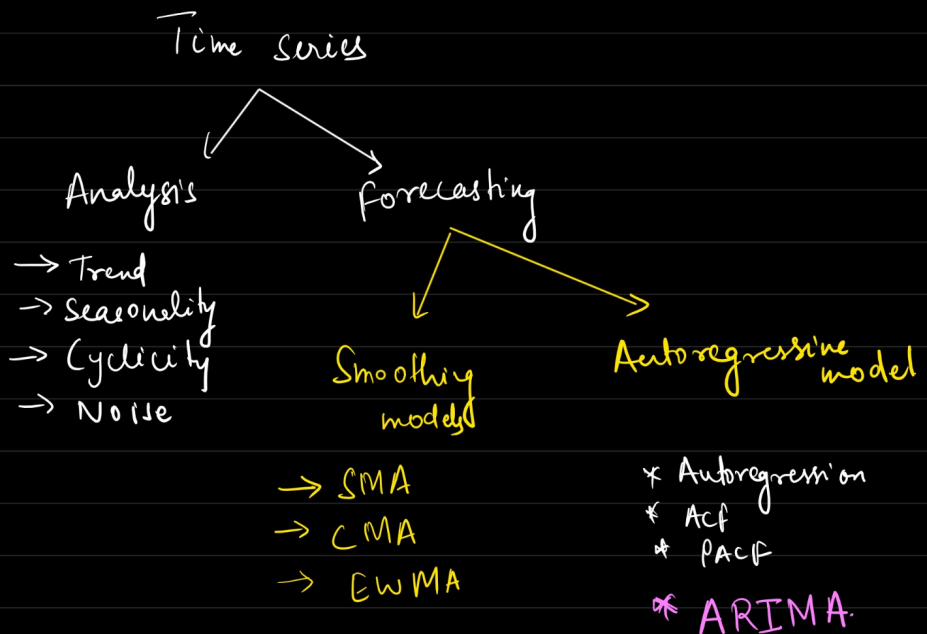
ARIMA



ML → Data → LR, SVR, DTR and so on.

TS data → Forecasting models. (Smoothing models already we have seen)

- → AR
- → MA
- → ARIMA
- → SARIMA
- → SARIMAX

AR

(Autoregression)

⇓

p

(0, 1, 2, ... n)

lag value

PACF

Partial autocorrelation function

(Correlelogram)

I

(Integrated)

⇓

q

(0, 1, 2, 3 .. n)

lag value

Differencing

(Stationary)

MA. model.

(Moving average)

⇓

q

(0, 1, 2, 3 .. n)

lag value.

ACF

(Autocorrelation function)

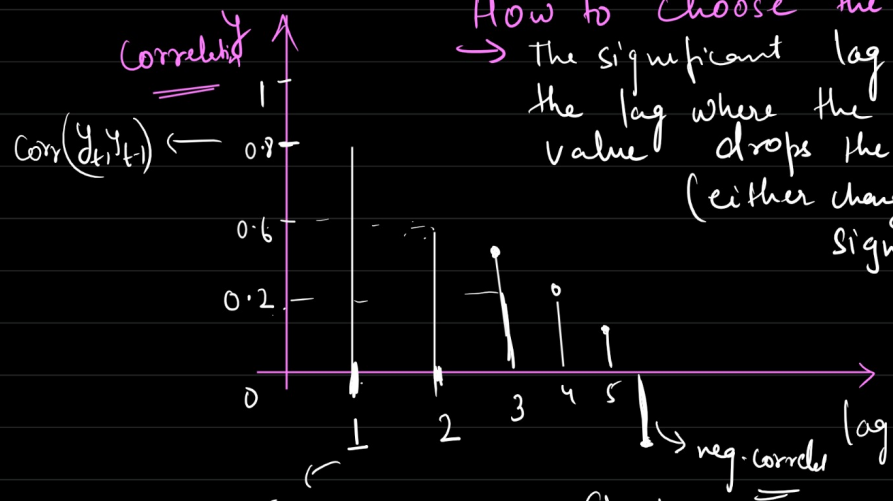
(Correlelogram)

ACF (for MA)

$Corr(y_t, y_{t-1})$ ✓

$Corr(y_t, y_{t-2})$

$Corr(y_t, y_{t-3})$

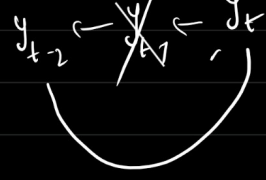


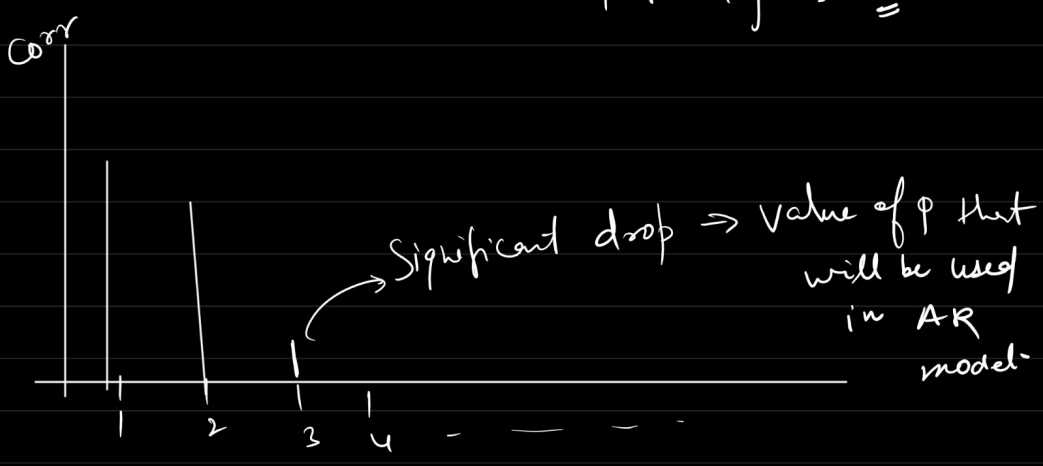
How to choose the lag?

