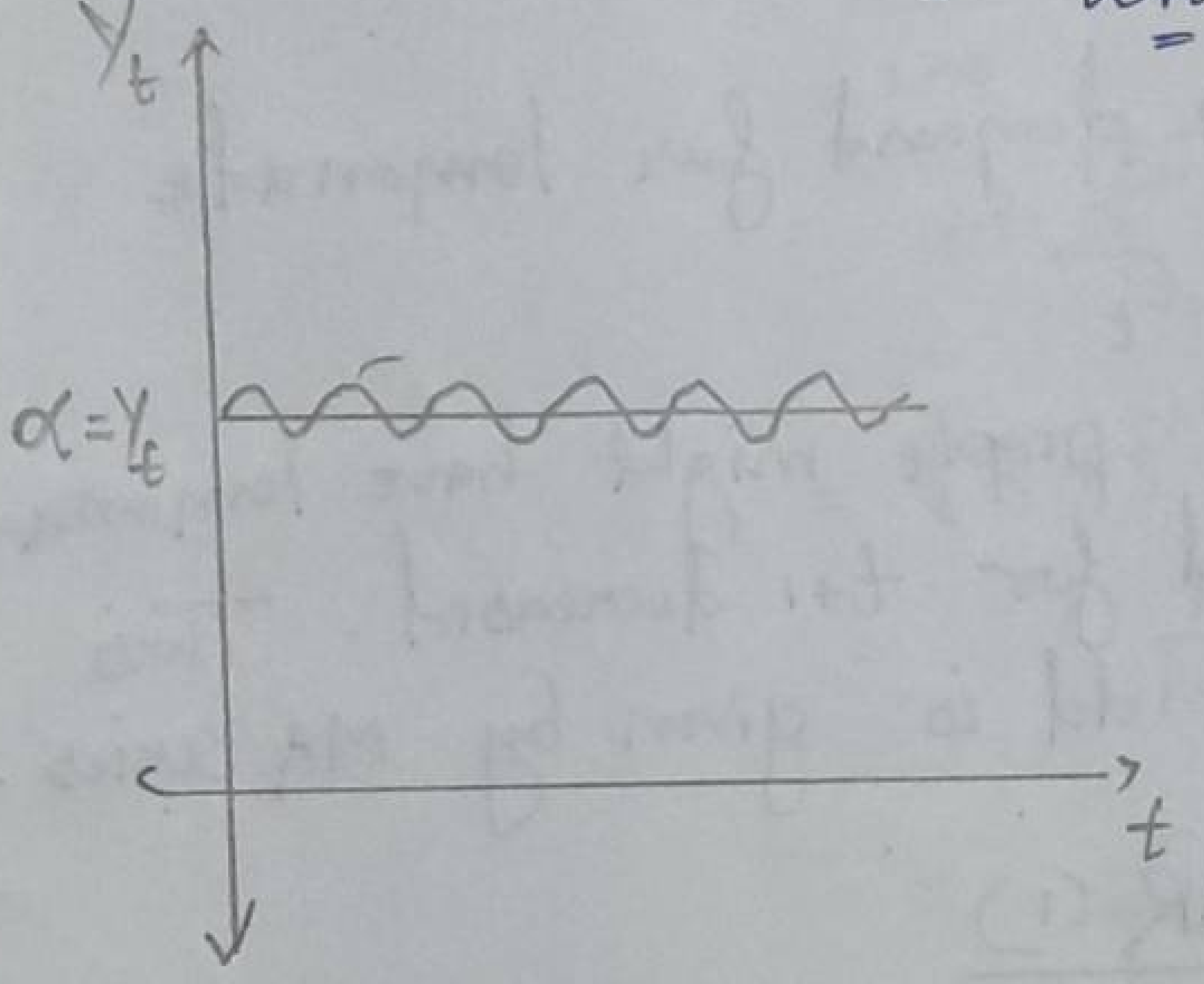
ii) $V(Y_t) = V(\alpha t) + V(\epsilon_t)$

$\Downarrow$ 0 $\quad$ $\Downarrow$ σ^2 (const.)

iii) $\text{cov}(Y_t, Y_{t+h}) = \text{cov}(Y_t \cdot \alpha t + \epsilon_t, \alpha(t+h) + \epsilon_{t+h})$

$$= \underbrace{\text{cov}(\alpha t, \alpha(t+h))}_{=0} + \underbrace{\text{cov}(\epsilon_t, \epsilon_{t+h})}_{=0}$$

$$= \underline{\text{const}}$$



However

For a model like

$$Y_t = \alpha + \epsilon_t \quad -(i)$$

$$\& \quad Y_t = \alpha \quad -(ii)$$

eqn (i) satisfies all 3 conditions of stationarity.

Now the reason that we are learning time trend is that; if time trend is added to a non-stationarity series, it helps to control its Random walk.

★ <u>Moving</u> <u>Average</u> <u>Process</u> (account for short term process

$$X_t \text{ or } Y_t = \epsilon_t + \theta \underbrace{\epsilon_{t-1}}_{\text{lagged Error}} \quad ; \epsilon_t \sim iid(0, \sigma^2) - MA-(1)$$

eg of MA (1)

change in demand of Lemonade.

$$\Delta \text{ Lemonade}_t = \epsilon_t - 0.5\, \epsilon_{t-1}$$

let ϵ_t represent Δ temp.

$\therefore$ if $\epsilon_t > 0$, then Δ lemonade$_t$ $\uparrow$.

Now suppose there is no change in temp. in t+1 i.e.

$$\Delta \text{ lemonade}_{t+1} = \underset{\overset{\|}{0}}{\epsilon_{t+1}} - 0.5\,\epsilon_t$$

$$= -0.5\,\epsilon_t$$

What does this say? i.e demand for lemonade as time t+1, ↓ due to $-0.5\,\epsilon_t$.

This can be explained that people might have lemonade left from time t, thus dd for t+1 decreased. This sudden & short term change in dd is given by MA series.

<u>Checking Stationarity & WD for AR (1)</u>

AR (1)

$$Y_t = \rho Y_{t-1} + \epsilon_t$$

now for a general form

let $y_{t-1} = f(Y_{t-2})$

$$\vdots$$

$$Y_t = \rho\,(\rho Y_{t-2} + \epsilon_{t-1}) + \epsilon_t$$

$$Y_t = \rho^2 Y_{t-2} + \rho^1 \epsilon_{t-1} + \rho^0 \epsilon_t$$

$$\boxed{Y_t = \rho^t Y_0 + \sum_{i=0}^{t-1} \rho^i \epsilon_{t-i}} \quad \text{— (1)}$$

•) $\underline{E(Y_t)} = E(\rho^t Y_0) + E\left(\sum_{i=0}^{t-1} \rho^i \epsilon_{t-1}\right)$

$$= \rho^t \underset{\overset{\|}{0}}{E(Y_0)} + 0$$

$$= 0 \quad \text{(constant)}$$

2) $V(Y_t) = V(\rho^t Y_0) + V\left(\sum_{i=0}^{t-1} \rho^i \epsilon_{t-i}\right)$

$$= \overset{\parallel}{0} + \sum_{i=0}^{t-1} \rho^i V(\epsilon_t - i)$$

$$= \sum_{i=0}^{t-1} \rho^i \sigma^2$$

$$= \sigma^2 \sum_{i=0}^{t-1} \rho^i$$

$$= (\rho^0 + \rho^1 + \rho^2 \dots \infty)\sigma^2$$

$$= \left(\frac{1}{1-\rho}\right)\sigma^2$$

$$= \frac{\sigma^2}{1-\rho} \begin{cases} \xrightarrow{\rho=1} \infty \\ \xrightarrow{\rho<1} \text{const.} \end{cases}$$

3) $\operatorname{cov}(Y_t, Y_{t+h})$

$$Y_{t+h} = \rho Y_{t+h-1} + \epsilon_{t+h}$$

$$\vdots$$

$$Y_{t+h} = \rho^h Y_t + \sum_{i=0}^{h-1} \rho^i \epsilon_{t+h-i}$$

$$\operatorname{cov}\left(Y_t, \rho^h Y_t + \sum_{i=0}^{h-1} \rho^i \epsilon_{t+h-i}\right)$$

$$= \operatorname{cov}(Y_t, \rho^h Y_t) + \operatorname{cov}\left(Y_t, \sum_{i=0}^{h-1} \rho^i \epsilon_{t+h-i}\right)$$

$$= \rho^h \frac{\sigma^2}{1-\rho} + 0.$$

$$\operatorname{cov}(Y_t, Y_{t+h}) = \frac{\rho^h \sigma^2}{1-\rho} \begin{cases} \xrightarrow{h\to\infty,\ \rho=1} \text{not const.} \\ \xrightarrow{h\to\infty,\ \rho<1} \text{const.} \end{cases}$$

Weak Dependency

$$\text{corr}(Y_t, Y_{t+h}) = \frac{\text{cov}(Y_t, Y_{t+h})}{\sqrt{V(Y_t)\,V(Y_{t+h})}}$$

as $h \to \infty$

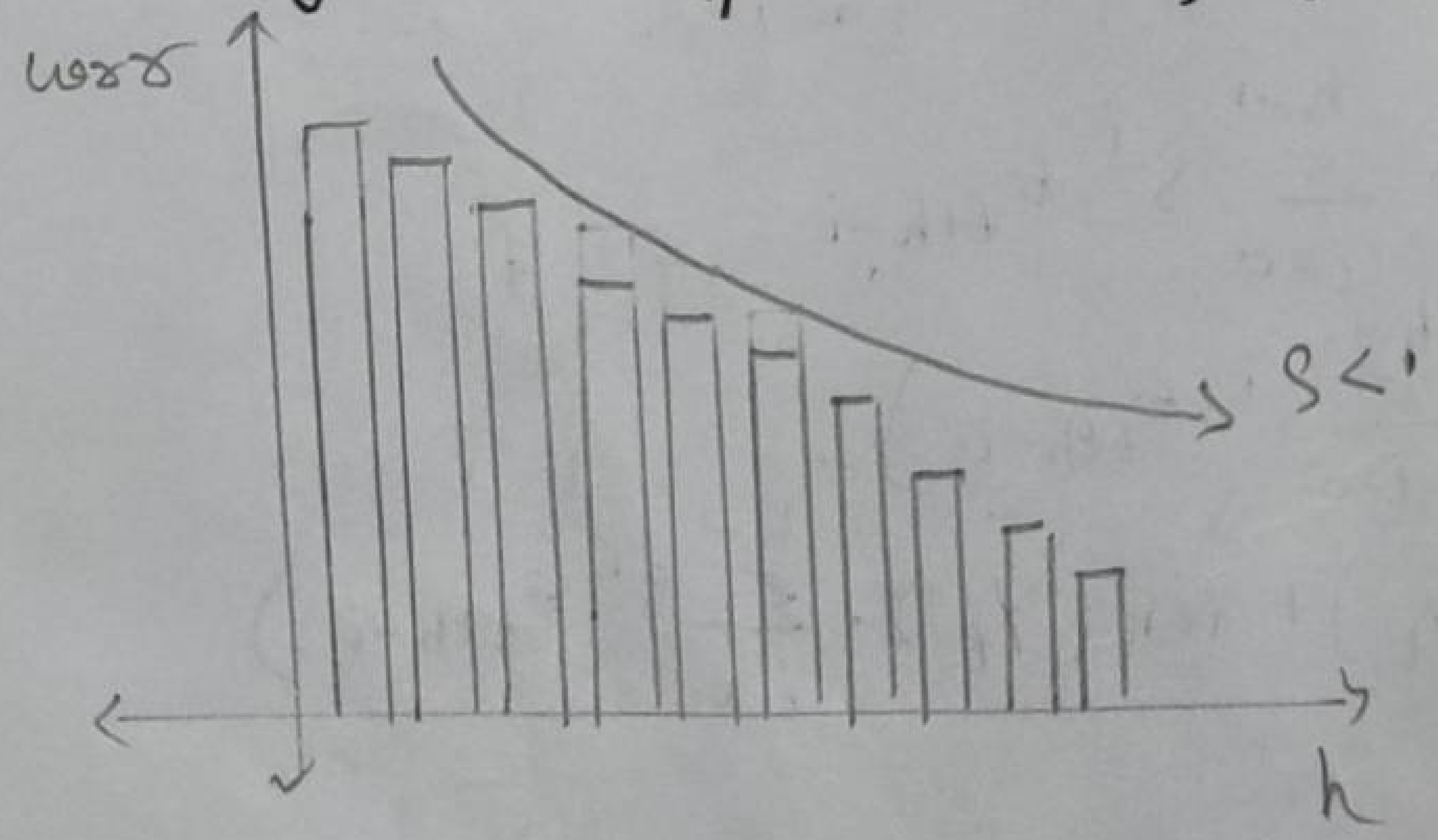
$$\Rightarrow \frac{\rho^h \frac{\sigma^2}{1-\rho}}{\sqrt{\frac{\sigma^2}{1-\rho}\,\underbrace{V(Y_{t+h})}_{}}} \Rightarrow \frac{\rho^h \frac{\sigma^2}{1-\rho}}{\frac{\sigma^2}{1-\rho}} = \rho^h$$

↳ we assume $Y_t = Y_{t+h}$

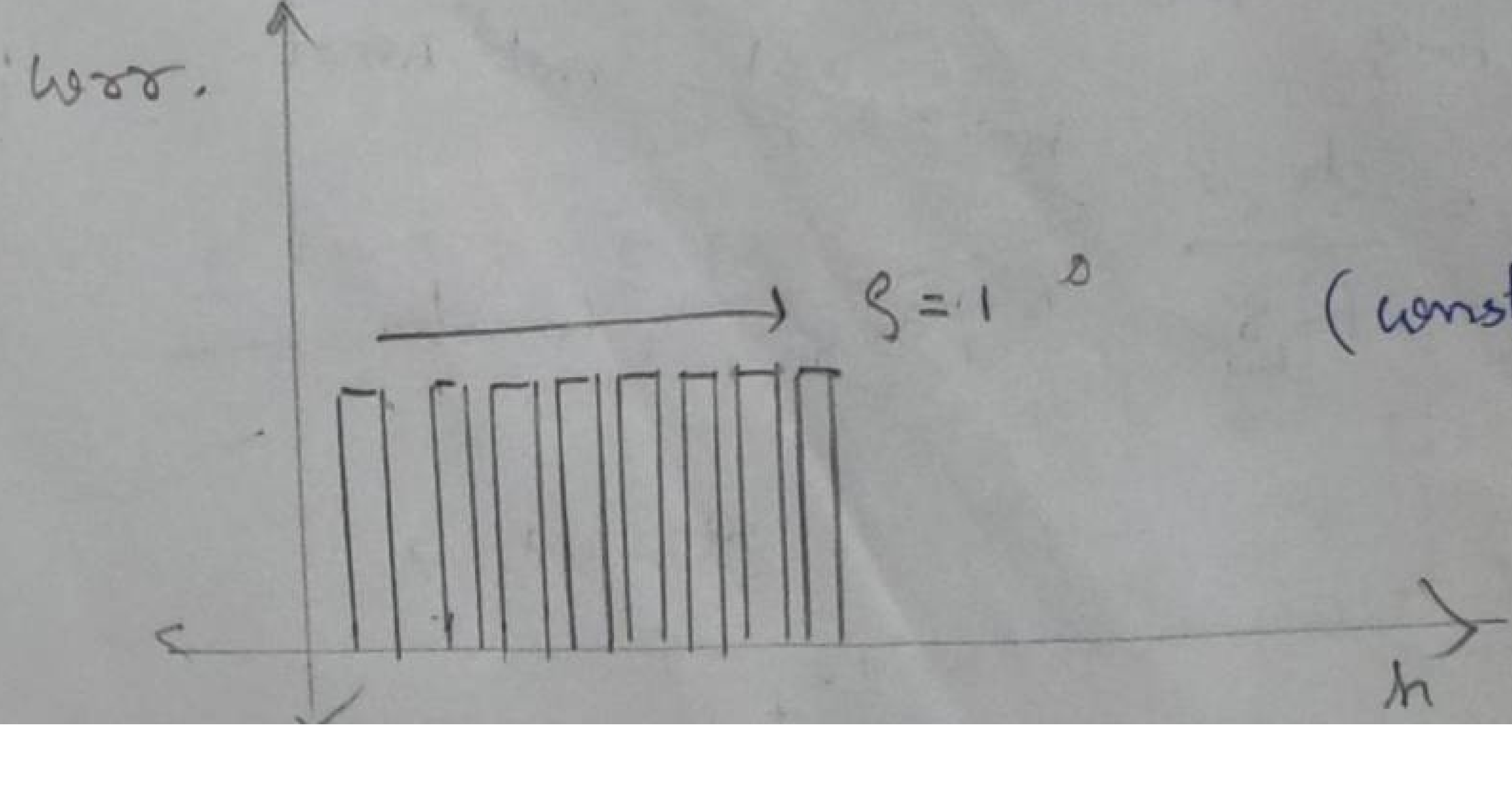
$\therefore V(Y_t) = V(Y_{t+h})$

as $h \to \infty$
- $\rho = 1$ —— $= 1$
- $\rho > 1$ —— $= \infty$
- $\rho < 1$ —— $= 0$

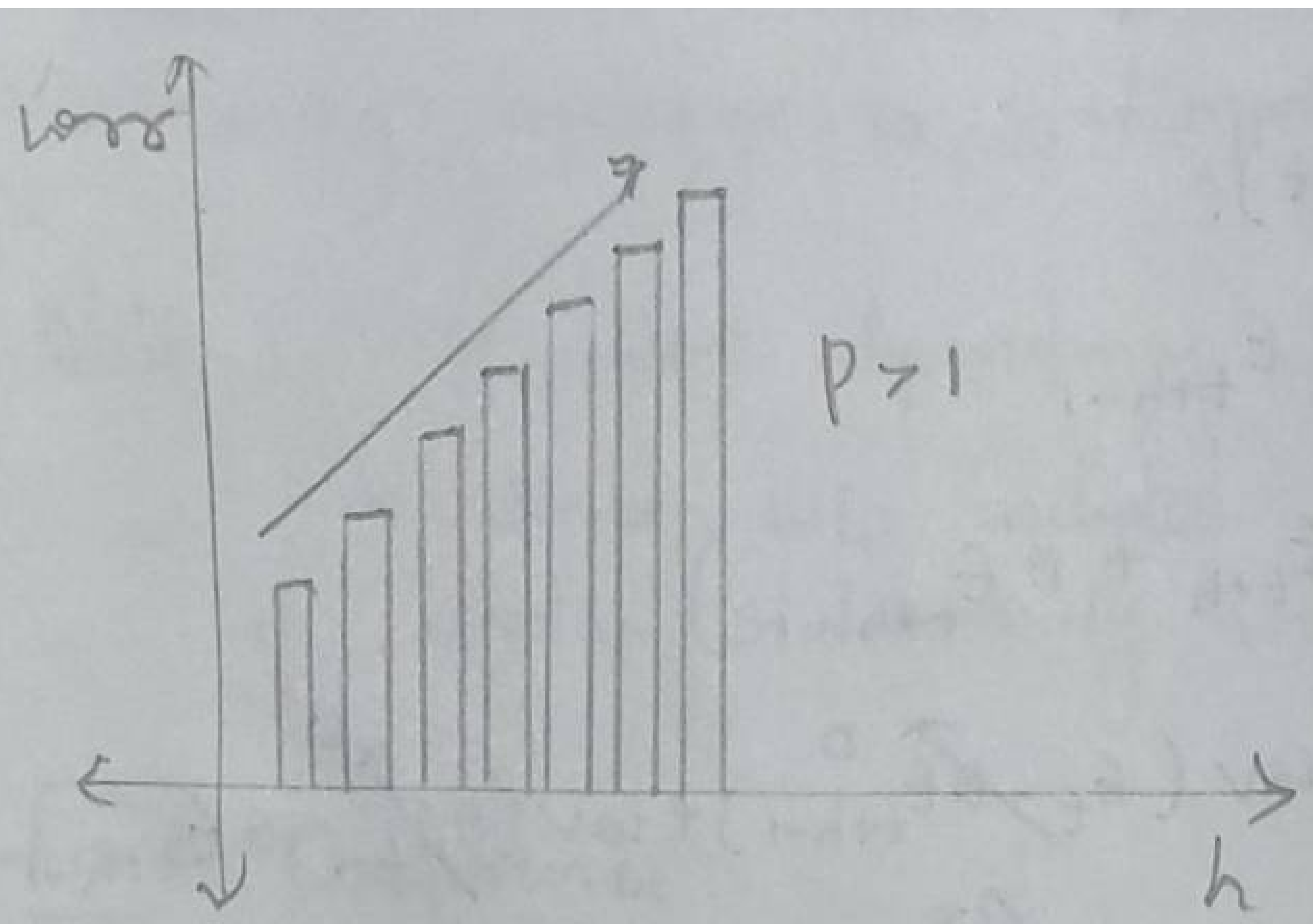
Correlogram Representation (AR-1)



Stationarity as corr $\to 0$



(const. but not corr $\to 0$)
Violates W.D



$P > 1$ corr $\to \infty$ (Violates W.D).

