PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

Semester : VI   Subject : CSC601 Data Analytics and Visualization   Academic Year: 2023- 2024

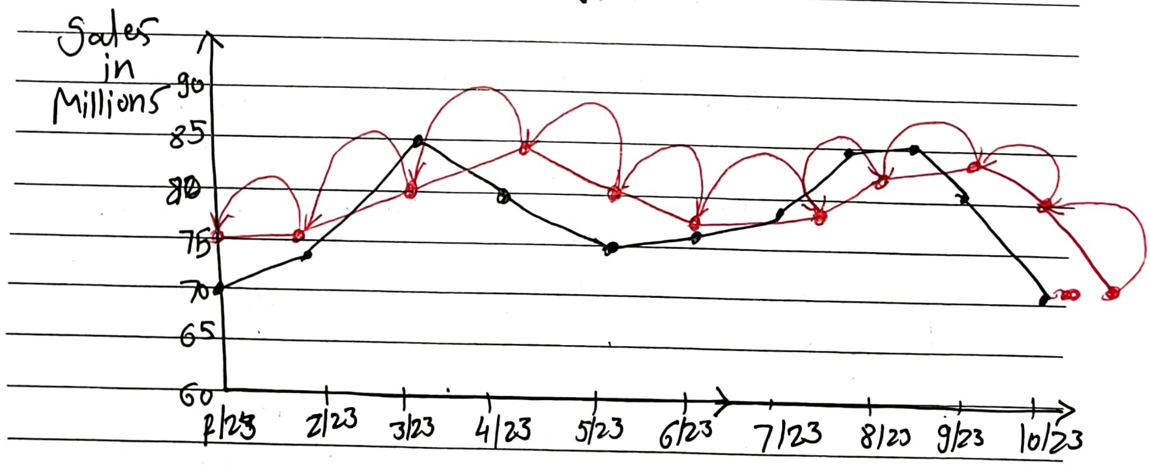
Moving Average Model (MA)

MA model forecast a series based solely on the past errors in the series, called error lags.

Model that depends only on one lag of error in the past:

$$Y_t = w + \theta e_{t-1} + e_t$$

Lagged error



Here — line → predicted values
— line → error

~~With~~ The idea of previous values of error is essentially how wrong you were yesterday. some unknown predicted shock that Shifted you to where you expected to be is what is actually affecting the current time. point here. ~~so the literal error from yesterday.~~

So the literal error from yesterday & the time period of before affects the current observation the current value of Y.

These errors do not last long into the future. Hence ~~so~~ it is called short memory model

$$Y_{t-1} = w + \theta e_{t-2} + e_{t-1}$$ (value for Yesterday)

$$Y_t = w + \theta e_{t-1} + e_t$$ (value for today)

These errors don't last long into the future.

$$Y_{t+1} = w + \theta e_t + e_{t+1}$$ (Value for tomorrow)

Here in the equation the error e_{t-1} is ~~gone~~ not considered.

PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

CSE DATA SCIENCE

Semester : VI   Subject : CSC601 Data Analytics and Visualization   Academic Year: 2023- 2024

In MA models, the effect of shocks have No effects on the present if they happened long enough ago.

In AR model ~~wt~~ we recursively look back in time, here the first observation matters, even if it is a small amount it still matters.

In MA model on the other hand as long as you far enough into the future observations and errors will literally have no effect.

A time series that is a linear function of q past error values plus current error is called a moving average process of order q - MA(q)

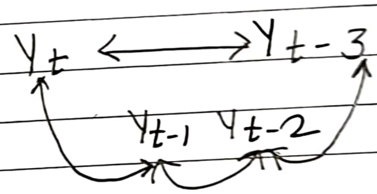
$$\boxed{Y_t = w + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q} + e_t}$$

Semester : VI Subject :CSC601 Data Analytics and Visualization Academic Year: 2023- 2024

ACF and PACF

Autocorrelation Function Plot:
ACF is all inclusive in measuring autocorrelation as the lag varies.

If we are trying to find the correlation between Y_t and Y_{t-3}, This correlation includes Y_{t-1} & Y_{t-2}. as there is dependency of these two variables.



day	~~Predicted~~ Value	lag 1	lag 2
d1	1	–	–
d2	2	1	–
d3	3	2	1
d4	4	3	2
d5	5	4	3
d6	?	5	4

$y \sim f(\text{lag1}\,y,\ \text{lag2}\,y, \ldots\ldots)$

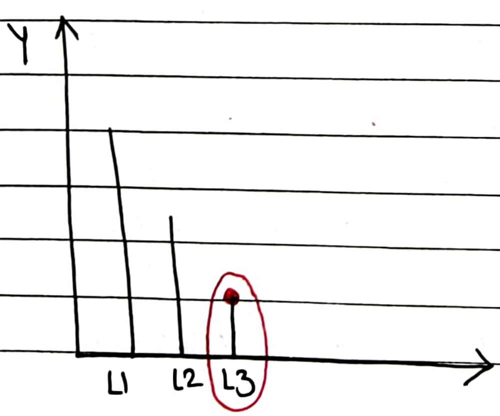
This is the concept behind auto regressive model.

The lag values crossing the significance line is said to have statistically significant relationship with Y.

Upto lag L2 the lag values are statistically significant relationship with Y as per the given ACF plot.

PACF (Partial Auto correlation Function)
The strength of correlation between the lags of Y ~~and~~ ~~with~~ and Y itself. But PACF additionally describes strength of any lag with Y after removing the effects of lags in between.