→ The significant lag will be the lag where the correlation value drops the most. (either changes the sign, dropped significant)

Significant lag is - 5.

PACF → p → AR





* Autoregression

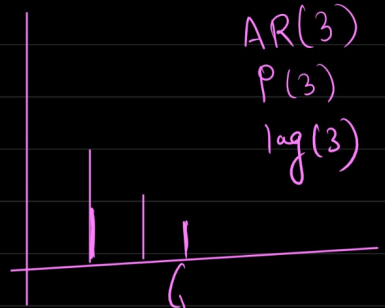
$$y_t = \psi_1 y_{t-1} + \psi_2 y_{t-2} + \cdots \psi_n y_{t-n} + C$$

Till how much lag we have to model??

PACF ⇒ Significant lag → P

Day	value	y_{t-1}	y_{t-2}	y_{t-3}
1	10			
2	20	↓		
3	30			
4				
5	40	20		
6	~~50~~			
7	60	20		

(y_t, y_{t-1}) , (y_t, y_{t-2}) (y_t, y_{t-3})



AR(3)
P(3)
lag(3)

$$y_t = \psi_1 y_{t-1} + \psi_2 y_{t-2} + \psi_3 y_{t-3} + C$$

for 2 lags ($P=2$)

$$y_t = \psi_1 t_{-1} + \psi_2 y_{t-2} + c$$

* We decide p (PACF) of AR, by seeing PACF plot. and q (ACF) of MA by seeing ACF plot.

ACF	lag 1	lag 2
10	Na	Na
20	10	Na
30	20	10
40	30	20
50	40	30

* Integrated - Differencing

D_0

D_1 D_2 D_3 D_4 D_5

d_1 d_1 d_1

D_2

y_t y_{t-1}

$(y_t - y_{t-1}) \rightarrow D = 1$

$(y_t - y_{t-2}) = D = 2.$

* Moving Error :-

(No same as Smoothy model average)

→ It models the error.

$$y_t = \varepsilon_{t-1} \psi + c$$

		Pred
D_1	10	8
D_2	15	?
D_3	20	
D_4	25	
D_5	30	

Last error $= 10 - 8 = 2$

$$y_t = \varepsilon_{t-1} \psi_{t-1} + \varepsilon_{t-2} \psi_{t-2} + \varepsilon_{t-3} \psi_{t-3} \cdots \varepsilon_{t-n} \psi_{t-n} + c$$

$\varepsilon \rightarrow$ error

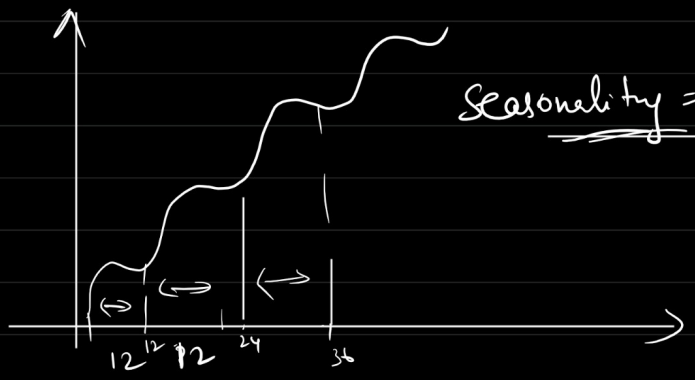
$$ARIMA = \underbrace{y_{t-1}\psi_{t-1} + y_{t-2}\psi_{t-2} + \cdots + y_{t-n}\psi_{t-n} + c}_{AR(p)} + \underbrace{(y_{t-1} - y_{t-2} - y_{t-3} \cdots - y_{t-n})}_{\text{Integrated } d} + \underbrace{\varepsilon_{t-1}\psi_{t-1} + \varepsilon_{t-2}\psi_{t-2} \cdots \varepsilon_{t-n}\psi_{t-n} c}_{MA(q) \rightarrow \text{Acf.}}$$

AR(1)
I(2)
MA(1)
} → ARIMA

$$y_t = (y_{t-1}\psi_{t-1} + c) + (y_t - y_{t-1} - y_{t-2}) + \varepsilon_{t-1}\psi_{t-1} + c$$

SARIMA.

Seasonality



$$(p,d,q)(P,D,Q)_s$$

$$\Rightarrow (1,1,2)(1,0,2)_{12}$$

→ Same p, d, q for seasonal component.

SARIMAX
↓
Seasonal

Exogenous variable (outlier, extra effect)
Noise

$$(p,d,q)(P,D,Q)_s$$
$$(0,1,2)(1,0,2)_{12} \cdot y_t \text{ (some extra variable)}$$