Moving Average : Stationarity & WD

$$Y_t = \epsilon_t + \theta \epsilon_{t-1}$$

1) $E(Y_t) = E(\epsilon_t) + \theta E(\epsilon_{t-1}) = 0$ (const)

$$E(\epsilon_t) = 0, \quad E(\epsilon_{t-1}) = 0$$

2) $V(Y_t) = V(\epsilon_t) + \theta V(\epsilon_{t-1})$

$= \sigma^2 + \theta \sigma^2$

$= \sigma^2(1+\theta)$ const.

3) $cov(Y_t, Y_{t-1}) = cov(\epsilon_t + \theta\epsilon_{t-1}, \epsilon_{t-1} + \theta \epsilon_{t-2})$

for $h = 1$

$\Rightarrow \overset{\approx 0}{cov(\epsilon_t, \epsilon_{t-1})} + \overset{\approx 0}{cov(\epsilon_t, \theta \epsilon_{t-2})} + cov(\theta\epsilon_{t-1}, \epsilon_{t-1})$

$+ \overset{\approx 0}{cov(\theta\epsilon_{t-1}, \theta\epsilon_{t-2})}$

assuming no autocorrelation.

$cov(Y_t, Y_{t-1}) = cov(\theta \epsilon_{t-1}, \epsilon_{t-1})$

$= \theta \sigma^2$ (const).

Now $\operatorname{cov}(Y_t, Y_{t+h})$ (const).

$$Y_{t+h} = \epsilon_{t+h} + \theta\,\epsilon_{t+h-1}$$

$$\Rightarrow \operatorname{cov}(\epsilon_t + \theta\,\epsilon_{t-1},\ \epsilon_{t+h} + \theta\,\epsilon_{t+h-1})$$

$$\Rightarrow \operatorname{cov}(\epsilon_t, \cancel{\epsilon_{t+h}})^{\to 0} + \operatorname{cov}(\epsilon_t, \cancel{\theta\epsilon_{t+h-1}})^{\to 0} + \operatorname{cov}(\theta \cancel{\epsilon_{t-1}}, \theta\,\epsilon_{t+h-1})^{\to 0}$$

$$+ \operatorname{cov}(\theta\,\epsilon_{t-1}, \theta \cancel{\epsilon_{t+h-1}})^{\to 0} = 0$$

$$\therefore \operatorname{cov}(Y_t, Y_{t+h}) \begin{cases} \xrightarrow{h=1} \theta\sigma^2 \text{ (const)} \\ \xrightarrow{h>1 \text{ or } h\to\infty} 0 \text{ (const)} \end{cases}$$

So the <u>moving average</u> <u>satisfies</u> all the <u>three conditions</u> of stationarity.

- <u>Weak Dependency</u>

$$\operatorname{corr}(Y_t, Y_{t+h}) = \frac{\operatorname{cov}(Y_t + Y_{t+h})}{\sqrt{V(Y_t)\,V(Y_{t+h})}}$$

for $h = 1$

$$\operatorname{corr} = \frac{\theta\sigma^2}{\sigma^2(1+\theta)} \quad \text{since } V(Y_t) = V(Y_{t+h})$$

$$= \frac{\theta}{1+\theta}$$

for $h > 1$ or $h \to \infty$

$$\operatorname{corr} = 0$$

$$\therefore \operatorname{corr}(Y_t, Y_{t+h}) \begin{cases} \xrightarrow{h=1} \theta/_{1+\theta} \text{ Voilates} \\ \xrightarrow{h>1} 0 \text{ satisfies} \end{cases}$$

Moving average is stationary & a short-term series.

Note : If weak dependency is Voilated & stationarity satisfy
→ Since WD includes 3^{rd} & 2^{nd} property of stationarity
it implies stationarity in some way is also being Voilated

first Difference.

we need to transform a non stationary series into stationary Hence we use the first difference.

$$Y_t = Y_{t-1} + \epsilon_t \longrightarrow \underline{N.S}$$

$$Y_t - Y_{t-1} = \epsilon_t$$

$$\Rightarrow \Delta Y_t = \epsilon_t$$

stationarity check.

1) $E(\Delta Y_t) = 0$ (const.)

2) $V(\Delta Y_t) = V(\epsilon_t) = \sigma^2$

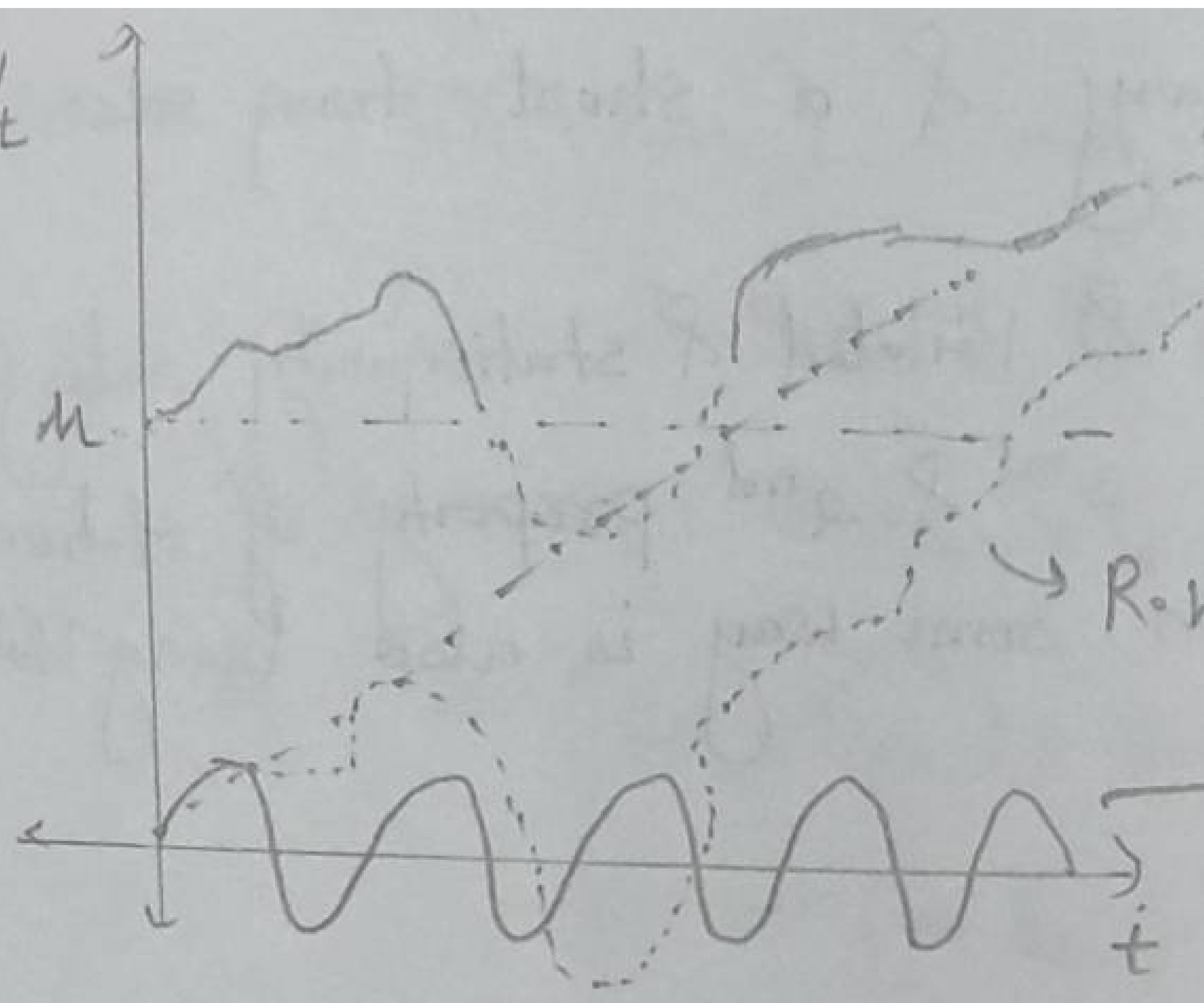
3) $\boxed{cov(\Delta Y_t, \Delta Y_{t+h}) = cov(\epsilon_t, \epsilon_{t+h}) = 0}$ → **imp**

no autocorr.

Weak dependency

$$\lim_{h \to \infty} corr(\Delta Y_t, \Delta Y_{t+h}) = \frac{cov(\Delta Y_t, \Delta Y_{t+h})}{\sqrt{V(\Delta Y_t)\, V(\Delta Y_{t+h})}}$$

$$= 0$$

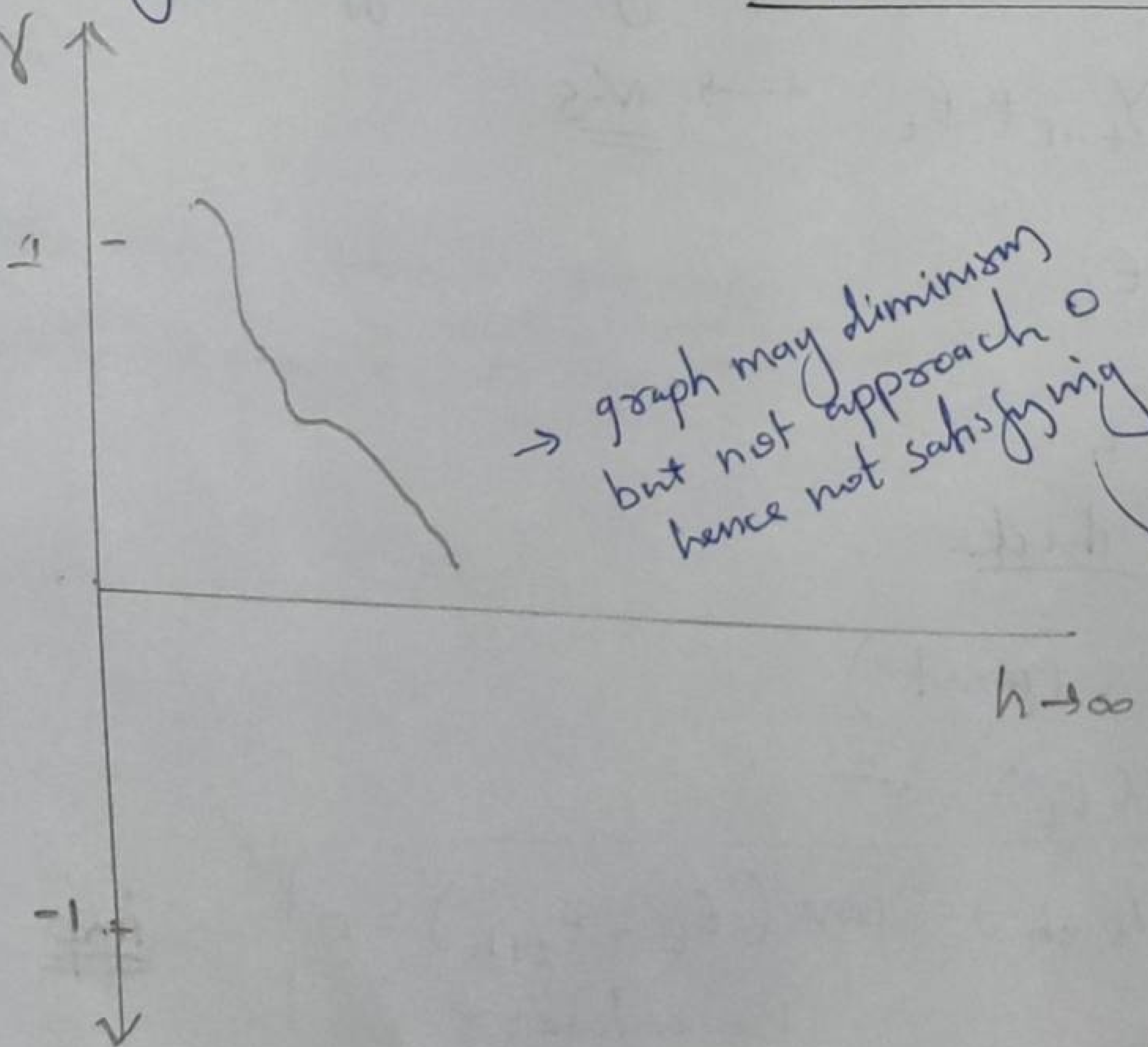


R.W (with drift)
↳ $y_t = \alpha + Y_{t-1} + \epsilon_t$

R.W (without drift) > $y_t = y_{t-1} + \epsilon_t$

first diff. for

Graph of correlationship with / without draft for $\boxed{R.W}$



→ graph may diminish but not approach 0 hence not satisfying

$\{-1 \leq \sigma \leq 1\}$

Since $corr(Y_t, Y_{t+h}) \neq 0$ as $\to \infty$

↓

It happen because $cov(Y_t, Y_{t+h}) \neq 0$

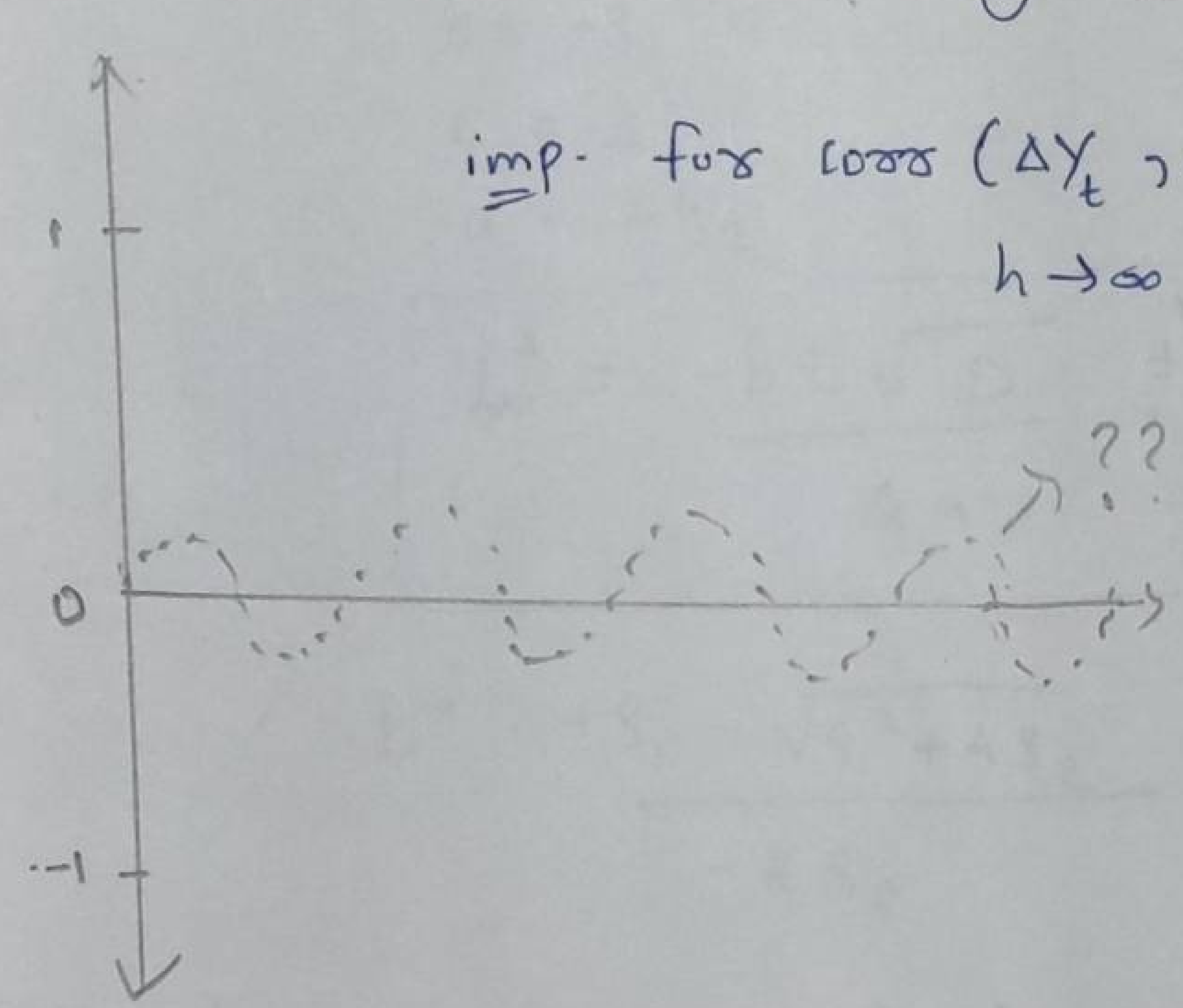
& this happens because

lets say R.W without drift

$cov\left(Y_t, Y_t + \sum_{i=0}^{h-1} \epsilon_{t+h-i}\right)$ or

$corr(\epsilon_t, \epsilon_{t+h}) \neq 0$, $cov(Y_t, Y_t) = \sigma^2 t$.

Graphical Representation of correlationship for 1st difference



imp. for $\underset{h \to \infty}{corr}(\Delta Y_t, \Delta Y_{t+h}) = \underset{h \to \infty}{corr}(\epsilon_t, \epsilon_{t+h}) = 0$

$(-1 \leq r \leq 1)$

Like we take 1st difference for AR(1), we take 2nd difference for AR(2)

i.e.,

$$Y_t = \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \epsilon_t \text{ ——— AR(2).}$$

We denote the differenced series by

I (1) → first diff AR(1)

I (2) → second diff AR(2)

I (3) → third diff AR(3)

* finding L i.e. Lag multiplier roots.

L = lag multiplier.

we take a mathematical model and assume

$Y_{t-1} = L Y_t$ {previous period is the lag multiplied of Y_t

$Y_{t-2} = L Y_{t-1} = L \cdot L Y_t = L^2 Y_t$

⋮

so on

for AR(1)

$$Y_t = \rho Y_{t-1} + \epsilon_t$$

Now since

$$Y_{t-1} = L Y_t$$

$$\Rightarrow Y_t = \rho L Y_t + \epsilon_t$$

$$\Rightarrow Y_t - \rho L Y_t = \epsilon_t$$

$$\Rightarrow Y_t (1 - \rho L) = \epsilon_t$$

Now if we assume $\epsilon_t = 0$ {under minimising ϵ_t }

$$\Rightarrow \epsilon_t = 0 \text{ if } Y_t = 0 \text{ or } (1 - \rho L) = 0$$

Suppose $(1 - \rho L) = 0$

then $\boxed{\frac{1}{\rho} = L^*}$ → Roots of AR(1).