PACF plot generally falls faster compared to ACF plot



Strength of correlation between L3 & Y after removing the effect of L1 & L2.
So PACF describes contribution of L3 only.

~~In PACF~~

If we represent Y using linear regression the equation we get is

$$y = \ln(\underbrace{y_1 + y_2 + y_3 + y_4}_{\text{lags}})$$

To find the PACF value of lag 4, we will fit this model in linear regression model.

$$Y = \theta_0 + \theta_1 Y_1 + \theta_2 Y_2 + \theta_3 Y_3 + \boxed{\theta_4} Y_4$$

PACF of Y_4

The coefficient θ_4 is the PACF value of Y_4

Now what will be the value of PACF for L2.
Here we can not take same value of θ_2 from above equation. We need to write the new equation with only Y_1 & Y_2.

$$y = \ln(y_1 + y_2)$$
$$y = \theta_0 + \theta_1 Y_1 + \boxed{\theta_2} Y_2$$

PACF of Y2

The θ_2 in both equation have different values.

PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

ARMA Model

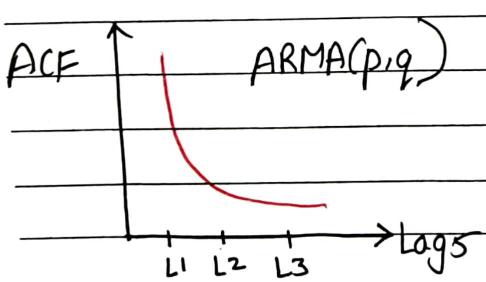
(Auto Regressive Moving Average Model)

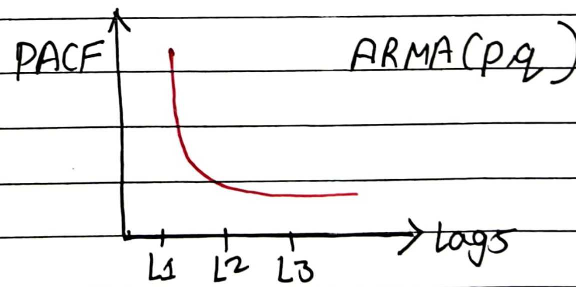
This model is combination AR series and MA series. This model has benefits of long memory model and short memory model.

ARMA (p,q) Series

$$Y_t = \mu + \underbrace{\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi Y_{t-p}}_{\text{AR Terms}} + \underbrace{\theta_1 U_{t-1} + \dots \theta_p U_{t-q}}_{\text{MA Terms}} + U_t$$

The ARMA model is used for the time series where exhibits properties of auto regression and moving average. To detect of the given time series is an ARMA model we need to draw ACF & PACF curve.



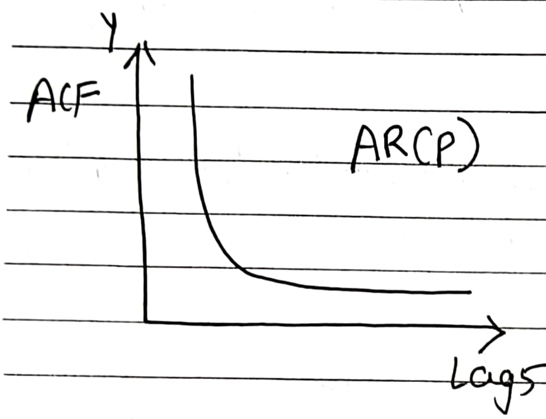


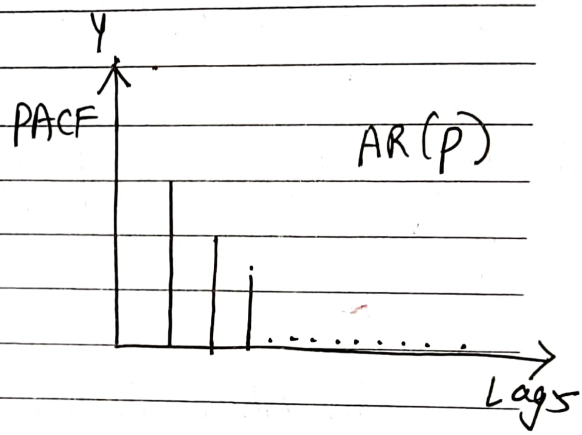
If the ACF & PACF curves are declining with

p - number of lags for AR terms
q - number of lags/error values for MA terms.

the increasing values of lags. In such case ARMA needs to applied.

The ACF of AR model is also declining but the values of PACF plot are different for AR model & ARMA model.





Semester : VI Subject :CSC601 Data Analytics and Visualization Academic Year: 2023- 2024

The estimation can be calculated using ordinary least square method.

Step3

- Check the accuracy of the model.
- In this step we try to check if the model fit is good.
- The conditions to be checked are overfitting and residual diagnostics

- Overfitting is the condition when the model ~~training~~ works accurately for training dataset but for testing dataset the performance is not upto the mark.

- In underfitting condition we can not identify the pattern in the given dataset. This results into autocorrelation of error terms.

Box - Jenkins Methodology

The two scientists Box & Jenkins have defined the ~~mod~~ methodology to design ARMA model.

This method consists of three steps
1. Identification
2. Estimation
3. Diagnostic Checking

Step 1 Identification

- Determine the order of the model
  We identify the orders of AR series & MA series.

- Calculate ACF & PACF for p and q

Step 2 Estimation

Now as number of lags to be used for AR series & MA series are known

$$\boxed{Y_t = \phi_1 Y_{t-1} + \Theta_1 u_{t-1} + u_t}$$ ARMA(1, 1) Model.