Lag roots for AR(2)

$$Y_t = \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \epsilon_t$$

From mathematical model we assumed

(i) $Y_{t-1} = L Y_t$

(ii) $Y_{t-2} = L^2 Y_t$

Now

$$Y_t = \rho_1 L Y_t + \rho_2 L^2 Y_t + \epsilon_t$$

$$\Rightarrow Y_t (1 - \rho_1 L - \rho_2 L^2) = \epsilon_t$$

solving quadratic equation $(1 - \rho_1 L - \rho_2 L^2) = 0$

$D = b^2 - 4ac = \cancel{\beta}\ \rho_1^2 + 4\rho_2$

$b = -\rho_1$

$c = 1$

$a = -\rho_2$

$$\therefore \boxed{L^* = \frac{-b \pm \sqrt{D}}{2a} = \frac{+\rho_1 \pm \sqrt{\rho_1^2 + 4\rho_2}}{-2\rho_2}} \quad \text{for } \epsilon_t = 0$$

$$\therefore L^* = \frac{+\rho_1 - \sqrt{\rho_1^2 + 4\rho_2}}{-2\rho_2} \text{ or } \frac{\rho_1 + \sqrt{\rho_1^2 + 4\rho_2}}{-2\rho_2} \quad \checkmark$$

Derivation AR(1) $\longrightarrow$ MA(∞)

$AR = \rho Y_{t-1} + \epsilon_t = Y_t$

since $Y_{t-1} = L Y_t$

$\Rightarrow Y_t = \rho L Y_t + \epsilon_t$

$\Rightarrow Y_t - \rho L Y_t = \epsilon_t$

$\Rightarrow Y_t (1 - \rho L) = \epsilon_t$

$\Rightarrow Y_t = \dfrac{\epsilon_t}{1 - \rho L}$

here $\dfrac{1}{1-\rho L} = \{ (\rho L)^0 + (\rho L)^1 + (\rho L)^2 \cdots \infty \}$

$\Rightarrow Y_t = \{ (\rho L)^0 + (\rho L)^1 \cdots \} \epsilon_t$

$\Rightarrow Y_t = \epsilon_t + \rho L \epsilon_t + \rho^2 L^2 \epsilon_t \cdots \infty$

$\Rightarrow Y_t = \epsilon_t + \rho \epsilon_{t-1} + \rho^2 \epsilon_{t-2} \cdots \infty$

where $\epsilon_{t-1} = L \epsilon_t$

$\epsilon_{t-2} = L^2 \epsilon_t$

$$\Rightarrow \quad Y_t = \rho^0 \epsilon_t + \rho \epsilon_{t-1} \cdots \infty$$

$$\boxed{Y_t = \sum_{i=0}^{\infty} \rho^i \epsilon_{t-i}}$$

MA (1) $\longrightarrow$ AR (∞).

$$Y_t = \epsilon_t + \theta \epsilon_{t-1}$$

we assume

$$\epsilon_{t-1} = L \epsilon_t$$

$$Y_t = \epsilon_t + \theta L \epsilon_t$$

then $Y_t = \epsilon_t (1-\theta L)$

$$\Rightarrow \quad \frac{Y_t}{1+\theta L} = \epsilon_t \qquad \therefore \quad \frac{1}{1+\theta L} = \left\{ (\theta L)^0 + (-\theta L)^1 + (-\theta L)^2 \cdots \infty \right\}$$

odd powers remain negative.

$$\Rightarrow \quad \left\{ (-\theta L)^0 + (-\theta L)^1 + (-\theta L)^2 \cdots \infty \right\} Y_t = \epsilon_t$$

$$\Rightarrow \quad Y_t - \theta L Y_t + \theta^2 L^2 Y_t - \theta^3 L^3 Y_t \cdots \infty = \epsilon_t$$

$$\Rightarrow \quad Y_t = \theta L Y_t - \theta^2 L^2 Y_t + \theta^3 L^3 Y_t \cdots \infty + \epsilon_t$$

Since

$$Y_{t-1} = L Y_t$$

$$\vdots$$

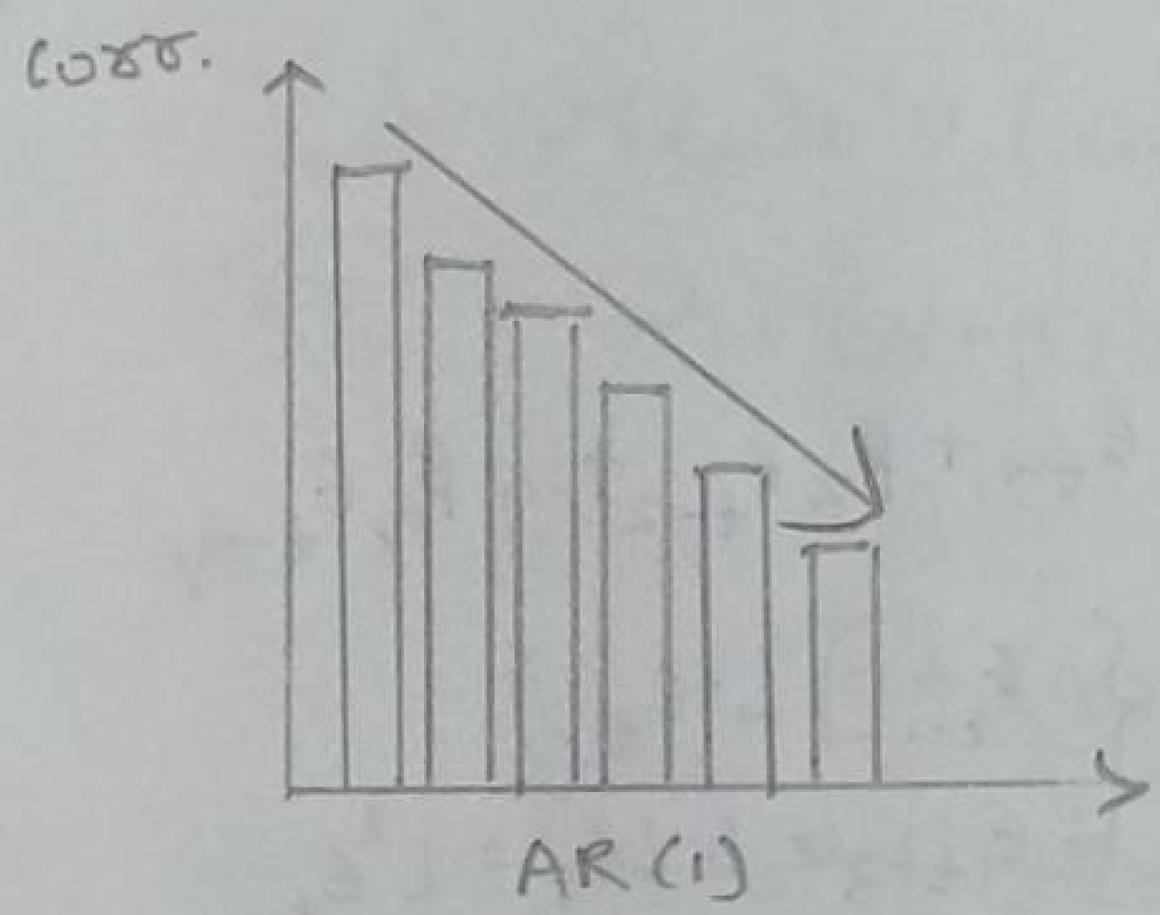
$$Y_{t-h} = L^h Y_t$$

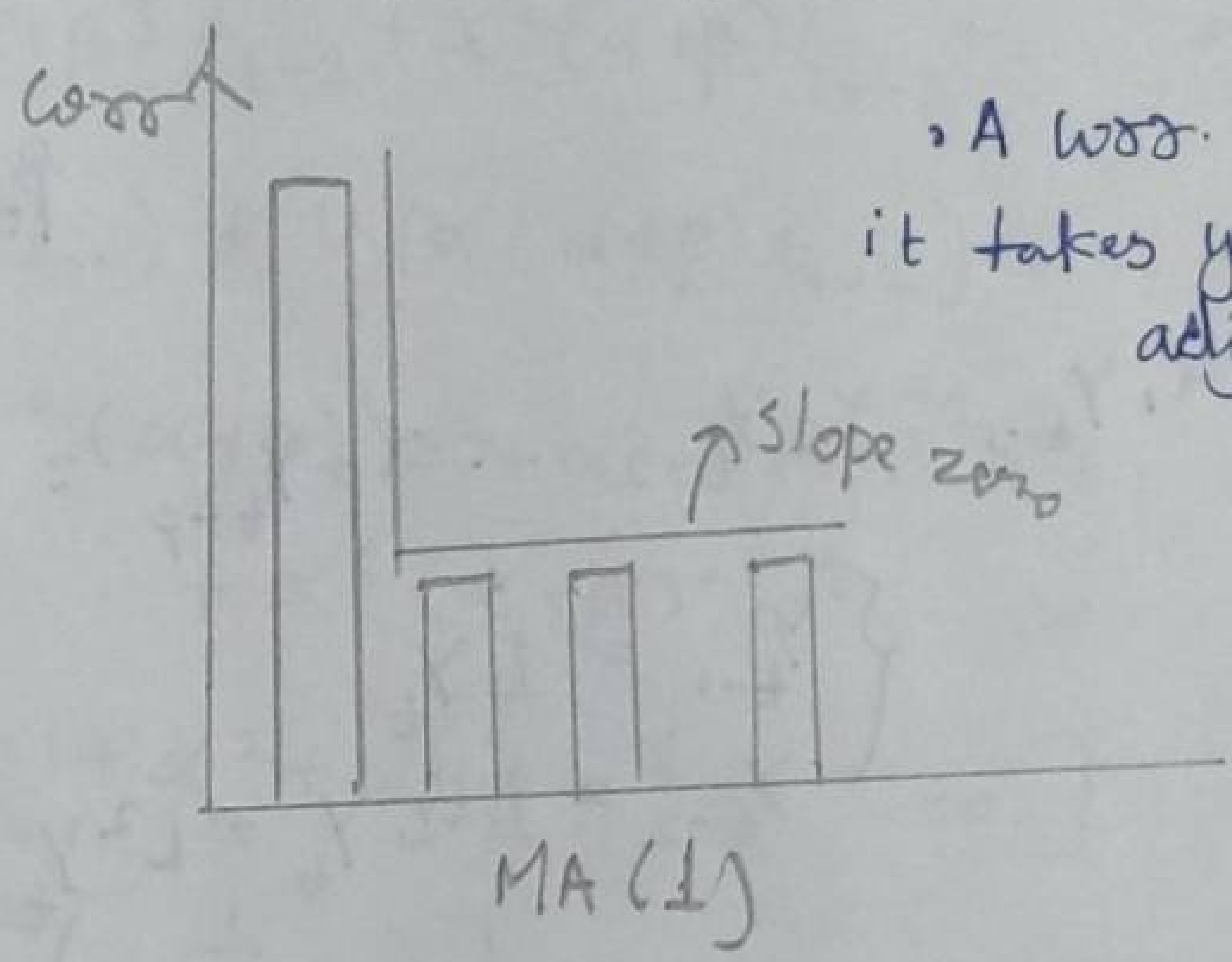
$$\rightarrow \quad Y_t = \theta Y_{t-1} - \theta^2 Y_{t-2} \cdots \infty + \epsilon_t$$

$$\boxed{Y_t = \sum_{i=1}^{\infty} (-\theta)^i Y_{t-i} + \epsilon_t} \longrightarrow \text{MA}(\infty)$$

* $Y_t = \rho Y_{t-1} + \epsilon_t$ —— AR (1) LR

$Y_t = \epsilon_t + \theta \epsilon_{t-1}$ —— MA (1) SR





• A corr. is high hence it takes year to adjust itself.

• If it took more than a one year then AR

• In one year adjusted itself then MA.

MA ⇒ In case of $Y_t = \epsilon_t + \theta \epsilon_{t-1}$ (Here ϵ_t contains all Y_{t-i})

AR → $Y_t = \rho Y_{t-1} + \epsilon_t$ (Here ϵ_t contains all ϵ_{t-i})

* All the Independent Variable has a lag behaviour AR Model.

* **AR-MA**

AR (p) → $Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \epsilon_t$

MA (q) → $Y_t = \beta_0 \epsilon_{t-0} + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + \dots + \beta_q \epsilon_{t-q}$

$[\beta_0 = 1]$

* ϵ_t in AR & $\beta_0 \epsilon_t$ in MA are different

* ARMA (p, q)

$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + \dots + \beta_q \epsilon_{t-q}$.

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} \dots + \beta_q \epsilon_{t-q}$$

$$\epsilon_t \downarrow \beta_0 \epsilon_{t-0}$$

$$Y_t - \alpha_1 Y_{t-1} - \alpha_2 Y_{t-2} \dots \alpha_p Y_{t-p} = \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + \dots \beta_q \epsilon_{t-q}$$

$$\begin{cases} Y_{t-1} = L Y_t \\ Y_{t-2} = L \cdot L Y_t = L^2 Y_t \end{cases} \qquad \begin{cases} \epsilon_{t-1} = L \cdot \epsilon_t \\ \epsilon_{t-2} = L \cdot L \epsilon_t = L^2 \epsilon_t \end{cases}$$

$$Y_t - \alpha_1 L Y_t - \alpha_2 L^2 Y_t \dots \alpha_p L^p y_t = \epsilon_t + \beta_1 L \epsilon_t + \beta_2 L^2 \epsilon_t + \dots \beta_q L^q \epsilon_t.$$

$$\Rightarrow Y_t (1 - \alpha_1 L - \alpha_2 L^2 \dots - \alpha_p L^p) = (1 + \beta_1 L + \beta_2 L^2 + \dots + \beta_q L^q) \epsilon_t$$

$$\Rightarrow Y_t = \left[\frac{1 + \beta_1 L + \beta_2 L^2 + \dots + \beta_q L^q}{1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_q L^q} \right] \cdot \epsilon_t$$

* ARMA (1.1)

$$Y_t = \left(\frac{1 + \beta L}{1 - \alpha L} \right) \epsilon_t$$

$$Y_t = (1 + \beta L) \left(\frac{1}{1 - \alpha L} \right) \epsilon_t$$

$\because S = 1 + r + r^2 + r^3 \dots + \infty$

$\text{Sum} = \frac{1}{1-r}$

$$Y_t = (1 + \beta L)(1 + \alpha L + \alpha^2 L^2 + \alpha^3 L^3 + \dots) \epsilon_t$$

$$Y_t = \left[\begin{matrix} 1 + \alpha L + \alpha^2 L^2 + \alpha^3 L^3 + \dots \\ \beta L + \alpha \beta L^2 + \alpha^2 \beta L^3 + \alpha^3 \beta L^4 + \dots \end{matrix} \right] \epsilon_t$$

$$Y_t = [1 + (\alpha + \beta) L + \alpha (\alpha + \beta) L^2 + \alpha^2 (\alpha + \beta) L^3 + \dots \infty] \epsilon_t$$

$$\hookrightarrow [V(\epsilon_t) = V(\epsilon_{t-i}) = \sigma^2]$$

$$Y_t = \epsilon_t + (\alpha+\beta).L\epsilon_t + \alpha(\alpha+\beta)L^2\epsilon_t + \alpha^2(\alpha+\beta)L^3 + \ldots \infty$$

$$Y_t = \epsilon_t + (\alpha+\beta)\epsilon_{t-1} + \alpha(\alpha+\beta)\epsilon_{t-2} + \alpha^2(\alpha+\beta)\epsilon_{t-3} + \ldots \infty$$

$$v(Y_t) = v(\epsilon_t) + v((\alpha+\beta)\epsilon_{t-1}) + v(\alpha(\alpha+\beta)\epsilon_{t-2}) + \ldots \infty$$

$$v(Y_t) = \sigma^2 + (\alpha+\beta)^2\sigma^2 + \alpha^2(\alpha+\beta)^2\sigma^2 + \alpha^4(\alpha+\beta)^2\sigma^2 + \ldots$$
$$\ldots + \alpha^6(\alpha+\beta)^2\sigma^2 \ldots \infty$$

$$= \sigma^2\left[1 + (\alpha+\beta)^2(1 + \alpha^2 + \alpha^4 + \alpha^6 + \ldots \infty)\right]$$

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

$$v(Y_t) = \sigma^2\left[1 + \frac{(\alpha+\beta)^2}{(1-\alpha^2)}\right].$$

$$Y_{t-1} = L\,Y_t$$

$$= L(\epsilon_t + (\alpha+\beta)L\epsilon_t + \alpha(\alpha+\beta)L^2\epsilon_t + \alpha^2(\alpha+\beta)L^3\epsilon_t + \ldots \infty)$$

$$= L\epsilon_t + (\alpha+\beta)L^2\epsilon_t + \alpha(\alpha+\beta)L^3\epsilon_t + \ldots$$

$$= \epsilon_{t-1} + (\alpha+\beta)\epsilon_{t-2} + \alpha(\alpha+\beta)\epsilon_{t-3} + \ldots$$

$\times$ $cov(Y_t, Y_{t-1})$

$$= cov\left[\epsilon_t + (\alpha+\beta)\epsilon_{t-1} + \alpha(\alpha+\beta)\cdot\epsilon_{t-2} \ldots, \epsilon_{t-1} + (\alpha+\beta)\epsilon_{t-2} + \alpha(\alpha+\beta)\epsilon_{t-3}\right.$$

$$= (\alpha+\beta)\,cov(\epsilon_{t-1}, \epsilon_{t-1}] + \alpha(\alpha+\beta)^2 cov(\epsilon_{t-2}, \epsilon_{t-2}) + \alpha^3(\alpha+\beta)^2 cov(\epsilon_{t-3}, \epsilon_{t-3}) + \ldots$$

$$= (\alpha+\beta)\,v(\epsilon_{t-1}) + \alpha(\alpha+\beta)^2 v(\epsilon_{t-2}) + \alpha^3(\alpha+\beta)^2 v(\epsilon_{t-3}) + \ldots$$

$$= (\alpha+\beta)\sigma^2 + \alpha(\alpha+\beta)^2\sigma^2 + \alpha^3(\alpha+\beta)^2\sigma^2 + \ldots$$

$$= \left[(\alpha+\beta) + \alpha(\alpha+\beta)^2 + \alpha^3(\alpha+\beta)^2 + \alpha^3(\alpha+\beta)^2 + \ldots\right]\sigma^2$$

$$= \left[(\alpha+\beta) + \alpha(\alpha+\beta)^2(1 + \alpha^2 + (\alpha^2)^2 + (\alpha^2)^3 + \ldots\right]\sigma^2$$

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

* ARMA (p, q)

$$Y_t = \left(\sum_{i=1}^{p} \alpha_i L^i\right) Y_t + \left(\sum_{i=0}^{q} \beta_j L^j\right) \epsilon_t$$

$$\left[1 + \sum_{i=1}^{p} \alpha_i L^i\right] \cdot Y_t = \left(\sum_{j=0}^{r} \beta_j L^j\right) \epsilon_t \quad - (1)$$

— ARMA (1, 1)

$$Y_t = \alpha Y_{t-1} + \epsilon_t + \beta \epsilon_{t-1}$$

$$Y_t = \frac{(1+\beta L)}{(1-\alpha L)} \epsilon_t$$

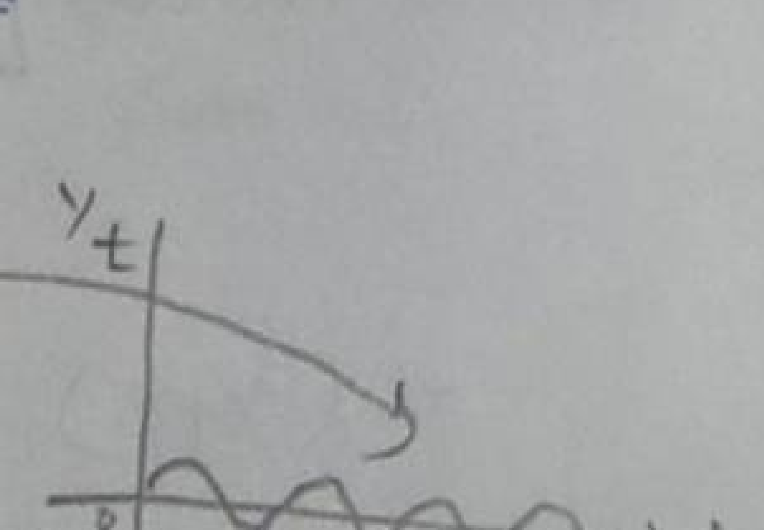
- Testing the Stationarity & WD.

i) Mean: $E(Y_t) = E\left\{\frac{(1+\beta L)}{(1-\alpha L)} \epsilon_t\right\}$

or

$$E(Y_t) = \frac{1+\beta L}{1-\alpha L} \underbrace{E(\epsilon_t)}_{=0}$$

$$\boxed{E(Y_t) = 0} = \mu.$$



ii) Variance

$$V(Y_t) = \sigma^2 \left\{1 + \frac{(\alpha+\beta)^2}{(1-\alpha)^2}\right\} = \text{constant}$$

$$\left[\begin{array}{l} \because \sigma^2 = \text{const} \\ \alpha, \beta = \text{const} \end{array}\right]$$

* Deriving Variance

$$Y_t = \{1 + (\alpha+\beta)L + \alpha(\alpha+\beta)L^2 + \alpha^2(\alpha+\beta)L^3 \dots\} \epsilon_t$$

$$L\epsilon_t = \epsilon_{t-1}$$

$$L^2 \epsilon_t = L \cdot L \epsilon_t$$

$$= L \cdot \epsilon_{t-1} \text{ or } \epsilon_{t-2}$$

$$L^3 \epsilon_t = \epsilon_{t-3}.$$

$$V(Y_t) = V(\epsilon_t + (\alpha+\beta)\epsilon_{t-1} + \alpha(\alpha+\beta)\epsilon_{t-2} + \alpha^2(\alpha+\beta)\epsilon_{t-3})$$

$$\left[\begin{array}{l} \therefore V(ax) = a^2 V(x) \\ V\{(\alpha+\beta)\epsilon_{t-1}\} \\ = (\alpha+\beta)^2 V(\epsilon_{t-1}) \\ = (\alpha+\beta)^2 \sigma^2 \end{array}\right]$$

$$V(\epsilon_t) = V(\epsilon_{t-1}) = V(\epsilon_{t-2}) = \dots = \sigma^2$$

$$(1 + r + r^2 \dots r) \to S = \frac{1}{1-r}.$$

$$V(Y_t) = \{\sigma^2 + (\alpha+\beta)^2\sigma^2 + \alpha^2(\alpha+\beta)^2 + \alpha^4(\alpha+\beta)^2\sigma^2 \dots +]$$

$$= \sigma^2 [1 + (\alpha+\beta)^2 \{1 + (\alpha)^2 + (\alpha^2)^2 \dots \infty\}]$$

$$V(Y_t) = \sigma^2 \left\{1 + \frac{(\alpha+\beta)^2}{(1-\alpha)^2}\right\} = \text{constant}.$$

* Covariance.

$$\text{cov}(Y_t, Y_{t-1}) = \text{cov}[\epsilon_t + (\alpha+\beta)\epsilon_{t-1} + \alpha(\alpha+\beta)\epsilon_{t-2} + \alpha^2(\alpha+\beta)\epsilon_{t-3} \dots]$$

$$\epsilon_{t-1} + (\alpha+\beta)\epsilon_{t-2} + \alpha(\alpha+\beta)\epsilon_{t-3} \dots + \}$$

$$\Rightarrow \{(\alpha+\beta)\sigma^2 + \alpha(\alpha+\beta)^2 + \alpha^3(\alpha+\beta)^2\sigma^2 + \dots\}.$$

$$\Rightarrow \sigma^2 [(\alpha+\beta) + \alpha(\alpha+\beta)^2 \{1 + \alpha^2 + \alpha^3 \dots\}]$$

$$\Rightarrow \sigma^2\left[(\alpha+\beta)+\alpha\frac{(\alpha+\beta)^2}{1-\alpha^2}\right] = \sigma^2\left[\frac{(\alpha+\beta)(1-\alpha^2)+\alpha(\alpha+\beta)^2}{1-\alpha^2}\right]$$

$$= \sigma^2\left[\frac{(\alpha+\beta)(1-\alpha^2)+\alpha^3+\alpha\beta^2+2\alpha\beta}{(1-\alpha^2)}\right]$$

$$= \frac{\sigma^2\left[\alpha+\beta-\alpha^3-\alpha^2\beta+\alpha^3+\alpha\beta^2+2\alpha\beta\right]}{1-\alpha^2}$$

$$= \sigma^2\left[\frac{\alpha+\beta+\alpha\beta^2+\alpha^2\beta}{1-\alpha^2}\right]$$

$\text{cov}(Y_t, Y_{t-1})$ or $\gamma = \sigma^2\left[\frac{(\alpha+\beta)(1+\alpha\beta)}{1-\alpha^2}\right]$ = constant.

$\Rightarrow$ In book, it is given like this.

for $\text{cov}(Y_t, Y_{t+h}) = \gamma\rho(h-1)$ when $h \geq 2$.

Q. Take a time trend and a first difference of order 2.

$\rightarrow$ $Y_t = \alpha_0 + \alpha_1 t + \epsilon_t$ —— (1) simple time trend model.

ΔY_t is the 1st diff.

$\Delta Y_t = (Y_t - Y_{t-1})$ — I (1) if series is stationary.

2nd diff.

$\Delta^2 Y_t = \{(Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})\}$ — I (2) if stationary.

d order difference

$\Delta^d Y_t = \{(Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) - (Y_{t-2} - Y_{t-3}) \dots (Y_{t-\alpha+1} - Y_{t-\alpha})]$

$\rightarrow Y_{t-1} = \alpha_0 + \alpha_1(t-1) + \epsilon_{t-1}$ [Slutsky effect]

$\Delta Y_t = (Y_t - Y_{t-1}) = (\alpha_0 - \alpha_0) + (\alpha_1 t - \alpha_1 t) - \alpha_1 + \epsilon_t - \epsilon_{t-1}$

$\Delta Y_t = -\alpha_1 + \epsilon_t - \epsilon_{t-1}$ —— (2)

$$\boxed{\Delta Y_t = -\alpha_1 + \epsilon_t + \theta\, \epsilon_{t-1}}$$ - when $\theta = -1$

It has MA components $\therefore Y_t = f(\epsilon_t, \epsilon_{t-1}, \dots)$

$Y_t = \epsilon_t + \theta(\epsilon_{t-1}$ <u>MA(1)</u>

* **ARIMA**

1st difference.

$\Delta Y_t = (Y_t - Y_{t-1}) = \omega_{t-1}$ (only assumed)

$\omega_t = \alpha_1 \omega_{t-1} + \alpha_2 \omega_{t-2} + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2}$

1) ARIMA $(\underset{AR}{2} \cdot \overset{\text{diff (Integration)}}{1} \cdot \underset{MA}{2})$

$$\Delta Y_t = \alpha_1 \Delta Y_{t-1} + \alpha_2 \Delta Y_{t-2} + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} \dots$$

2) 2nd diff. order.

ARIMA $(2, 2, 2)$

$$\Delta^2 Y_t = \alpha_1 \Delta^2 Y_{t-1} + \alpha_2 \Delta^2 Y_{t-2} + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2}$$

$$\Delta^2 Y_t = (Y_t - Y_{t-1})$$

$$\Delta^2 Y_t = (Y_t - Y_{t-2}$$

$$\Delta^2 Y_t = (Y_t - Y_{t-3})$$

$I(1) \rightarrow \Delta^1 Y_t = Y_t - Y_{t-1}$

$I(2) \rightarrow \Delta^2 Y_t = Y_t - Y_{t-2} = Y_t - Y_{t-2} - Y_{t-1} - Y_{t-1}$

$= (Y_t - Y_{t-1}) + (Y_{t-1} - Y_{t-2})$

$I(3) \Rightarrow \Delta^3 Y_t = Y_t - Y_{t-3} = (Y_t - Y_{t-1}) + (Y_{t-1} - Y_{t-2}) + (Y_{t-2} - Y_{t-3})$

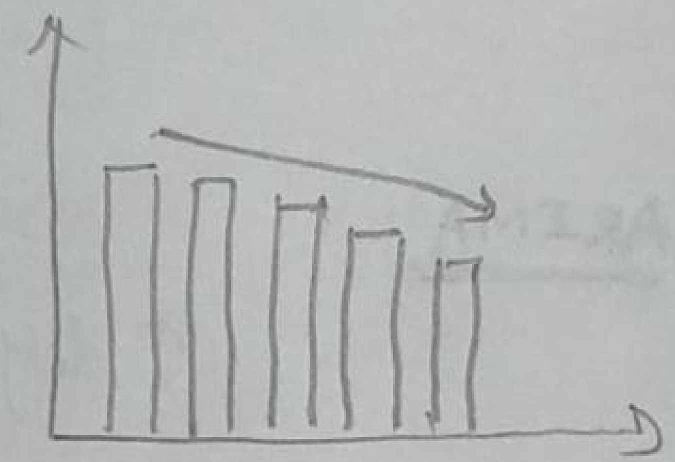
Integration of 3rd order - * We use integration if we face a non-stationary data

* first we check for stationarity and WD and if there is WD, then they are independent.

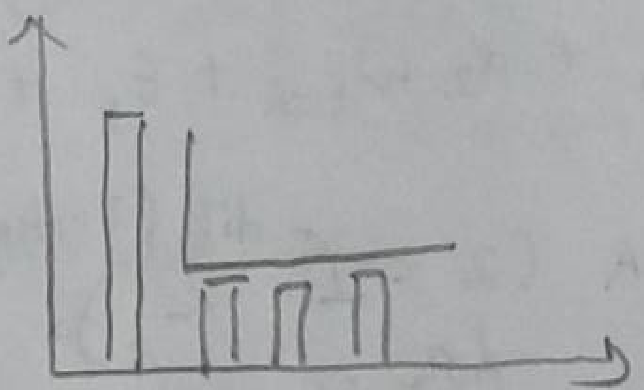
Then if we take lag

$corr(Y_t, Y_{t+\underline{h}})$

↳ lag



If we find W.D, we should go for Asymptotic least square method.



Example : Simple Model

If we take $Y_t = \alpha_0 + \alpha_1 t + \epsilon_t \longrightarrow$ OLS

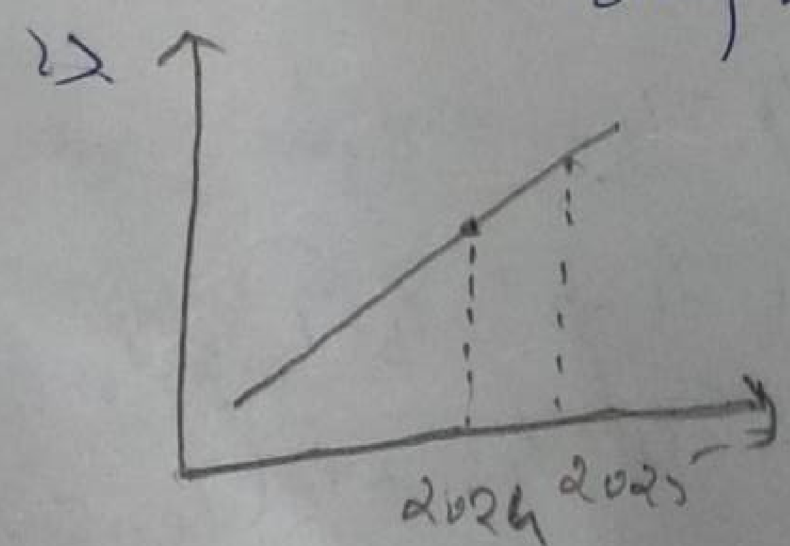
$\hat{\alpha}_0$ & $\hat{\alpha}_1$

$\therefore Y_t = \hat{\alpha}_0 + \hat{\alpha}_1 t + \epsilon_t \doteq 0$

↳ $t = 2025$

Box-Jenkins method of forecasting

↳ In R & phython its only way of forecast



* **Estimation.**

* Time Trend Model $- Y_t = \alpha_0 + \alpha_1 t + \epsilon_t$

  a) OLS

  b) Estimate $\hat{\alpha}_0$, $\hat{\alpha}_1$

  c) Y_{t+1} we can estimate with given $(Y_t = \hat{\alpha}_0 + \hat{\alpha}_1 t)$

We have data GDP of India.

* India ⇒

	GDP(Y_t)	[Independent] t
$t = 2000, \dots, 2023$	Y_1	2000
$t = 1, 2, \dots, 24$	Y_2	2001
	⋮	⋮
	Y_t	2023

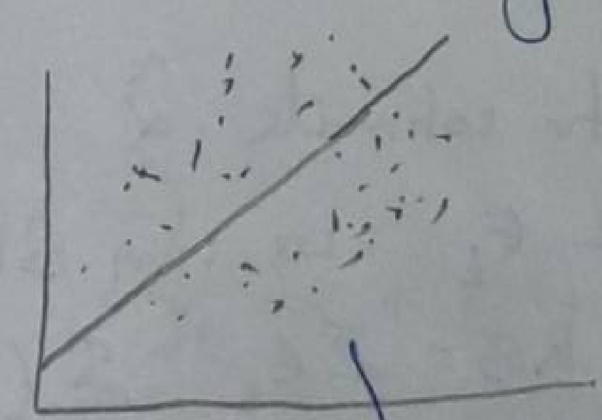
we run an estimation

we get $\hat{\alpha}_0 = 20$

$\hat{\alpha}_1 = 1$

How we get we run OLS on data with GDP (dep) & t (independent) and we get model $Y_t = 20 + 1t$

Now my objective is to predict GDP at $t = 2030$??



$$Y_t = 20 + 1 \times (2030)$$
$$= 20 + 2030$$
$$Y_t = 2050$$

we fitted the model then we extended it to get 2030's GDP.

We have to check from the model if its following stationarity & WD. if its following then we will go for OLS. & then only we can go for forecast.

{we also need to check autocorrelation, Multicollinearity}

⇒ If we fail to apply OLS, you wont be able to forecast.
↓
So if it doesn't fit, fit it forcefully, then we can go for least squares or max^m likelihood.

We have done a) b) c) } → But before it we need to check for stationarity & WD

So we go from one model to another model to satisfy the assumption, once it satisfies, we go for forecasting.

* In research → before going to ARIMA – you should justify why no AR, MA, ARMA
  - so we need to ensure, all the assumptions are satisfied when we OLS or AR or MA etc.

* <u>Estimation of AR.</u>

- AR (1)

$$Y_t = \alpha Y_{t-1} + \epsilon_t \quad \text{— AR (1)}$$

estimate → $\hat{\alpha}$ — How to estimate $\hat{\alpha}$

we have ϵ_t. we can estimate

$RSS = \sum \epsilon_t = \sum (Y_i - \hat{Y}_i)^2$

we need to minim (optimise) or OLS we get $\hat{\alpha}$

From $\hat{\alpha}$ we get eq^n like

$$Y_{2023} = \hat{\alpha} Y_{2022} + 0$$
$$Y_{2024} = \hat{\alpha} Y_{2023}$$
$$Y_{2025} = \hat{\alpha} Y_{2024}$$
$$Y_{2026} = \hat{\alpha} Y_{2025}$$

The more lag value we bring in, the more parameter

* forecasting *

* Estimation of MA

$$Y_t = \epsilon_t + \theta \epsilon_{t-1} \quad \text{— MA (1)}$$

can we run OLS ? → No we cannot, because in OLS we need to minimize $\sum \epsilon_t^2$.

$$Y_t = \alpha_0 + (\epsilon_t + \theta \epsilon_{t-1}) \quad \text{— MA (1)}$$

So which error term would we minimize ϵ_t or ϵ_{t-1}.

It will be complicated we can't run OLS. We cannot find $\hat{\alpha}_0$ and we cannot α_0 forecast.

To solve it first we alter the eq^n to estimate ϵ_t.

$$\epsilon_t = Y_t - \alpha_0 - \theta \epsilon_{t-1}$$

if $t = 1$, $\epsilon_1 = Y_1 - \alpha_0 - \theta \epsilon_{1-1}$ {In period 1 we assume $\epsilon_{1-1} = \epsilon_0 = 0$}

$$\epsilon_1 = Y_1 - \alpha_0$$

$$\vdots$$

$$\epsilon_2 = Y_2 - \alpha_0 - \theta \epsilon_1$$

$$\vdots$$

$$\epsilon_m = Y_m - \alpha_0 - \theta \epsilon_{m-1}$$

- MA (2)

$$Y_t = \alpha_0 + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2}$$

$$\epsilon_t = Y_t - \alpha_0 - \beta_1 \epsilon_{t-1} - \beta_2 \epsilon_{t-2}$$

Now,

$$\epsilon_1 = Y_1 - \alpha_0 - \beta_1 \epsilon_0 - \beta_2 \epsilon_{-1}$$

$$\epsilon_1 = Y_1 - \alpha_0$$

{$\because \epsilon_0 = \epsilon_{-1} = 0$} ϵ_{-1} & ϵ_0 has no meaning

Now

$\epsilon_2 = Y_2 - \alpha_0 - \beta_1 \epsilon_1$

$\epsilon_3 = Y_3 - \alpha_0 - \beta_1 \epsilon_2 - \beta_2 \epsilon_1$

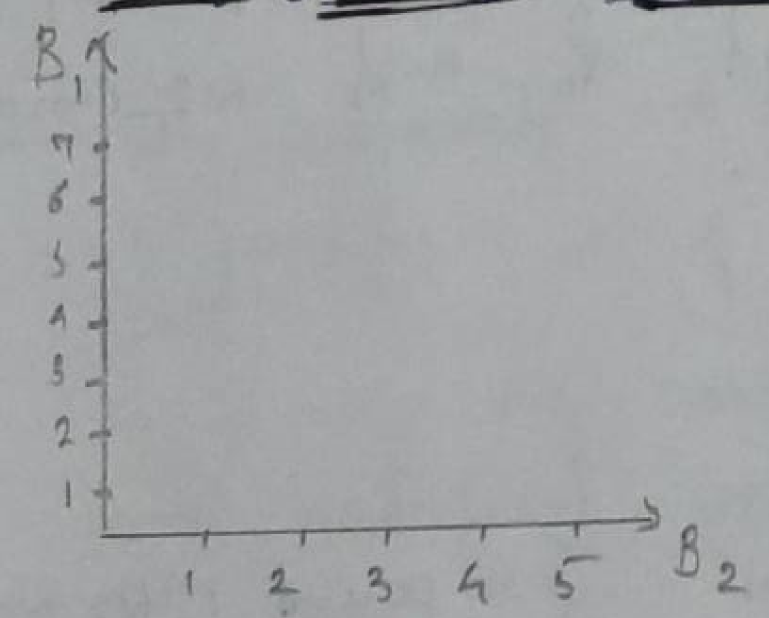
$\epsilon_4 = Y_4 - \alpha_0 - \beta_1 \epsilon_3 - \beta_2 \epsilon_2$

But the problem is even if there Y_i, we don't have β_1 & β_2 & we cannot estimate ϵ_t & if we cannot get ϵ_t we cannot calculate $\sum \epsilon_t^2 \rightarrow$ min.

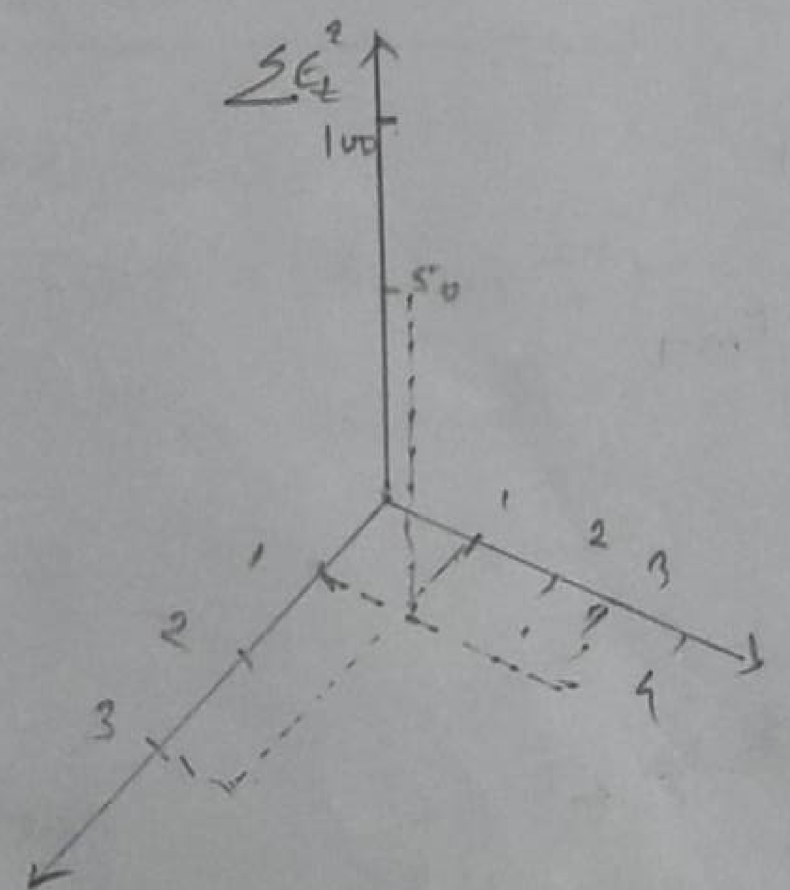
So we cannot run OLS we cannot get estimated value.

Hence we come up to Box - Jenkin method.

* **Box Jenkin Method** [also known as Grid search method]



From this graph we can take values of β_1 & β_2 & we can calculate ϵ_t's and we can plot it on graph.



$\therefore$ min $\sum \epsilon_t^2 \rightarrow \theta \; (\beta_1, \beta_2)$

- when the parameter increase from more than 2

$\rightarrow$ n dimension, B-J fails.

This is like trial error method we need to plot $\sum \epsilon_t^2$ for all θ β_1's & β_2's and from all the plotted $\sum \epsilon_t^2$ we need to find the min^m $\sum \epsilon_t^2$

But the thing is when we further add a drift to the model, we need to find α_0, β_1 & β_2.

Then we will have 4-Dimensions and in 4-D we don't have GRID. its hypothetical situation. Here it becomes complicated & we cannot find $\min \sum \epsilon_t^2$.

Here Box-Jenkins method becomes obselete.

- MA method is a trap, to come out of it, we use B-J
- So B-J is effective till $MA(2)$

Something is better than nothing

Theory

Basic Info (for test II)

Stationarity - Property where statistical characteries (such as μ, Var, autocovariance) of data do not change over time. (easier to model & forecast this way)

Types — Strict → entire distribution remains same over time
pdf same.

— weak - only the first two moments i.e., mean & Varian are constant over time, the covar. betn two time periods depends only on lag Variable betn them.

Why stationarity imp?

Many time series models, including ARMA & ARIMA assume that underlying data is stationary because it amplifies the modelling process. Non-stationary data often leads to inaccurate or unreliable models, so its crucial to make the data Stationary before applying these models.

- How to check Stationary
  - * Visual Inspectn - plot data, observe trends, seasonal patterns.
  - Statistical tests - ADF (augm. dickey fuller) test.
    - └ checks if unit root present in data
    - KPSS test

* Making non stationary data stationary:
  - Differencing - Subs. previous from current. to remove trends.
  - Transformatn - Apply log, sq. root etc.

★ ARMA (Auto Regressive Moving avg).

∟ used for stationary time series data. Combines 2 components

1. AR (H) comp.

In AR model, Value at time 't' is regressed on its own previous Values. Order of AR, denoted as 'p' refers to no. of time periods the current Value depends on.

- AR(p): $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots \phi_p Y_{t-p} + \epsilon_t$.

2. MA comp

In MA model the Value at time 't' depends on residual (error terms) from previous time periods. The order of MA denoted as 'q', refers to no. of lagged forecast errors that are included in the model.

- MA(q): $Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$.

- ARMA (p,q) Model – combines both comp.

  - ARMA (p,q): $Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots \theta_p \epsilon$

★ ARIMA. → Integrated

∟ This model extends ARMA to handle non-stationarity data by introducing differencing component (I).

1. AR – Same as ARMA.

2. I → Represent no. of time data need to be differenced to become stationary. The order of differencing is denoted by 'd'

- If data diff. once: $Y_t' = Y_t - Y_{t-1}$

3) MA comp – Same as ARMA.

★ ARIMA (p,d,q) Model: combines AR, differencing & MA.

- ARIMA (p,d,q): $Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$

★ Key Differences:

feature	R.W without drift	RW with drift.
Eqn	$Y_t = Y_{t-1} + \epsilon_t$	$Y_t = \alpha + Y_{t-1} + \epsilon_t$
Trend	No deterministic trend (pure randomness)	Deterministic trend driven by α
Expected Value	$E(Y_t) = Y_0$	$E(Y_t) = Y_0 + \alpha t$.
Variance	Increases over time	Increases over time.
Direction	Randomly fluctuats with no systematic bias	fluctuates with a systematic upward/downward bias (trend)
long term behaviour	Wanders randomly with no tendency to rise or fall	Tends to rise (or fall) over time due to [illegible] $\cdot\,\alpha$.

★ Key Differences:

Feature	R.W without drift	RW with drift.
Eqn	$Y_t = Y_{t-1} + \epsilon_t$	$Y_t = \alpha + Y_{t-1} + \epsilon_t$
Trend	No deterministic trend (pure randomness)	Deterministic trend driven by α
Expected Value	$E(Y_t) = Y_0$	$E(Y_t) = Y_0 + \alpha t$.
Variance	Increases over time	Increases over time.
Direction	Randomly fluctuats with no systematic bias	fluctuates with a systematic upward/downward bias (trend)
long term behaviour	Wanders randomly with no tendency to rise or fall	Tends to rise (or fall) over time due to $t \cdot \alpha$.

★ Dickey fuller & Augmented D.F. test.

tests used to check for the presence of unit root in time series when is an indications of whether the series is non-stationary.

- Dickey fuller – used to test null hypothesis that a time series has unit root, meaning its non stationary. It tests rejects the null hypothesis, it suggests the series is stationary.

- Augmented D.F.

  — is an extension of ~~A~~DF that address the issue of auto correlation in the residuals. By adding lagged differences of the ~~in~~dependent variables to the regression, the ADF test accounts for autocorrelation and provides more reliable results.

Aspect	D.F test	ADF test
Purpose	test unit root (non stationary)	test unit root (non stat.) which accounting for autocorrelation.
Model	Basic AR(1)	Extended the DF by adding lag diff. AR2, AR3...
Autocorr in residuals	Assumes no auto-corrlⁿ in error term	Accounts for autocorr. by adding lagg diff of dependent variables.
Applicability	Best suited for data with no serial correlatⁿ in residuals	Handles wide range of data with serial corr. as well.

Matrix form Representation

Cross section : (Univariate case)

$$
\underset{n\times 1}{\overset{Y_i}{\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \end{bmatrix}}} = \underset{n\times 2}{\overset{X_i}{\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ & \vdots \\ & X_n \end{bmatrix}}} \underset{2\times 1}{\overset{\beta_i}{\begin{bmatrix} \beta_0 \\ \beta_2 \end{bmatrix}}} \underset{n\times 1}{\overset{u}{\begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix}}}
$$

$$Y = X\beta + U.$$

Multivariate case

$$
\underset{n\times 1}{\overset{Y_i}{\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}}} = \underset{n\times (K+1)}{\overset{X_i}{\begin{bmatrix} 1 & X_{11} & \cdots & X_{1K} \\ 1 & X_{21} & \cdots & X_{2K} \\ 1 & & & \vdots \\ \vdots & & & \vdots \\ 1 & & & X_{nK} \end{bmatrix}}} \underset{(K+1)\times 1}{\overset{\beta}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}}} + \underset{n\times 1}{\overset{U_i}{\begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix}}}
$$

$$\underset{n\times 1}{Y} = \underset{n\times 1}{X\beta} + \underset{n\times 1}{U}$$

For timeseries and cross section only.

$U^T U$ = RSS in matrix = salar matrix

$$Y = X\beta + U$$

$$\Rightarrow Y - X\hat{\beta} = \hat{U}$$

$$\therefore RSS_{1\times 1} = \hat{U}^T\hat{U} = (Y - X\hat{\beta})^T (Y - X\hat{\beta})$$

$$= (Y^T - \hat{\beta}^T X^T)(Y - X\hat{\beta})$$

$$= Y^T Y - \underbrace{Y^T X}_{A}\underbrace{\hat{\beta}}_{X} - \underbrace{(\hat{\beta})^T}_{X'} \underbrace{X^T Y}_{A} + \hat{\beta}^T \underbrace{X^T X}_{A} \hat{\beta}$$

X' X ?

$Y = AX$

$\frac{\partial Y}{\partial X} = A^T$

$\therefore$ FOC

$$\frac{\partial RSS}{\partial \hat{\beta}} = 0 - (Y^T X)^T - (X^T Y) + 2\hat{\beta} X^T X$$

$$= -X^T Y - X^T Y + 2X^T X\hat{\beta}$$

$$\Rightarrow -2X^T Y = 2X^T X\hat{\beta}$$

$$\Rightarrow (X^T X)^{-1} X^T Y = (X^T X)(X^T X)^{-1}\hat{\beta}$$

$$\Rightarrow \hat{\beta} = (X^T X)^{-1}(X^T Y)$$

$\therefore$ SOC

$$\frac{\partial^2 RSS}{\partial \hat{\beta}^2} = 0 - 0 + 2X^T X$$

$$\Rightarrow X^T X > 0,\ 2 > 0$$

$$\therefore \frac{\partial^2 RSS}{\partial \hat{\beta}^2} > 0 \quad \text{(convex or minima)}$$

order of $U^T = 1 \times n$ & order of $U = n \times 1$

$\therefore U^T U = [\ ]_{1\times 1}$.

Panel Data

	Consumptn		Income	
S. no	2015	2020	2015	2020
1	70	80	90	100
2	40	50	60	70
3	80	90	110	120

→ dependent Variable.

Construction of Matrix

$$\underset{6\times 1 \text{ or } NT\times 1}{\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix}} = \underset{6\times 2}{\begin{bmatrix} 1 & X_{11} \\ 1 & X_{12} \\ 1 & X_{21} \\ 1 & X_{22} \\ 1 & X_{31} \\ 1 & X_{32} \end{bmatrix}} \underset{2\times 1}{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}} + \underset{6\times 1}{\begin{bmatrix} U_{11} \\ U_{12} \\ U_{21} \\ U_{22} \\ U_{31} \\ U_{32} \end{bmatrix}}$$

$C_{it} = X_{it}\beta + U_{it}$ (Univariate case)

In non-matrix form

$Y_{it} = \beta_0 + \beta_1 X_{it} + U_{it}$ (Univariate case)

Panel Data - Multivariate case

$$C_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Y_{it} + U_{it}$$

In Matrix :

$$\underset{(N\times t)\times 1}{\begin{bmatrix} C_{11} \\ C_{12} \\ C_{21} \\ C_{22} \\ \vdots \\ C_{Kt-K} \\ \vdots \\ C_{Nt} \end{bmatrix}} = \underset{NT\times(K+1)}{\begin{bmatrix} 1 & X_{11} & Y_{11} & Z_{11} & \cdots & t_{11} \\ 1 & X_{12} & Y_{12} & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & & \vdots \\ \vdots & X_{21} & Y_{21} & & & \vdots \\ \vdots & X_{22} & Y_{22} & & & \vdots \\ \vdots & \vdots & \vdots & & & \vdots \\ & X_{Nt} & Y_{Nt} & Z_{Nt} & & t_{Nt} \end{bmatrix}} \underset{(K+1)\times 1}{\begin{bmatrix} \beta_0 \\ \vdots \\ \beta_K \end{bmatrix}} + \begin{bmatrix} U \\ \vdots \\ U_{nK} \end{bmatrix}$$

$$= NT\times 1 = NT\times 1$$

So concluding
Matrix forms (Univariate)

cross Section =

$$\underset{n\times 1}{\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}} = \underset{n\times 2}{\begin{bmatrix} 1 & X_1 \\ \vdots & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}} \underset{2\times 1}{\begin{bmatrix} \beta_0 \\ \beta \\ \vdots \\ \beta_n \end{bmatrix}} + \begin{bmatrix} U \\ \vdots \\ \vdots \end{bmatrix}$$

Multivariate :

$$\underset{n\times 1}{\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}} = \underset{n\times(K+1)}{\begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{nn} \\ & \ddots & & & \vdots \\ & & & & X_{\cdot n} \end{bmatrix}} \underset{(K+1)\times 1}{\begin{bmatrix} \beta_0 \\ \vdots \\ \beta_n \end{bmatrix}} + \begin{bmatrix} U \\ \vdots \\ \vdots \end{bmatrix}$$

time series samp

Panel series — Univariate done ✓ / Multivariate done ✓

Degrees of Freedom (??)

DoF In econometrical analysis are the independent values or quantities that can vary in a statistical calculations with violating any constraints.

D.f of cross sectional data sets are

$$df = n - (K+1)$$

n: no. of observation; K: slope parameters; 1: intercept parameter

$$df_{cross} = n-(K+1) = n-K-1$$

- **D.f of panel data sets.**

$$df_{panel} = NT - (K+1)$$

NT: no. of obs. at diff time.

and in panel data $NT > N$.

Thus, why do we use panel data instead of time series & cross section. ?

∵ it has more degrees of freedom compare to other data analysis method.

- **Panel Data.**

The panel data i.e, different observations at different time periods, has 3 models within it.

1. Pooled regression model
2. Fixed effect model
3. Random effect model

But, only one among them is used, depending on the data and the testing of hypothesis.

- Panel data representation.

$$Y_{it} = \beta_0 + \beta_1 X_{it} + U_{it}$$

Y_{it} = dependent Var.
X_{it} = independent Var
β_0 = intercept
β_1 = slope parameter

i = obs
t = time
it = i^{th} observatn at time t.

Error term U_{it} under panel data consists of 2 components.

i.e

U_{it}:
- α_i = individual specific heterogenity
- η_{it} = individual at time specific heterogenity
  - ↳ both individual and time heterogenity.

These models can be represented as:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + \eta_{it}$$

Note

correlation between X_{it} & α_i

if $E(X_{it}, \alpha_i) = 0 \Rightarrow$ {we use Random effect model REM}

$E(X_{it}, \alpha_i) \neq 0 \Rightarrow$ {we use fixed effect model FEM}

REM $= 0 \Rightarrow E(\alpha_i, X_{it}) = 0$

FEM $\neq 0 \Rightarrow E(X_{it}, \alpha_i) \neq 0$

$E(X_{it}, \alpha_i)$
- $= 0 \rightarrow$ REM
- $\neq 0 \rightarrow$ FEM

Diagrammatic model of selection of FEM & REM while Running Panel Data.

Pooled Regression

H_0: Pooled regression
H_1 : REM
Test significance through LM test

H_0 : PRM
H_2 : FEM
Test significance through restricted F test

REM ← If both are significant in ↑ → FEM

Run Hausman test to check which is apt for this specific dataset

⇓

H_0 : REM
H_2 : FEM

Given the significance of hausman test, we select an appropriate model to run a panel data.

- **ASSUMPTIONS OF OLS**

i) $E(U_i | X_i) = 0 \quad \forall i = 1 \dots n$

ii) $\sum_{i=1}^{n} X_i U_i = 0 \quad cov(X_i, U_i) = 0$ — if violated causes endogenity problem

iii) $\sum_{i=n}^{n} U_i = 0, \quad \forall i = 1 \dots n.$

iv) Zero autocorrelation i.e., $corr(U_i, U_j) = cov(U_i, U_j) = 0$

v) $V(U_i) = \sigma_u^2$ { if violated, causes heteroskedasticity }

vi) No multicollinearity.

5 Important OLS Assumptions

i) $\sum_{i=1}^{n} U_i = 0$ i.e sum of error = 0

ii) $\sum_{i=n}^{n} X_i U_i = cov(X_i, U_i) = 0$ { no endogenity }

iii) $cov(U_i, U_j) = 0 \ \forall \ i \neq j$ i.e { zero autocorrelation } = $E(U_i U_j) = 0$

iv) $V(U_i) = \sum_{i=1}^{n} (U_i - \bar{U})^2 = \sigma_U^2$ { homoskedasticity } $E(U_i, U_j) = \sigma^2$ $i = j$

v) No multicollinearity.

vi)* $E(U_i \mid X_i) =$ Zero conditional mean = 0

or

$E(U_i) = 0$

Matrix Derivation of BLUE

We know, from $\hat{U}^T \hat{U}$ i.e RSS of order 1×1

$$\hat{U} = Y - X\hat{\beta}$$

$$\hat{U}^T = Y^T - X^T \beta^T$$

$\hat{U}^T \hat{U} = RSS$ But $\hat{U} \hat{U}^T =$ Variance - covariance matrix.

Error = $Y_i - \bar{Y}$

Residual = $Y_i - \hat{Y}$ → estimated

Sample or actual

• Derive the Variance & co-variance Matrix

V-C Matrix $= E(\hat{U}\hat{U}^T) = \sigma_n^2$

$$\hat{U} = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{bmatrix}_{N\times 1} \qquad \hat{U}^T = \begin{bmatrix} U_1 & \cdots & U_N \end{bmatrix}_{1\times N}$$

$$\therefore \hat{U}\hat{U}^T = \begin{bmatrix} U_1 \\ \vdots \\ U_N \end{bmatrix}_{N\times 1} [U_1, \ldots, U_N]$$

$$= \begin{bmatrix} U_1U_1 & U_1U_2 & U_1U_3 & \cdots & U_1U_N \\ & \ddots & & & \vdots \\ U_NU_1 & U_NU_2 & \cdots & \cdots & U_NU_N \end{bmatrix}$$

Now

$$E[\hat{U}\hat{U}^T]$$

$$\rightarrow E \begin{bmatrix} U_1U_1 & \cdots & U_1U_N \\ \vdots & \ddots & \\ U_NU_1 & & U_NU_N \end{bmatrix}$$

$\therefore$ The diagonal elements of the Matrix, represent $E(U_iU_j) \forall i=j$
$= \sigma^2$

While off diagonal elements i.e,
$E(U_iU_j) \forall i \neq j = 0$
$= cov(U_j, U_i) = 0$

$$VC = \begin{bmatrix} E(U_1U_1) & \cdots & E(U_1U_N) \\ \vdots & & \\ E(U_NU_1) & \cdots & E(U_NU_N) \end{bmatrix}$$

Now

Under 5 main OLS assumption

1) $\Sigma U_i = 0$
2) $E(U_i, U_j) = cov(U_i U_j) =$
3) $V(U) = E(U_iU_j) \forall i=j = \sigma$
4) No multicollinearity
5) $E(U_i, X_i) = cov(U_i X_i)$
$= 0$

$$E(\hat{U}\hat{U}^T) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & & \vdots \\ \vdots & \vdots & & \\ 0 & & \cdots & \sigma^2 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \sigma^2 I_n$$

ANNOVA

$$\left.\begin{aligned} TSS &= \text{Total sum of squares} = \sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2 \\ RSS &= \text{Residual sum of squares} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \\ ESS &= \text{Explained sum of squares} = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y}_i)^2 \end{aligned}\right\} \text{non-matrix representation.}$$

Matrix Representation

Since TSS, RSS, ESS are all scalar values

$$\therefore \; TSS = [T^T T]_{1\times1} = \sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2$$

$$RSS = [U^T U]_{1\times1} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$ESS = [E^T E]_{1\times1} = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y}_i)^2.$$

- **Steps to calculate ESS, TSS & RSS.**

1) Run OLS on $Y = XB + U$
2) obtain $\hat{\beta} = (X'X)^{-1}X'Y$
3) estimate $\hat{Y} = X\hat{\beta}$

Run OLS on $Y = X\beta + U$
est $\hat{\beta} = (X'X)^{-1}X'Y$
est $\hat{Y} = X\hat{\beta}$ (use it in RSS)

Suppose $\hat{Y} = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$ & $Y_i = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ $\therefore \bar{Y}_i = \frac{\sum Y_i}{n} = 2$

$\therefore TSS = T^T T$

$$\therefore T = \begin{bmatrix} 1-2 \\ 2-2 \\ 3-2 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

($Y_i - \bar{Y}$)

$$T^T = [-1 \ 0 \ 1]$$

$$\therefore T^T T = [-1 \ 0 \ 1] \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = 1+1=2$$

$$RSS = U^T U = \Sigma (Y_i - \hat{Y}_i)^2$$

$$Y_i = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \hat{Y}_i = \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}$$

$$\therefore (Y_i - \hat{Y}) = \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix}$$

$$\therefore U^T U = [-1 \ 0 \ -1] \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix} = 1+0+1 = 2$$

$ESS = (\hat{Y} - \bar{Y})$

let $E = (\hat{Y}_i - Y_i)$

$$\Rightarrow E = \begin{bmatrix} 2-2 \\ 2-2 \\ 4-2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

$$E^T E = [0 \ 0 \ 2] \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

$$= 0+0+4 = 4$$

~~Rove~~ PROVE

$TSS = ESS + RSS$.

ANNOVA TABLE (Application of ESS, TSS & RSS)

The application of ESS, TSS & RSS is called ANOVA, through which we derive test statistics.

ESS tells us no. of Variables.

K = no. of independent variables

n = no. of observations.

		d.f	Mean Square
ESS	$\Sigma(\hat{Y}_i - \bar{Y})^2$	K	$ESS/_{df} = \frac{\Sigma(\hat{Y}_i - \bar{Y})^2}{K}$
RSS	$\Sigma(Y_i - \hat{Y})^2$	n-K-1	$RSS/_{df} = \frac{\Sigma(Y_i - \hat{Y})^2}{n-K-1}$
TSS	$\Sigma(Y_i - \bar{Y})^2$	RSS+TSS = n-k-1+K = n-1	no need.

f statistics

$$f_{cal} = \frac{MSE}{MSR}$$

$$= \frac{\Sigma(\hat{Y}_i - \bar{Y})^2}{K} \times \frac{n-k-1}{\Sigma(Y_i - \hat{Y})^2}$$

$$= \frac{ESS}{K} \frac{(n-k-1)}{RSS}$$

$F_{K, N-K-1}$

Derivation $\hat{\beta}_0$ & $\hat{\beta}_1$ (matrix format)

$$RSS = \sum_{i=1}^{n} U_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \quad \text{(non matrix form)}$$

$$RSS = \hat{U}^T\hat{U} = (Y - X\beta)^T(Y - X\beta)$$

$$\sum_{i=1}^{n} U^TU = \sum_{i=1}^{n} (Y - X\beta)^T(Y - X\beta) = \sum (Y^T - \hat{\beta}^T X^T)(Y - X\hat{\beta})$$

$$= \sum (Y^TY - Y^TX\hat{\beta} - \hat{\beta}^TX^TY + \hat{\beta}^TX^TX\hat{\beta})$$

$$RSS = \sum Y^TY - Y^TX\hat{\beta} - \hat{\beta}^TX^TY + \hat{\beta}^TX^TX\hat{\beta}$$

$$\frac{\partial RSS}{\partial \hat{\beta}} = 0 - (Y^TX)^T - (X^TY) - 2(X^TX)\hat{\beta} = 0$$

$$\Rightarrow -(X^TY) - (X^TY) + 2(X^TX)\hat{\beta} = 0$$

$$\Rightarrow -2(X^TY) = -2(X^TX)\hat{\beta}$$

$$\Rightarrow (X^TY) = (X^TX)\hat{\beta}$$

$$\Rightarrow (X^TX)^{-1}(X^TY) = (X^TX)^{-1}(X^TX)\hat{\beta}$$

$$\Rightarrow \boxed{\hat{\beta} = (X^TX)^{-1}(X^TY)}$$

Unbiasedness

$$E(\hat{\beta}) = \hat{\beta}$$

$$E\{(X^TX)^{-1}(X^TY)\} \qquad Y = X\beta + u$$

$$\Rightarrow E\{(X^TX)^{-1}(X^TX\beta + X^TU)\}$$

$$\Rightarrow E\{(X^TX)^{-1}(X^TX)\beta + (X^TX)^{-1}X^TU\}$$

$$\Rightarrow E\{\beta + (X^TX)^{-1}X^TU\}$$

$$\Rightarrow \beta + (X^TX)^{-1}X^T\underbrace{E(U)}_{\to 0}$$

since $(X^TX)^{-1}X^T$ is non-stochastic

$\rightarrow \beta$

$$V(\hat{\beta}) = ?$$

from

$$(X^TX)^{-1}(X^TX\beta + X^TU) = \hat{\beta}$$

$$(X^TX)^{-1}(X^TX)\beta + (X^TX)^{-1}X^TU = \hat{\beta}$$

$$\beta + (X^TX)^{-1}X^TU = \hat{\beta}$$

$$\hat{\beta} = \beta + (X^TX)^{-1}X^TU$$

$$\hat{\beta} - \beta = (X^TX)^{-1}X^TU$$

$$V(\hat{\beta}) = E(\hat{\beta} - E(\hat{\beta}))^2$$

$$E(\hat{\beta} - \beta)^2 \Rightarrow E\{(X^TX)^{-1}X^TU\}^2$$

$$E(\hat{\beta} - \beta)^2 = E\{(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T\}$$

$$= E\{(X^TX)^{-1}X^TU)(U^TX(X^TX)^{-1}\}$$

$$= E\{(X^TX)^{-1}X^TUU^TX(X^TX)^{-1}\}$$

$$= (X^TX)^{-1}\cancel{X^T}(X^TX)^{-1}E(UU^T)$$

$$= (X^TX)^{-1}(X^TX)(X^TX)^{-1}\sigma_u^2$$

where $E(UU^T) = \sigma_u^2 I_n$

v-c matrix

$$= I_n (X^TX)^{-1}\sigma_u^2$$

$$\therefore \boxed{\text{Variance of } \hat{\beta} = (X'X)^{-1}\sigma_u^2 I_n}$$

understood

Unbiasedness for $\hat{\beta}_0$ & $\hat{\beta}_1$ in non matrix form.

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$\hat{\beta}_1 = \frac{cov(x_i, u_i)}{v(x_i)}$ or $\frac{cov(X_i, Y_i)}{v(X_i)}$.

$E(\hat{\beta}_1) = \beta_1$

$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$ let $\frac{(X_i - \bar{X})}{\sum (X_i - \bar{X})^2} = k_i$

$\therefore \hat{\beta}_1 = \sum k_i y_i$; where $y_i = (Y_i - \bar{Y})$

now
$E(\hat{\beta}_1) = E(\sum k_i y_i) \Rightarrow \sum k_i (Y_i - \bar{Y})$

$= \sum k_i (\beta_0 + \beta_1 X_i + U_i - \bar{Y})$

$= \beta_0 \sum k_i + \beta_1 \sum k_i X_i + \sum k_i U_i - \sum k_i \bar{Y})$

$= 0 + \beta_1 + \sum k_i U_i - \sum k_i \bar{Y}$

$\hat{\beta}_1 = \beta_1 + \sum k_i U_i - \sum k_i \bar{Y}$

$\hat{\beta}_1 = \beta_1 + \sum k_i U_i - \cancel{\bar{Y} \sum k_i}^{0}$

$\therefore E(\hat{\beta}_1) = \beta_1 + \sum \cancel{k_i} E(U_i)^{0}$

Variance

$V(\hat{\beta}_1) = E(\hat{\beta}_1 - \beta_1)^2$

$\rightarrow E(\hat{\beta}_1 - \beta_1)^2 = E(\sum k_i U_i)^2$ where $\frac{\sum (X_i - \bar{X})(U_i - \bar{U})}{\sum (X_i - \bar{X})^2} = \sum k_i U_i$

$\rightarrow E(\sum k_i^2 \sum U_i^2) + 2 \sum \cancel{U_i U_j}^{0} k_i$

$\rightarrow \sum k_i^2 \, E(U_i)^2 \quad 2k \sum E(\cancel{U_i U_j})^{0}$ → no autocorrel^n.

$\rightarrow \frac{\sigma_u^2}{\sum x_i^2}$

?, $\Rightarrow \frac{\sigma_u^2}{SST_x}$

Revise

Matrix Derivation of Efficiency to prove BLUE.

We now, $\hat{\beta} = (X'X)^{-1}X'Y$

To prove $V(\hat{\beta}) = \sigma^2 (X'X)^{-1}$ is minimum

for that lets suppose a $\hat{\beta}^* = (X'X)^{-1}X'Y + CY$ be another estimator B.

Now, Unbiasedness of $\hat{\beta}^*$

$$E(\hat{\beta})^* = E((X'X)^{-1}X'Y + CY)$$

$$= E((X'X)^{-1}X'(X\beta + U) + C(X\beta + U))$$

$$= E((X'X)^{-1}X'X\beta + (X'X)^{-1}X'U + CX\beta + CU)$$

$$= \beta + (X'X)^{-1}E(U) + E(CXB) + E(CU)$$

Assumption orthogonality or $X'C' = 0$

so, iff $CX = 0$, then $E(\hat{B})^* = \beta$.

$$= \beta \pm CE(U)$$

Variance of $\hat{\beta}^*$

$$V(\hat{\beta}^*) = E(\hat{\beta}^* - E(\hat{\beta}^*))^2$$

Now $\hat{\beta}^* = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'U + CX\beta + CU$

$$\hat{\beta}^* = \beta + (X'X)^{-1}X'U + CX\beta + CU$$

$$\rightarrow \hat{\beta}^* - \beta = (X'X)^{-1}X'U + CX\beta + CU$$

$$\rightarrow V(\hat{\beta}^*) = E(\hat{\beta}^* - \beta)^2$$

$$\rightarrow E\{(\hat{\beta}^* - \beta)(\hat{\beta}^* - \beta)^T\}$$

$$= E\{[(X'X)^{-1}X'U + CX\beta + CU][U'X(X'X)^{-1} + \beta'X'C' + U'C']\}$$

$$= E\{[(X'X)^{-1}X'U + CU][U'X(X'X)^{-1} + U'C']\}$$

$$= E[(X'X)^{-1}X'UU'X(X'X)^{-1} + (X'X)^{-1}X'UU'C' + CUU'X(X'X)^{-1} + CUU'C']$$

$$= (X'X)^{-1}\sigma^2 + (X'X)^{-1}\sigma^2 + 0 + \sigma^2 C^2$$

$$= (X'X)^{-1}\sigma^2 + \sigma^2 C^2$$

$$= V(\hat{\beta}) + \sigma^2 C^2$$

$$\therefore V(\hat{\beta}^*) > V(\hat{\beta})$$

Hence its not least.

BLUE in NON Matrix form. (why $\hat{\beta}_1$ is the most efficient estimator of $\hat{\beta}$)

$$\hat{\beta}_1 = \frac{\sum (Y_i - \bar{Y})(X_i - \bar{X})}{\sum (X_i - \bar{X})^2} \quad \text{or } \sum K_i Y_i \text{ for } K_i = \frac{x_i}{\sum x_i^2},$$

$$\hat{\beta}_1 = \sum k_i y_i \qquad y_i = (Y_i - \bar{Y})$$

$$\sum k_i = 0$$

Let $\hat{\beta}_1 = \sum w_i y_i$ $\qquad \sum k_i X_i = 1$ (with x_i written above k_i)

where $w_i = k_i + c_i$ $\qquad \frac{1}{\sum x_i^2} = \sum k_i^2$

$\sum w_i y_i$

$$= \sum w_i (Y_i - \bar{Y})$$

$$= \sum w_i \{(\beta_0 + \beta_1 X_i + U_i) - \bar{Y}\}$$

$$= \sum (w_i \beta_0 + w_1 \beta X_i + w_i U_i - w_i \bar{Y})$$

$$\rightarrow \sum w_i \beta_0 + w_1 \beta_1 X_i + w_i U_i - w_i \bar{Y}$$

$$\Rightarrow \beta_0 \sum w_i + \beta_1 \sum w_i X_i + \sum w_i U_i - \bar{Y} \sum w_i$$

now,

Assumption ÷ i) $\sum w_i = 0$ & if $\sum w_i = 0 \Rightarrow \sum c_i = 0$

ii) $\sum w_i X_i = 1 \Rightarrow \sum c_i X_i = 0$

$\beta_i \ \sum (k_i + c_i) \ X_i$

$\hat{} = \beta_i \ \sum (k_i X_i + c_i X_i)$

$= \beta_i \cdot 1 + \cancel{\sum c_i X_i}^{\,0}$

<u>Variance</u>

$\hat{\beta}_1^* = \beta_1 + \sum w_i U_i$

$\hat{\beta}_{11}^* - \beta = \sum w_i U_i$

$V(\hat{\beta}_1^*) = E(\hat{\beta}_1^* - \beta_1)^2$

$= E(\sum w_i U_i)^2$

$E\left(\sum (x_i + c_i) u_i)\right)^2$

$\Rightarrow E\{\sum k_i U_i + c_i U_i\}^2$

? $\Rightarrow E\{\sum k_i^2 U_i^2 + \sum c_i^2 U_i^2 + 2\sum k_i c_i U_i U_j$

$\sum U_i E(U_i^2) + \sum c_i^2 E(U_i^2) + 2\sum \cancel{k_i c_i} \overset{0}{E(U_i, U_j)}$

$\Rightarrow \dfrac{\sigma^2}{\sum x^2} + \sigma^2 \sum c_i^2$

$\Rightarrow \therefore V(\overset{*}{\beta}_1) > V(\hat{\beta}_1)$ ($\hat{\beta}_1$ is BLUE)

<u>Derivation. of α or β.</u>

$\beta_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

or $= (\beta_0 + \beta_1 \bar{X}_i + \bar{U}_i) - \hat{\beta}_1 \bar{X}$

$= \beta_0 + \bar{X}_1 (\beta_1 - \hat{\beta}_1) + \bar{U}_1$

or

$\hat{\beta}_0 = \beta_0 - \{\beta_1 - E(\hat{\beta}_1)\}\bar{X}_i + \bar{U}_i\}$ main condition

Unbiasedness of $\hat{\beta}_0$

$$E(\hat{\beta}_0) = \beta_0 - E\{\beta_1 - E(\hat{\beta}_1)\} - E(\bar{X}_1) + \cancel{E(\bar{U}_i)}^{0}$$

$$= \beta_0 - \{\cancel{E(\beta_1) - \beta_1}\}^{0} X_1$$

$$= \beta_0$$

Variance of $\hat{\beta}_0$

$$\hat{\beta}_0 = \beta_0 - \{\beta_1 - E(\hat{\beta}_1)\}\bar{X}_1 + \bar{U}_i$$

$$V(\hat{\beta}_0) = E(\hat{\beta}_0 - \beta_0)^2$$

or

$$\Rightarrow E\{-\{\beta_1 - E(\hat{\beta}_1)\}\bar{X}_i + \bar{U}_i\}^2$$

or

$$\Rightarrow E\{\bar{U}_i - \{\beta_i - E(\hat{\beta}_i)\bar{X}_i\}^2$$

$$\Rightarrow E\{\bar{U}_i^2 + \{\beta_1 - E(\hat{\beta}_1)\}^2 \bar{X}_i^2 + 2(\beta_1 - E(\hat{\beta}_1)\bar{X}_i \bar{U}_i\}$$

$$\Rightarrow E(\bar{U}_i)^2 + E\{\beta_1 E(\hat{\beta}_1)\}\bar{X}_i^2 + 2\bar{X}_1 E\{\beta_i - E(\hat{\beta}_i)\}$$

$$\Rightarrow E\left(\frac{\Sigma U_i^2}{x_i^2}\right) + \bar{X}_i^2 E\{\beta_1 - E(\beta_1)\}^2 + 2\bar{X}_i E\{\bar{U}_i(\beta_1 - E(\hat{\beta}_1))\}$$

$$\Rightarrow \frac{1}{n^2}\Sigma E(U_i)^2 + \bar{X}_1^2 E(\beta_1 - E(\hat{\beta}_1))^2 + 2\bar{X} E\{\bar{U}_i(\beta_1 - E(\hat{\beta}_1)\}$$

$$\Rightarrow \frac{\sigma^2}{x} + \bar{X}_1^2 V(\hat{\beta}_1) + 2\bar{X}_i E\{\bar{U}_i(\beta_1 - E(\hat{\beta}_1)\}$$

if $E(\bar{U}_i(\beta_1 - E(\hat{\beta}_1)) = 0$ then

$$V(\hat{\beta}_0) = \frac{\sigma^2}{x} + \bar{X}_i \frac{\sigma^2}{\Sigma(x_i - \bar{x})^2}$$

$$\Rightarrow \sigma^2\left\{\frac{1}{x} + \frac{\bar{X}_i^2}{\Sigma(x_i - \bar{x})^2}\right\} = V(\hat{\beta}_0)$$

To prove

$$E(\bar{u}_i(\hat{\beta}_1 - E(\hat{\beta}))) = 0$$

Since $\hat{\beta}_1 - \beta_1 = \dfrac{cov(x_i, u_i)}{V(x_i)}$

$$\Rightarrow E(\bar{u}_i\{\hat{\beta}_1 - \beta_1\}) = 0$$

$$= E(\bar{u}_i \sum k_i u_i) = 0$$

$$\Rightarrow E\left(\frac{\sum u_i}{\cdot I} \frac{\sum(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} u_i\right)$$

Since $\dfrac{1}{\sum(x_i - \bar{x})^2}$ are non stochastic

$$\Rightarrow \frac{1}{\sum(x_i - \bar{x})^2} E\left(\sum u_i \sum(x_i - \bar{x}) u_i\right)$$

iff $E(\sum u_i \sum(x_i - \bar{x}) u_i) = 0$, then $E(u_i(\hat{\beta}_1 - E(\hat{\beta}_1))) = 0$

$$\Rightarrow E\{\sum u_i^2 (x_i - \bar{x}) + \sum u_i u_j (x_i - \bar{x})\}$$

$$\Rightarrow \{\sum E(u_i^2)(x_i - \bar{x}) + \sum(x_i - \bar{x}) E(u_i u_j)\}$$

↳ no autocorr-elatn

$$\Rightarrow \cancel{\sigma_u^2 \sum(x_i - \bar{x})} + 0$$

↳ sum of std deviatn from mean

$$\Rightarrow 0$$

Residual Maker Matrix / Projection Matrix / orthogonality conditions

Residual

$U = Y - \hat{Y} \quad \hat{\beta} = (X'X)^{-1}X'Y$

$Y = X\beta + U$

$\hat{Y} = X\hat{\beta}$

$U = Y - \hat{Y}$

$\Rightarrow Y - \cancel{X}\hat{\beta}$

$\Rightarrow Y - X\{(X'X)^{-1}X'Y\}$

$\Rightarrow Y - X(X'X)^{-1}X'Y$

$U \Rightarrow [I - X(X'X)^{-1}X']Y$

Let $M = [I - X(X'X)^{-1}X']$

$U = MY$

Projection

$M = [I - X(X'X)^{-1}X']$

now, $U = MY$

& $U = Y - \hat{Y}$

$\Rightarrow \hat{Y} = Y - u$

$\Rightarrow \hat{Y} = Y - MY = [I - M]Y$

$\Rightarrow \hat{Y} = [I - I + X(X'X)^{-1}X']Y$

$\Rightarrow \hat{Y} = [\underbrace{X(X'X)^{-1}X'}_{P}]Y$

Symmetric $= P = P^T$

$P^T = [X(X'X)^{-1}X']^T$

$= X(X'X)^{-1}X' = P$

Idempotent

$P.P = P \Rightarrow [X(X'X)^{-1}X'][X(X'X)^{-1}X']$

$= X(\underbrace{X'X)^{-1}X'X}_{In}(X'X)^{-1}X'$

$\Rightarrow X(X'X)^{-1}X'$

orthogonality

$PX = X(X'X)^{-1}X'X = X \neq 0$

- Properties of M

1) Idempotent $(M.M = M)$

2) Symmetric $(M^T = M) \rightarrow M^T = [I - X(X'X)^{-1}X']^T$

$= [I - X(X'X)^{-1}X']$

$= M$

3) orthogonality $(MX = 0)$

$\rightarrow M.M = [I - X(X'X)^{-1}X'][I - X(X'X)^{-1}X']$

$= I - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X\underbrace{(X'X)^{-1}X'X}_{I}(X'X)^{-1}X'$

$\Rightarrow I - 2X(X'X)^{-1}X' + X(X'X)^{-1}X'$

$\Rightarrow I - X(X'X)^{-1}X' = M$

$\rightarrow [I - X(X'X)^{-1}X']X = X - X[X'X]^{-1}X'X \rightarrow X - X = 0$

- The orthogonality condition to prove if $\hat{\beta}_1 \gtreqless \hat{\beta}_1^*$ & $\hat{\beta}_2 \gtreqless \hat{\beta}_2^*$

let $Y_1 = \beta_0 + \hat{\beta}_1 x_1 + u_1$

$Y_2 = \beta_0 + \hat{\beta}_2 x_2 + U_2$ be 2 Univariate regressions.

? & · $Y_{1i} = \beta_0 + \hat{\beta}_1 x_{11} + x\hat{\beta}_2 x_2 + u$ ⟶ multivariate case.

Under Matrix format (Univariate case).

$$\underset{Y_i}{\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}_{2\times1}} = \underset{X_i}{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}_{2\times2}} \underset{\beta_i}{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}} + \underset{U_i}{\begin{bmatrix} U_1 \\ U_2 \end{bmatrix}_{2\times1}}$$

or (multivariate case.)

?

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ x_2 & x_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$$

we have to show whether

$$\hat{\beta}_1 = (x'x)^{-1} x' Y_1$$

$$\hat{\beta}_2 = (x'x)^{-1} x' Y_2$$

we have to see whether.

$\hat{\beta}_1 = \hat{\beta}_1^*$ (Univariate case $\hat{\beta}_1$ = multivariate $\hat{\beta}_1^*$

$\hat{\beta}_2 = \hat{\beta}_2^*$ (" ")

or

$$\hat{\beta}_1 \gtrless \hat{\beta}_1^*$$

$$\hat{\beta}_2 \gtrless \hat{\beta}_2^*$$

Estimated multivariated regression matrix.

$$\Rightarrow \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & X_2 \\ X_1 & X_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

$\Rightarrow$ Multiplying X_1^T & X_2^T in 1st & 2nd

$$\begin{bmatrix} X_1^T X_1 \\ X_2^T Y_2 \end{bmatrix} = \begin{bmatrix} X_1^T X_1 & X_1^T X_2 \\ X_1^T X_1 & X_2^T X_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

$$= \begin{bmatrix} X_1^T Y_1 \\ X_2^T Y_2 \end{bmatrix} = \begin{bmatrix} X_1^T X_1 \hat{\beta}_1 + X_1^T X_2 \hat{\beta}_2 \\ X_2^T X_1 \hat{\beta}_1 + X_2^T X_2 \hat{\beta}_2 \end{bmatrix}$$

$$\Rightarrow X_1^T X_1 \hat{\beta}_1 + X_1^T X_2 \hat{\beta}_2 = X_1^T Y_1$$

$$X_2^T X_1 \hat{\beta}_1 + X_2^T X_2 \hat{\beta}_2 = X_2^T Y_2$$

Now

$$(X_1^T X_1)\hat{\beta}_1 = X_1^T Y_1 - X_1^T X_2 \hat{\beta}_2$$

$$(X_1^T X_1)^{-1}(X_1^T X_1)\hat{\beta}_1 = (X_1^T X_1)^{-1}\{X_1^T Y_1 - X_1^T X_2 \hat{\beta}_2\}$$

$$\hat{\beta}_1^* = \underbrace{(X_1^T X_1)^{-1}(X_1^T Y_1)}_{\beta_1} - (X_1^T X_1)^{-1}\overset{0}{\cancel{(X_1^T X_2)}}\hat{\beta}_2$$

$$\hat{\beta}_1^* = (X_1^T X_1)^{-1}(X_1^T Y_1)$$

$$= \hat{\beta}_1$$

now $(X_1^T X_2) = 0$ {serves as correlation btn X_1 & X_2}

iff X_1, X_2 corr $= 0$, i.e, X_1 X_2 are independent

then only $\hat{\beta}_1$ is free of multicollinearity.

Similarly

$$\hat{\beta}_2^* = \hat{\beta}_2 - (X_2^T X_2)\underbrace{(X_2^T Y_1)}_{\to 0}\hat{\beta}_1$$

$$\hat{\beta}^* = \hat{\beta}_2.$$

- **Problem of Heteroskedasticity**

Heteroskedasticity occurs when the variance of errors is not constant across all levels of independent variables.

It can occur due to

→ model misspecification
↳ ommited variable, incorrect functional form.

→ presence of outliers.

Now,

How to check if heteroskedasticity is present ?

→ BP test
→ white Test [WHITE TEST]
→ GOLDFIELD QUANTOM TEST.

- **BP Test**

Suppose we want to check if the following model

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad — (1)$$

has heteroskedasticity problem.

To apply BP TEST, we run OLS on (1)

we estimate $\hat{\beta}_0$ & $\hat{\beta}_1$ from OLS process through which Y_i can be derived as $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ & from Y_i

now

$\hat{U}_i$ can be derived from $Y_i - \hat{Y}_i$

Here, $\hat{U}_i^2$ can be expressed as a regression model.

Such that

$$\hat{U}_i^2 = \delta_0 + \delta_1 X_i + V_i \quad \text{--- (II)}$$

↳ error term.

we run OLS on eqⁿ (II) & estimate F calculated value.

$$F_{cal} = \frac{MSE}{MSR} = \frac{ESS/K}{RSS/n-k-1} = \frac{ESS}{K} \cdot \frac{n-k-1}{RSS}$$

if $F_{cal} > F_{tab}$, we reject null hypothesis.

our null hypothesis : $H_0 : \delta_1 = 0$ {i.e coeff. of $X_i = 0$} homoskedastic

$H_1 : \delta_1 \neq 0$ {heteroskedasticity}.

if null is rejected

$\hat{U}_i^2$ becomes the function of X_i

WHITE TEST (A better version of BP test).

→ $\hat{U}_i^2$ may be a non-linear function of X_j, not only a linear function of X_i, hence can put a forceful homoskedasticity result..

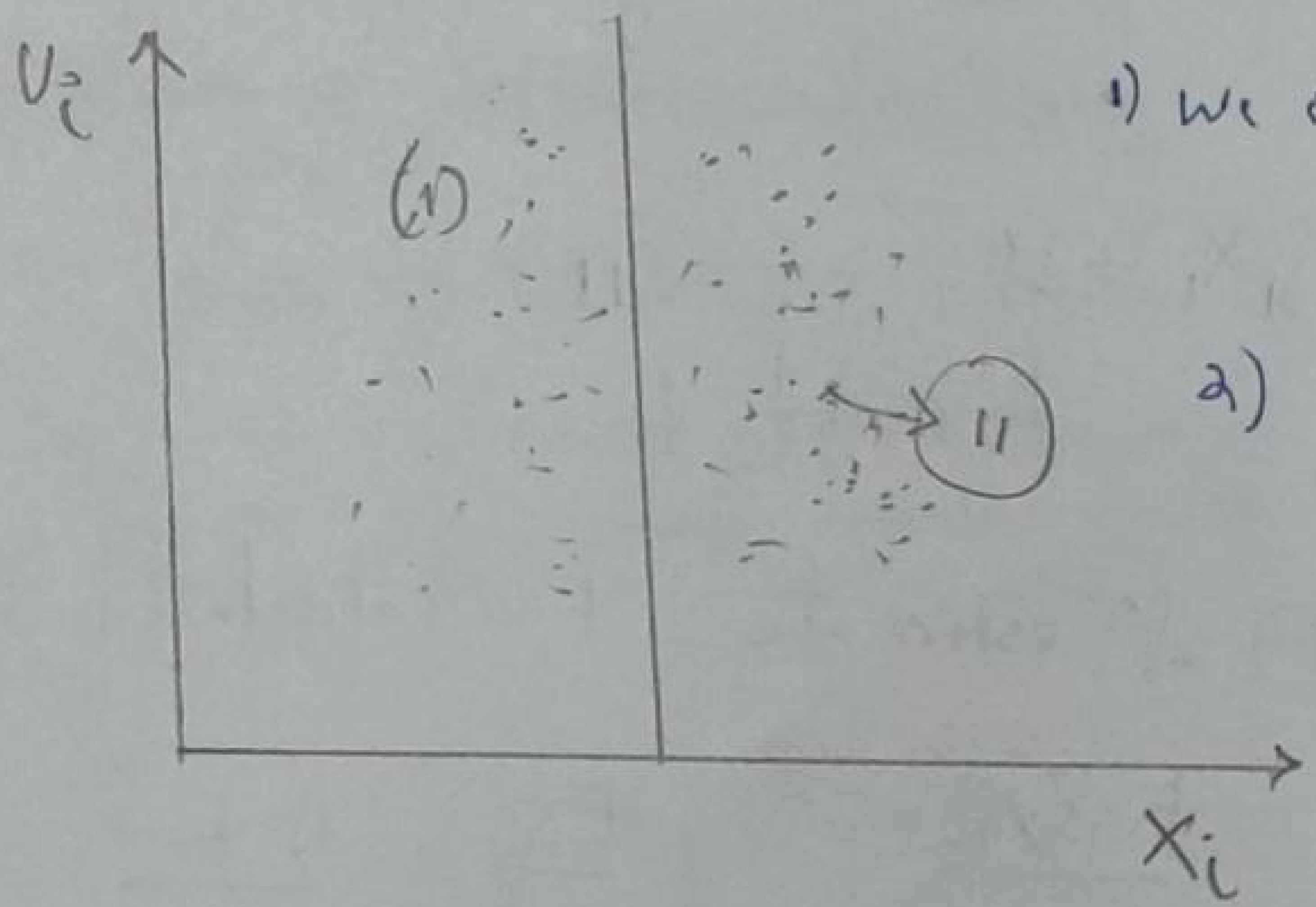
WHITE TEST

Step - 1 → Same

Step - 2 → $\hat{U}_i^2 = \delta_0 + \delta_1 X_i + \delta_2 X_i^2 + U_i$ ~ U_i is non-linear form of X_i

$$F = \frac{ESS/K}{RSS/n-K-1} = \frac{\frac{ESS}{TSS}/K}{\frac{RSS}{TSS}/n-K-1} = \frac{R^2/K}{1-R^2/n-k-1}$$

- **GOLD FIELD QUANTUM TEST**



1) we observe a positive relationship of U_i

2) we divide the dataset into 2 segments.

3) we run OLS seperately for each of these two segments

4) obtain RSS from segment 1 & 2 i.e., RSS_1 & RSS_2.

5) Use F statistic i.e $F = \frac{RSS_2/df_1}{RSS_1/df_2}$

with hypothesis ; H_0 : Homo i.e, $\delta_1 = 0$

H_2 : Hetero i.e., $\delta_1 = 0$

? if $F > F_{critical}$, reject H_0

- **AUTOCORRELATION**

A VIOLATION OF OLS ASSUMPTION IS AUTOCORRELATION PROBLEM.

Here $Cov(U_i, U_j) \neq 0 \;\forall\; i \neq j$

It can occur in timeseries or cross section dataset

Under timeseries ; it is called serial autocorrelation

Under <u>cross section</u>, it is called <u>spatial correlation</u>.

Now, the problem of autocorrelation effects the efficiency of a OLS estimator β, making it not the least variance estimator.

Under V-C Matrix $= E(UU^T) = E\begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix}[U_1 \cdots U_n] = E\begin{bmatrix} U_1U_1 & U_1U_2 & \cdots \\ \vdots & & \vdots \\ \vdots & & \vdots \end{bmatrix}$

$$= \begin{bmatrix} E(U_1U_1) & \cdots & E(U_1U_n) \\ \vdots & \ddots & \vdots \\ \vdots & & \vdots \\ \cdots & \cdots & E(U_nU_n) \end{bmatrix}$$ for $cov(U_i, U_j) = \sigma_U^2 \; \forall i = j$

but if $cov(U_i, U_j) \neq 0 \; \forall i \neq j$

then $V(\hat{\beta}) \neq \sigma_u^2 a (X'X)^{-1}$ {hence not efficient}

not being an efficient estimator, hence affects the process of hypothesis test methods, like F test & T-test

since standard Error $=$ SE $= \sqrt{V(\hat{\beta})}$

not having a least variance Estimator causes the ~~out~~ outcome of calculated test statistics to fall in case of t-test.

i.e. $t_{cal} = \dfrac{\hat{\beta}}{\sqrt{SE}}$, if $V(\hat{\beta})\uparrow$, $t_{cal}\downarrow$

hence, if $t_{cal} < t_{tab}$, we fail to reject null hypothesis.

Autocorrelⁿ in cross sectⁿ dataset

$$Y_i = \beta_0 + \beta_1 x_i + U_i$$

Y_i = consumptⁿ

X_i = income

U_i = (age, locatⁿ, religion, price, etc)

&

$$Y_1 = \beta_0 + \beta x_1 + U_1$$
$$Y_2 = \beta_0 + \beta x_2 + U_2$$

Both error terms can have commodities with each other in terms of the hidden Variable stored in them.

In such case,

$cov(U_i, U_j) \neq 0 \ \forall i \neq j$

This is hence called spatial correlation.

How to detect AUTOCORRELATION

DW TEST {Durbin Watson Test}

AUTOCORRELATION IN TIMESERIES DATA

Model of time series

$$Y_t = \beta_0 + \beta_1 x_t + U_t \quad \forall t = 0, 1, \ldots$$

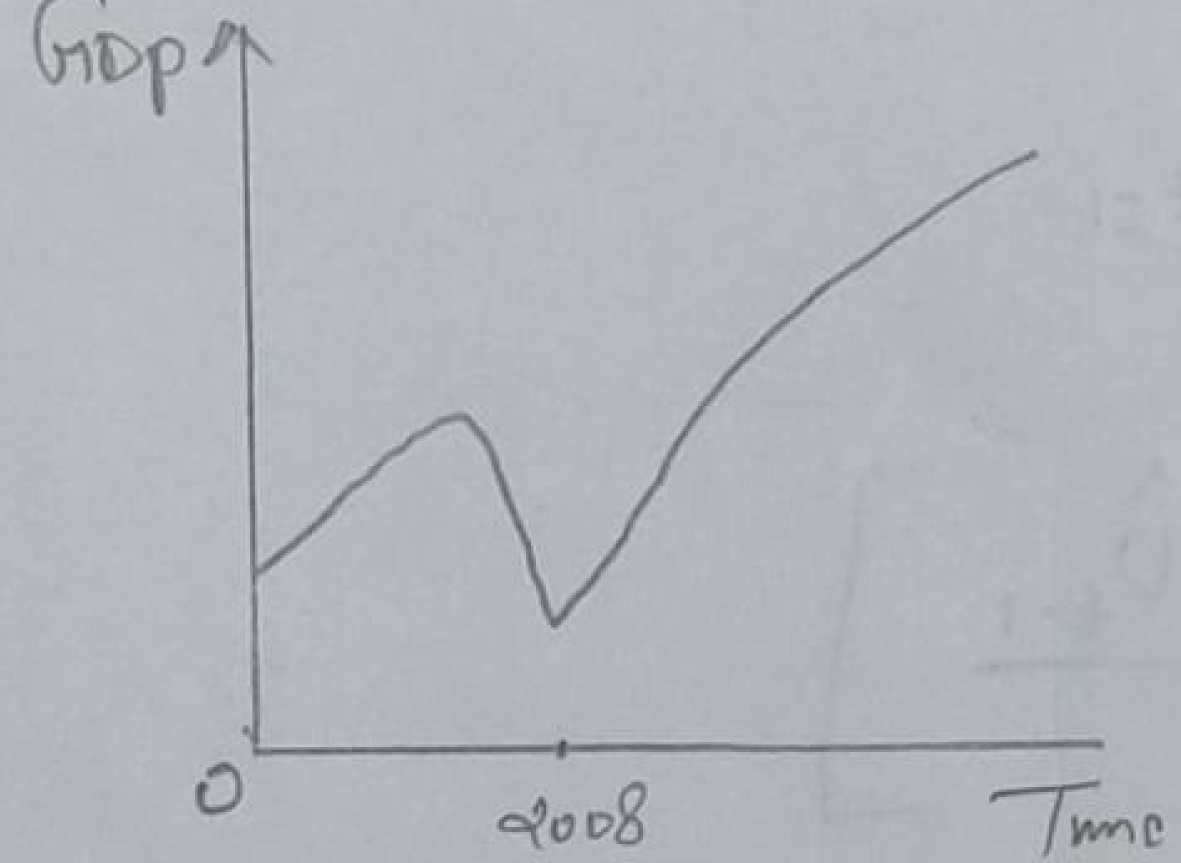
Autocorr. of time series dataset $cov(U_t, U_{t-1}) \neq 0$

This can occur due to

→ model misspecification.

→ External shock

eg: due to 2008 crash, GDP growth suffered through a decline.



So; $cov(U_{2008}, U_{2009}) \neq 0$

$cov(2U_{2008}, U_{2010}) \neq 0$

⋮

i.e., all error terms post 2008 are somehow affected due to error term of 2008.

Hence one can conclude, Under autocorr.

$\boxed{U_t = f(U_{t-1})}$ {and under Autoregressiveness $X_t = F(X_{t-1})$}

sub. to

$U_t = \rho_0 + \rho_1 U_{t-1} + V_t$ {this is different from model of heteroskedasticity note}

To test for autocorrelation, we run DW Test Statistics

$$d = 2(1-\rho)$$

Prove.

$$d = DW = \frac{\Sigma(\hat{U}_t - \hat{U}_{t-1})^2}{\Sigma \hat{U}_t^2} = 2[1-\rho]$$

$$d = \frac{\Sigma \hat{U}_t^2 + \Sigma \hat{U}_{t-1}^2 - 2\Sigma \hat{U}_t \hat{U}_{t-1}}{\Sigma \hat{U}_t^2}$$

now

if $t \to \infty$; $\sum \hat{U}_t^2 = \sum \hat{U}_{t-1}$ { law of large no) ?

CLT

$$d = \frac{2\sum \hat{U}_t^2 - 2\sum \hat{U}_t \hat{U}_{t-1}}{\sum \hat{U}_t^2}$$

or

$$d = \frac{2\sum \hat{U}_{t-1}^2 - 2\sum \hat{U}_t \hat{U}_{t-1}}{\sum \hat{U}_{t-1}^2}$$

$$\Rightarrow \quad = 2\left[\frac{\sum \hat{U}_{t-1}^2 - \sum \hat{U}_t \hat{U}_{t-1}}{\sum \hat{U}_{t-1}^2}\right]$$

$$\Rightarrow \quad = 2\left[1 - \frac{\sum \hat{U}_t \hat{U}_{t-1}}{\sum \hat{U}_{t-1}^2}\right] = 2\left[1-\rho\right]$$

let $\rho = \dfrac{\sum \hat{U}_t U_{t-1}}{\sum \hat{U}_{t-1}^2} \approx \dfrac{\sum x_i y_i}{\sum x_i^2}$

if $\rho = 1 \Rightarrow d = 0 \rightarrow$ (perfect autocorrl^n)
if $\rho = 0 \Rightarrow d = 2 \rightarrow$ (No autocorr)
if $\rho = -1 \Rightarrow d = 4$ (perfect -ive autocorr)

Internal-3

- (1) Var Model → why VAR over ARIMA → lag selection
  → steps to go for VAR → to satisfy VAR
  (3) → Problems with VAR → BVAR
  (2) → cointegration → what is cointegration
  → How to do or check.
  → VEC.
  (4) → VAR & Cointegration
  (5) → Granger causality → Pair wise
  → individual
  (6) → ARDL → (7) SVAR → improved over VAR
  → impulse response
  → Variance decomposition
  → Shocks.

- Better analysis through software (R)

- Analysis of single time series Variable → AR
  → MA
  → ARIMA
  → ARMA

However analysis of multivariate or simple Bivariate regression timeseries model need more complex model operations

- → VAR
- → BVAR
- → VARMA
- → ARDL
- → SVAR

for explaining and analysis of simultaneously related variables (with lags) in time series where we take into account interdependence; we estimate the simultaneous equation. Such model is called dynamic simultaneous equation model.

↳ The method of doing so i.e, to explain the interdependence between two endogenous variables in a simultaneous equation model; is done by Vector Autoregressive Model (VAR). VAR is used as it can run OLS estimation.

eg: 3 endogenous variables i.e, y_t, x_t, I_t will have 3 eqns with their respective lags.

NOTE

$y_t = \beta y_{t-1} + \epsilon_t$ is not bivariate model. It is Univariate model with Y_t's lag. This is AR(1)

However

$y_t = \beta_1 y_{t-1} + \beta_2 X_{t-1} + \epsilon_t$ is multivariate model with Y_t's 1st lag & 2nd lag of some other variable. This is still AR(1)

However

$y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 X_{t-1} + \epsilon_t$ is a multivariate, AR(2) model

↳ because it has 2 different Variables Y_t & X_t interacting & Y_{t-1} & Y_{t-2} as Variables

Why VAR ?

Suppose: Y_{1t} represent money supply

Y_{2t} represents GNP deflator

such

$$Y_{1,t} = \alpha_{11} Y_{1,t-1} + \alpha_{12} Y_{2,t-1} + \epsilon_{1t}$$

$$Y_{2,t} = \alpha_{21} Y_{2,t-1} + \alpha_{22} Y_{1,t-1} + \epsilon_{2t}$$

$Y_{i,t-j}$ ↗ represents ith

$i = 1, 2, 3 \ldots$

$j = 1, 2, \ldots$

↳ ~~No~~ time lag

Here

$Y_{1,t}$ & $Y_{2,t}$ are 2 endogenous Variables.

If there are more than 2 endogeneous Variables; then suppose there are are k endogeneous Variables & lags. then it would be more complicated.

But for simplicity we consider our equation to be 2 endogeneous Variables with AR(1)

2 Unknowns, hence 2 eqns.

$\therefore$ Money supply ⟵ $Y_{1,t} = \alpha_{01} + \alpha_{11} Y_{1,t-1} + \alpha_{12} Y_{2,t-1}, \epsilon_{1,t}$

$Y_{2,t} = \alpha_{02} + \alpha_{21} Y_{2,t-1} + \alpha_{22} Y_{1,t-1} + \epsilon_{2,t}$

↓

GNP deflator

Then to solve this, we represent through Matrix method

$$\underset{2\times2}{\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix}} = \underset{2\times1}{\begin{bmatrix} \alpha_{01} \\ \alpha_{02} \end{bmatrix}} + \underset{2\times2}{\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}} \underset{2\times1}{\begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix}} + \underset{2\times1}{\begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}}$$

we can write this system of equation in the form of lag L Operators.

i.e., $Y_{1,t-1} = L\,Y_{1,t}$

$Y_{1,t-2} = L^2\,Y_{1,t}$

$\vdots$

So on.

$$\therefore \underset{2\times1}{\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix}} = \underset{2\times1}{\begin{bmatrix} \alpha_{01} \\ \alpha_{02} \end{bmatrix}} + \underset{2\times2}{\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}} \underset{2\times1}{\begin{bmatrix} LY_{1,t} \\ LY_{2,t} \end{bmatrix}} + \underset{2\times1}{\begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}}$$

In compressed form;

$$\underset{\substack{\downarrow \\ \text{Endogenous} \\ \text{Var. matrix}}}{[Y]} = \underset{\substack{\downarrow \\ \text{intercept} \\ \text{matrix}}}{[K]} + \underset{\substack{\downarrow \\ \text{coeff.} \\ \text{matrix}}}{[\beta]} \underset{\substack{\downarrow \\ \text{lag} \\ \text{matrix}}}{[L]} + \underset{\substack{\downarrow \\ \text{error} \\ \text{matrix}}}{[\epsilon]}$$

Taking the lag terms to LHS

$$\underset{2\times1}{\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix}} - \underset{2\times2}{\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}} \underset{2\times1}{\begin{bmatrix} LY_{1,t} \\ LY_{2,t} \end{bmatrix}} = \underset{2\times1}{\begin{bmatrix} \alpha_{01} \\ \alpha_{02} \end{bmatrix}} + \underset{2\times1}{\begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}}$$

$$\Rightarrow \begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix}_{2\times1} - \begin{bmatrix} \alpha_{11} L Y_{1,t} + \alpha_{12} L Y_{2,t} \\ \alpha_{21} L Y_{1,t} + \alpha_{22} L Y_{2,t} \end{bmatrix}_{2\times1} = \begin{bmatrix} \alpha_{01} \\ \alpha_{02} \end{bmatrix}_{2\times1} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}_{2\times1}$$

$$\Rightarrow \begin{bmatrix} Y_{1,t} - (\alpha_{11} L Y_{1,t} + \alpha_{12} L Y_{2,t}) \\ Y_{2,t} - (\alpha_{21} L Y_{1,t} + \alpha_{22} L Y_{2,t}) \end{bmatrix} = [K] + [\epsilon]$$

$$\Rightarrow \underbrace{\begin{bmatrix} 1-\alpha_{11} L & -\alpha_{12} \\ -\alpha_{21} & 1-\alpha_{22} L \end{bmatrix}}_{A} \begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = [K] + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = [A]^{-1}_{2\times2} [K]_{2\times1} + [A^{-1}]_{2\times2} [\epsilon]_{2\times1}$$

$$\Rightarrow \begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = [A]^{-1} [K] + \begin{bmatrix} 1-\alpha_{11} L & -\alpha_{12} \\ -\alpha_{21} & 1-\alpha_{22} L \end{bmatrix}^{-1}_{2\times2} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

Now,

$$[A^{-1}] = \frac{\text{Adj } A}{|A|}$$

Adj. A for 2×2: $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = A$

$$\therefore \text{Adj. } A = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$[Y] = [A^{-1}][K] + \frac{\text{Adj } A}{|A|} [\epsilon]$$

$$\Rightarrow [Y] = [A^{-1}][K] + \frac{\begin{bmatrix} 1-\alpha_{22} L & \alpha_{12} \\ \alpha_{21} & 1-\alpha_{11} L \end{bmatrix}}{|A|} [\epsilon]$$

<u>If you</u> <u>observe</u>

$y_{1,t} = f(\epsilon_{1,t}, \epsilon_{2,t})$ only after transformat^n

$y_{2,t} = f(\epsilon_{1,t}, \epsilon_{2,t})$ " "

(These are impulse response functions)

Very important while comparing the VAR & SVAR.

This implies

here $|A| = \begin{vmatrix} 1-\alpha_{11}L & -\alpha_{12} \\ -\alpha_{21} & 1-\alpha_{22}L \end{vmatrix}$

$= (1-\alpha_{11}L)(1-\alpha_{22}L) - \alpha_{12}\alpha_{21}$

$= 1-(\alpha_{11}+\alpha_{22})L + (\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21})L^2$

= Quadratic eq^n (2 roots).

Let these 2 roots are λ_1 & λ_2. Why are these 2 roots important? To check for stability of the model. Given its lags which were taken, is this model stable or not; we have to check. If its unstable; we drop the lags on by 1 to see stability.

<u>STABILITY CONDITIONS</u>

1) $|\lambda_1| < 1$ & $|\lambda_2| < 1$

2) $|A - \lambda I| = 0$ where A is matrix of lag coefficients.

$\hookrightarrow \begin{pmatrix} 1-\alpha_{11}L & -\alpha_{12} \\ -\alpha_{21} & 1-\alpha_{22}L \end{pmatrix}$

Once stability is achieved; $Y_{1,t}$ & $Y_{2,t}$ can be expressed as functions of its current and lagged Values of $\epsilon_{1,t}, \epsilon_{2,t}$. These are known as impulse response functions.

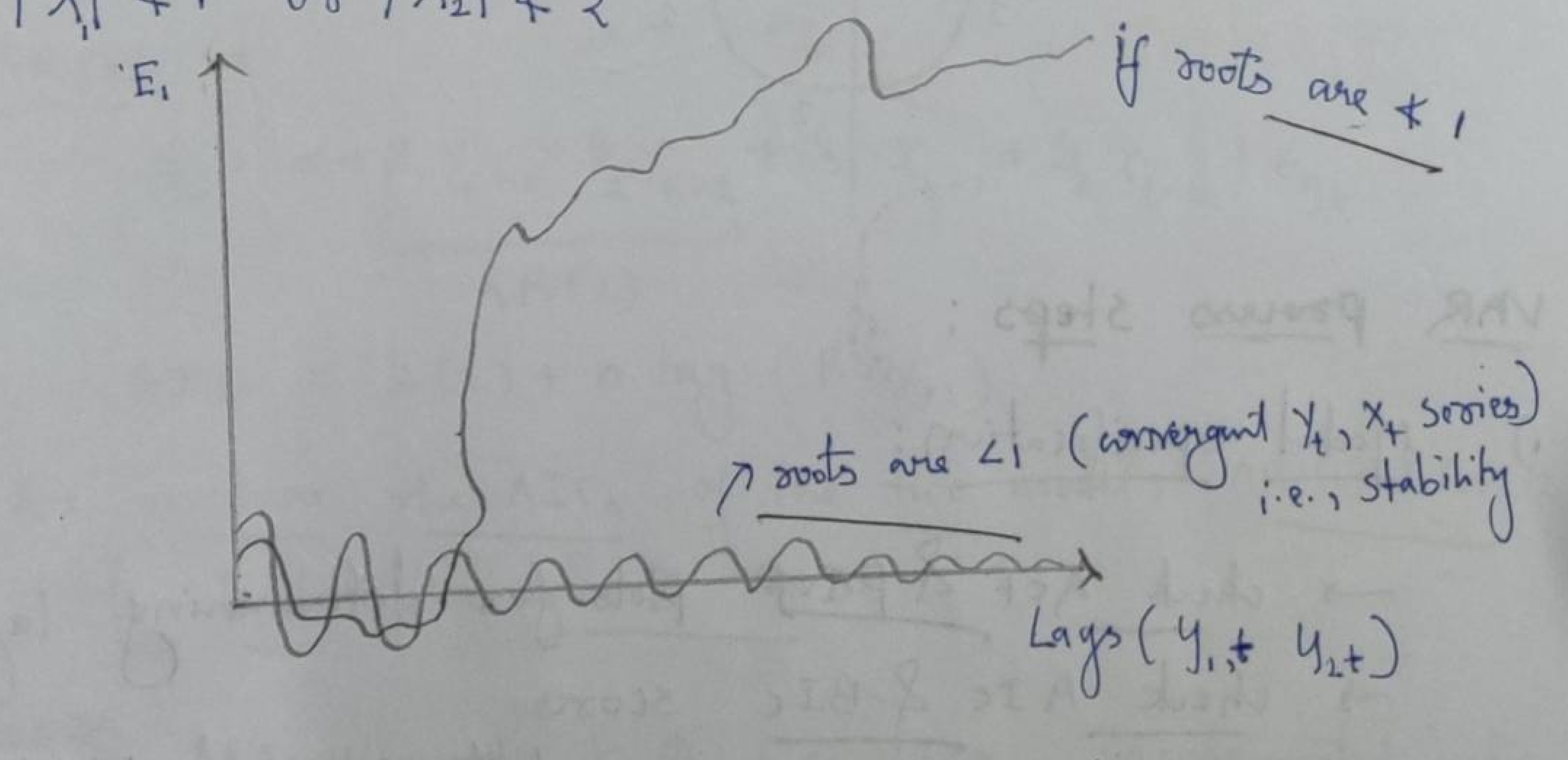
i.e $\boxed{Y_{1t} = [\lambda^{-1}][\epsilon_t]}$

NOTE

Error term has to be white noise for implementation of OLS. white noise → OLS Assumption + stationarity + weak dependency.

Suppose

$|\lambda_1| \nless 1$ or $|\lambda_2| \nless 2$

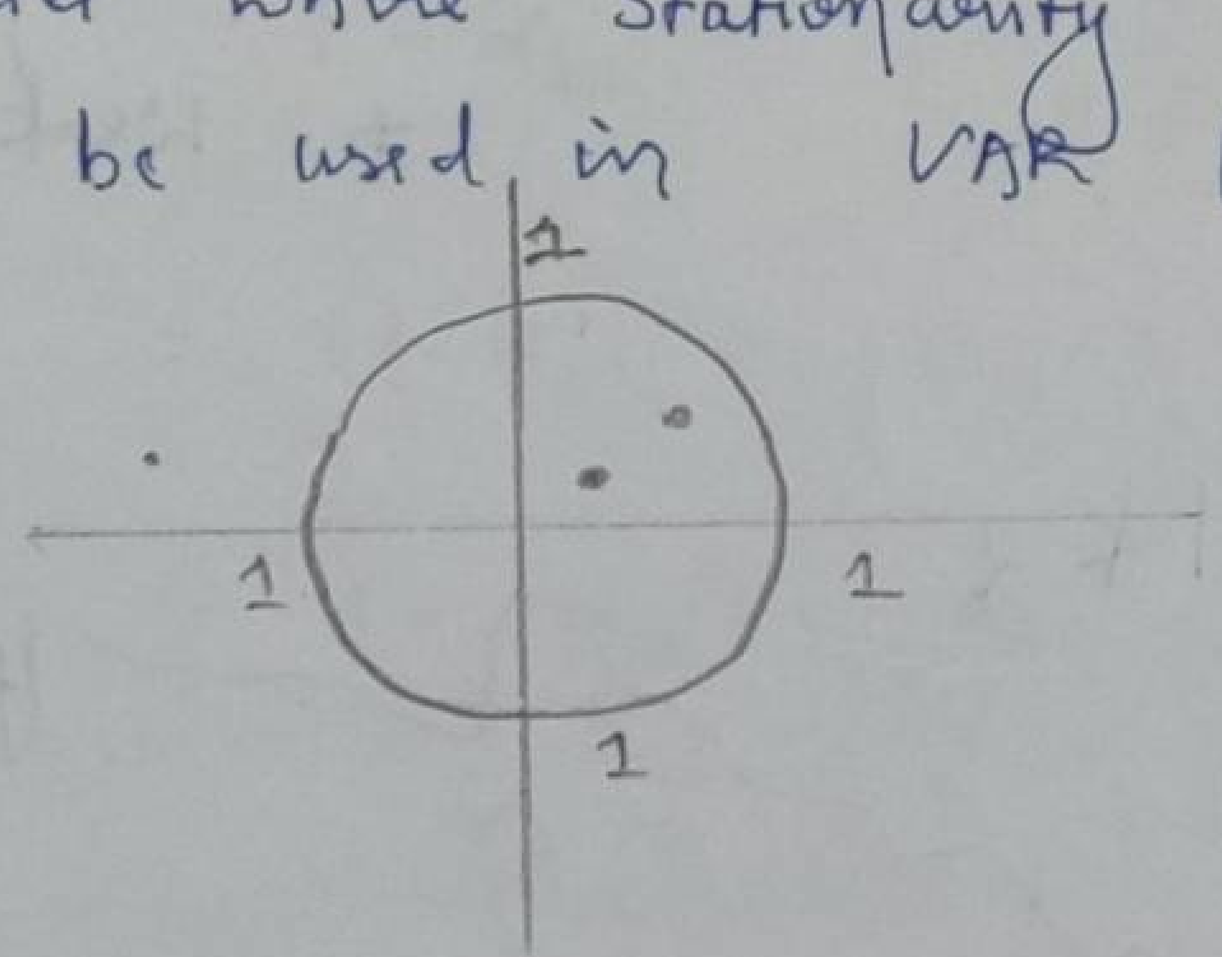


why is error effect when the roots aren't less than 1?

Because if $\lambda_i \nless 1$ $\forall$ $i = 1,2$ then $Y_{1,t}$ & $Y_{2,t}$ are not convergent. If $y_{1,t}$ & $y_{2,t}$ are convergent implies the incorporated shocks {seen through impulse response functn : $y_{1,t} = f(\epsilon_{1t}, \epsilon_{2t})$ and so far y_{2t} fade away over a period of time.

If the model is not converging ; implies the effect of past shocks impact the predictions of future. It means error term is non-stationary. No a non-stationary error term will not follow OLS or ALS. i.e asymptotic least square. So OLS estimation ~~on~~ method in a non-stationary VAR processes cannot be implemented.

① $|\lambda_1| \nless 1$ & $|\lambda_2| \nless 1$ mean ; the model's roots must be inside a Unit root circle. Then only we say a stable model where stationarity is there in Error term can be used in VAR processes.



• **VAR process steps :**

i) Model specification :

→ check ACF & PACF plots for determining lags

→ check AIC & BIC scores

↳ (more better to find right lag model)

$$AIC = 2K + n\log(RSS/n)$$

→ lags

↳ obs.

$$BIC = K\log(n) + n\log(RSS/n)$$

↳ lags ↳ obs

Suppose we have an AR (1) model.

$$\boxed{Y_t = \alpha + \underbrace{\beta_1 y_{t-1}}_{AR(1)} + \beta_2 X_{t-1} + \epsilon_{yt}}$$

And we run OLS
to estimate the parameters α, β_1, β_2 through Min RSS

Now

if we want to check AIC.

$$AIC = 2k + n \log (RSS/n)$$

$$k = 1 \quad (\text{lag} - 1)$$

$$AIC = 2(1) + n \log (RSS/n)$$

AR (2)

$$y_t = \alpha + \underbrace{\beta_1 Y_{t-1} + \beta_2 Y_{t-2}}_{AR(2)} + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \epsilon_{yt}$$

$$AIC = 2(2) + n \log (RSS/n)$$

we compare the AICs of the two models AR (1) & AR(2)
and check the lowest AIC.

However Even though k ↑ i.e. lags incorporated into the model ↑, thus AIC drops as RSS ↓. But this comes at the cost of degrees of freedom

A better method of check model fit is the BIC
i.e., Bayesian information criteria.

$$BIC = k \log (n) + n \log (RSS/n)$$

where k is the lags while as k lags are imported into our eqn; the impact for loss of obs. n-k is taken into account as well through log (n).

2) Maximum integration order of the endogenous Variables is I (1) i.e, y_t is either I (0) or I (1)

↳ stationary time series; ↳ stationary 1st diff.

similarly X_t.

NOTE

If both x_t & y_t are I (0), simple OLS can be applied to estimate parameters easily. No need for VAR. However, if the process.

Cointegration

If two time series are integrated at the same level & their linear combination is stationary, then the two series are cointegrated. co-integration is called co-trend or co-moving.

i.e, if $y_t \to I(1)$

$X_t \to I(1)$

$\therefore\ y_t = \beta x_t + U_t$ then $y_t - \beta x_t = U_t \to I(0)$.

6) Difference b/w correlation & cointegration

correlation is just linear association between two time series.

cointegration → no. of series can be more than 2

↳ needs same order of integration according to

Engel Granger.

- There is no metric to quantify co-integration
- cointegration is a long run concept. word^ is short run or long run analysis.
- cointegration requires a minimum of 20 or more observations
- co-integrat^n applies to time series only.

if $y_t \rightarrow I(0)$; $x_t \sim I(1)$

then linear combination of these 2 will be I(1) as the non-stationary I(1) will denote I(0).

- Prop. of cointegratn.

1) Series have common trends.
2) Series can be represented having moving average.
3) There exist an equillibrium model.

X_t	I(0)	I(1)	I(2)	I(1)	I(0)
y_t	I(0)	I(1)	I(2)	I(2)	I(1)
Possibilit-y of cointegrat^n	OLS	co-int possible	coint possible	not Possi-ble	coint. not possible
confirm-at^n of cointeg-rat^n	"	If error → term is → stationary → then yes	" " " "	X	X

NOTE

if Error — Stationary $\rightarrow I(0)$ → Apply VAR

— NST $\rightarrow I(1)$ → cannot apply VAR → we used VEC, which then can use OLS.

VEC (Vector Error correctn Model)

The VEC restricts the long run behaviour of the endogeneous Variables to converge to their cointegrating relationship. while allowing for short run adjustment. Hence its ~~res~~ restricted version of VAR which allows the use of non-stationary time series that are known to be cointegrated.

The cointegrating term because of which our error becomes stationary is called error correctn term.

VAR & VARMA (Problems with VAR)

Suppose we have 4 Variables and 4 eqn to run OLS → Suppose each Variable has 4 lags.

No. of parameters $= (g + kg^2) = \{4 + 4(4)^2\} = 68$

This is a problem as no. of estimates are too many

In case of VARMA; we consider both AR & MA.

for VARMA (1,1) {1 lag, each}

then ignoring intercepts;

there are 2 parameters to estimate. Suppose there are 4 Variables (endogenous), the 4 Variables, 1 lag each.

$\therefore$ $4 \times 2 = 8$ parameters { this is better than VAR }

but VARMA has more partial derivatives to calculate for its parameters as they can be huge in no.

Hence VAR is having more parameters to estimate $\Rightarrow$ Lower df than VARMA; but VARMA is very difficult to calculate.

GENERAL VAR (4) Eq^n

$$\boxed{Y_t = \sum_{i=1}^{4} \alpha_i \, y_{t-i} + \sum_{i=1}^{4} \beta_i \, X_{t-i} + \epsilon_t}$$

$\downarrow$ Endogenous Variable

$\downarrow$ y_t's Lags

$\downarrow$ endogenous Variable with its lags.

The eq^n has total 8 Unknown ~~parameters~~ Variables to estimate as there are 4 lags for each. 2 Variables.

$\therefore$ no. of parameters $= (g + kg^2)$

$= \{ 8(2)^2 + 2 \}$

$= 34$ param. or restrict^n

concept clearing (Imp)

- If there is an AR (2) eq^n of 2 Variables y_t, x_t

$$y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \alpha_1 X_{t-1} + \alpha_2 X_{t-2}$$

∴ Total no. of parameters or restrictions are

⇒ 4 lags × 2 Variables × 2 eqⁿ.

⇒ 16 restrictions (without intercepts)

- Suppose AR(6) & 4 Variables

$$y_t = \sum_{i=1}^{6} x_{t-i} + \sum_{i=1}^{6} y_{t-i} + \sum_{i=1}^{6} z_{t-i} + \sum_{i=1}^{6} e_{t-i} + \epsilon_t$$

6 lags × 4 Variables × 4 eqⁿˢ.

⇒ 16 × 6 = 96 parameters (without intercept)

Variables are considered here as Endogeneous Variables; not lag Variables to find how many equations we want.

BVAR

To solve the problem of over parameterizatⁿ, BVAR was developed.

→ over parameterization led to df ↓ : which is dangerous

How to solve.

1st lag of dependent Variables has $\mu = 1$ root

$\sigma^2 = \lambda < 1$

As from the 2nd lag

Lag 2 ; $\sigma^2 = \lambda^{Lag}$

ie $\sigma^2 = \lambda^2$

Lag 3 ; $\sigma^2 = \lambda^{lag} = \lambda^3$

$\sigma^2 = \lambda^4$

$\sigma^2 = \lambda^5$

⋮

now as $\sigma^2 = \lambda^{lag}$; lag $\uparrow$ $V(y_t)\downarrow$

as $V(Y_t)\downarrow$ it converges to mean value $\mu = 0$

$\therefore$ we eliminate all those lags in the eq^n. for

(Suppose g variables where $\sigma^2 \longrightarrow 0$.

SVAR

VAR doesn't explain contemporaneous Link:
eq^n i.e $\{$ impact of y_t due to itself. $\}$.

SVAR models rely on economic theory to sort out the contemporaneous Link bet^n the variables in our model.

A K variable ; P Lags representation of VAR.

$$[y_t] = [\beta_{0i}]_{K\times1} + [\beta_1]_{K\times K}[y_{i,t-1}]_{K\times1} + [\alpha]_{K\times K}[y_{i,t-2}]_{K\times1}$$
$$\cdots [\delta]_{K\times K}[y_{i,t-p}]_{K\times1} + [\epsilon_t]_{K\times1}$$

In a 2 variable ; 1 lag model VAR representation:

$$y_t = [\beta_0]_{2\times1} + [\beta]_{2\times2}[y_{t-i}]_{2\times1} + [\epsilon_{it}]_{2\times1}$$

$$y_{1,t} = \beta_{01} + \beta_{11} y_{1,t-1} + \beta_{12} y_{2,t-1} + \epsilon_{1t}$$
$$y_{2,t} = \beta_{02} + \beta_{21} y_{2,t-1} + \beta_{22} \cdot y_{1,t-1} + \epsilon_{2t}$$

$$\Rightarrow \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix}\begin{bmatrix} y_{1,t-1} \\ \cdots \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \cdots \end{bmatrix}$$

Now to incorporate contemporary effect i.e., the effect of $y_{1,t}$ & $y_{2,t}$. We create structural transformation.

$\Rightarrow$ We transform

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} \longrightarrow \begin{bmatrix} a_{11}\, y_{1,t} & a_{12} \cdot y_{2,t} \\ a_{21}\, y_{2,t} & a_{22}\, y_{1,t} \end{bmatrix}$$

For simplicity let $y_{1,t}$ & $y_{2,t}$ be y_t & G_t resp. & so are their lags y_{t-1} & G_{t-1}

$$\Rightarrow \begin{bmatrix} y_t \\ G_t \end{bmatrix} = \begin{bmatrix} a_{11}\, y_t & a_{12}\, G_t \\ a_{21}\, y_t & a_{22}\, G_t \end{bmatrix}$$

} where y_t & G_t are income & expenditure of Govt.

- How do we do that transformat^n ?

$$\begin{bmatrix} a_{11}=1 & a_{12} \\ a_{21} & a_{22}=1 \end{bmatrix} \begin{bmatrix} y_t \\ G_t \end{bmatrix} = \begin{bmatrix} \beta_{10} \\ \beta_{20} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ G_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

$\downarrow$

Structural matrix.

<u>Compressed Matrix form</u>.

$$[A]_{2\times2}\,[y]_{2\times1} = [\beta_0]_{2\times1} + [\beta]_{2\times2}\,[y_{t-i}]_{2\times1} + [\epsilon_t]_{2\times1} \rightarrow \text{SVAR}$$

<u>Why add</u>

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_t \\ G_t \end{bmatrix} = ?$$

$$\Rightarrow a_{11}\, y_t + a_{12}\, G_t = \ldots$$

} we include y_t & G_t in the previous all lag eq^n.

This form of SVAR (RAW) can't run OLS as it has 2 dependent variables on LHS. { you can't regress 2 dependent variables in OLS at once }

To be able to run OLS, we transform SVAR to a form of VAR. we do that by multiplying $[A]^{-1}$ to both sides; estimating the LHS's matrix. $[A]$ is what is going to do the difference incorporating shocks.

$$\underbrace{[A^{-1}][A]}_{I}[y_t] = [A^{-1}][\beta_0] + [A^{-1}][B][y_{t-1}] + [A^{-1}][\epsilon_t]$$

$$\boxed{[y_t] = [A^{-1}][\beta_0] + [A^{-1}][\beta][y_{t-1}] + [A^{-1}][\epsilon_t]}$$

$$[A^{-1}] = \begin{bmatrix} 1 & a_{12} \\ a_{21} & a_1 \end{bmatrix}^{-1} = \frac{\text{Adj } A}{|A|}$$

$$\text{Adj } A = \begin{bmatrix} 1 & -a_{12} \\ -a_{21} & 1 \end{bmatrix} \quad \& \quad |A| = 1 - a_{12}a_{21}$$

If we compare the VAR & SVAR model eq^n

this changed from ϵ_t to U_t.

$$[y_t]_{2\times1} = [\beta_0]_{2\times1} + [\beta]_{2\times2}[y_{t-1}]_{2\times1} + [U_t]_{2\times1} \rightarrow \text{VAR}$$

$$[y_t] = [A]^{-1}[\beta_0] + [A]^{-1}[\beta][y_{t-1}] + [A^{-1}][\epsilon_t] \rightarrow \text{VAR Type form SVAR.}$$

Problem arises when we compare VAR & VAR type. The no. of parameters inside VAR

as $\epsilon_{t}^{*} = f(\epsilon_{t_1}, \epsilon_{t_2})$
$t = f(\epsilon_{t_1}, \epsilon_{t_2})$

$$(\beta_{01}, \beta_{02}, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \sigma^2_{\epsilon_{t1}}, \sigma^2_{\epsilon_{t2}}, \text{cov}(\epsilon_{t_1}, \epsilon_{t_2}))$$

Through SVAR, we transform VAR to incorporate the X_t & G_t shocks $\therefore$ cov $(\epsilon_{t_1}, \epsilon_{t_2})$ eliminates how it adds 2 additional parameters i.e a_{12} & a_{21}.

When we count the total no. of parameter of the SVAR transformed matrix i.e VAR type :

$$(\beta_{01}, \beta_{02}, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, a_{12}, a_{21}, \sigma^2_{\epsilon_{t^1}}, \sigma^2_{\epsilon_{t^2}})$$

Now to be able to make any meaningful comparison bet^n the 2 models we need to eliminate the parameter that has no impact. As 10 parameters ; 9 parameters ; } can be compared.

Thus for our easiness

$$\begin{bmatrix} 1 & a_{12} \\ a_{21} & 1 \end{bmatrix} \rightarrow \text{make it } 0 \text{ (restriction)}$$

$$= \begin{bmatrix} 1 & 0 \\ a_{21} & 1 \end{bmatrix} \rightarrow \text{Now we have 9 parameters}$$

Suppose we have more variables; how to determine the number of <u>restrictions</u> we have to make ?

$$\frac{k^2 - k}{2} \quad \text{where } k = \text{Variables.}$$

When $k = 2$

$$\therefore \frac{4-2}{2} = 1 \text{ restriction.}$$

When $k = 3$

$$\therefore \frac{9-3}{2} = 3 \text{ restriction.}$$

with restriction in place we compare every matrix of VAR to matrix of VAR type ; but our motive is to See the effect of

$$[U_t] = [A^{-1}] [\epsilon_t]$$

↳ forecasting Error
↳ estimate through OLS

↳ Shock (we can find it once A^{-1} is knows)

To see the effect of the shocks
we use impulse response function.

$$\boxed{y_t = \mu + \sum_{i=0}^{\infty} \beta_j \epsilon_t}$$