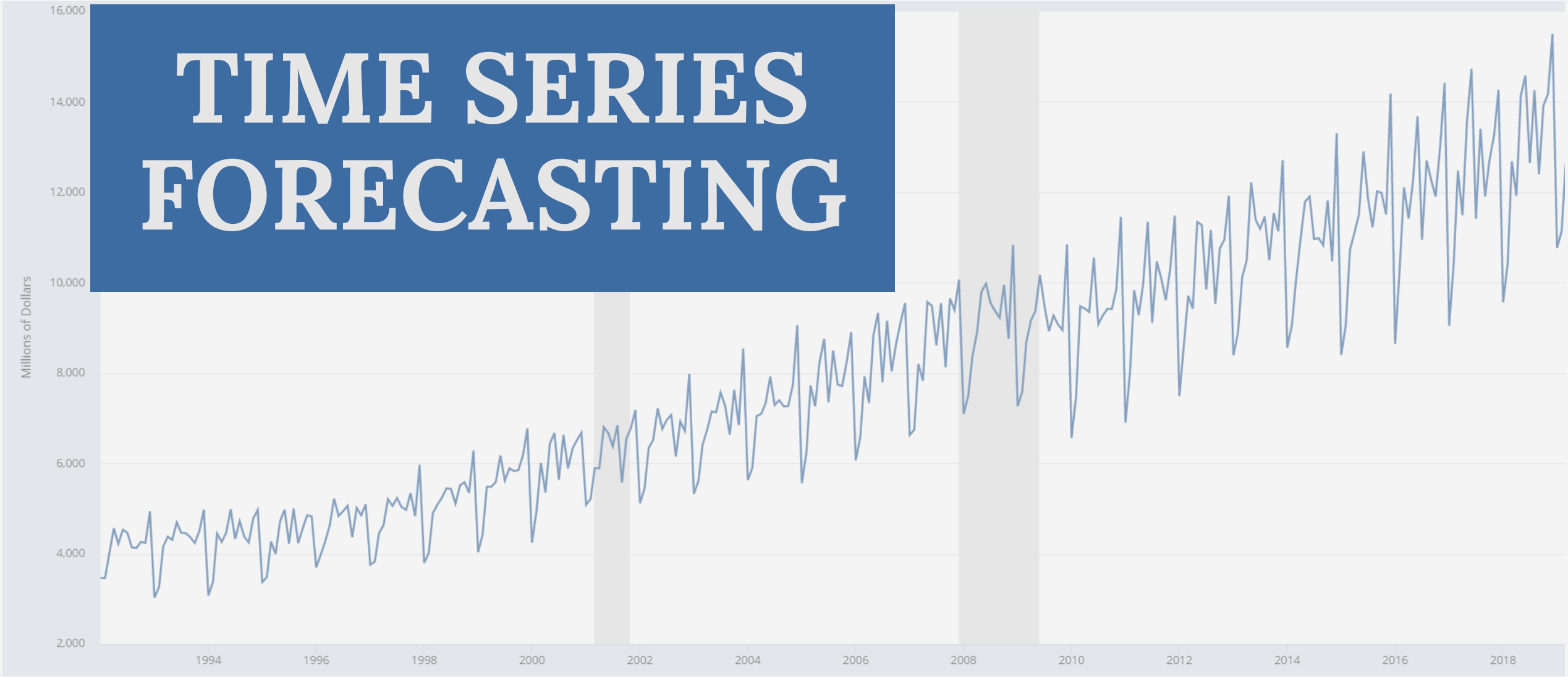
TIME SERIES FORECASTING



Contents

- What Is Time Series?
- Why Time Series Forecasting?
- Components of Time Series
  - Trend
  - Seasonality
  - Cyclical
  - Irregular
- Stationarity
- ARIMA model
- Seasonal ARIMA (SARIMA)

Time Series

- Time Series is data collected at ***constant time intervals***.
- Constant interval means – either the data should on daily, monthly, quarterly etc. levels.
- Time series data is analyzed to forecast the future values based on the observed values.
- Example – Stock prices, Daily sales at a coffee shop etc..

Time Series Data

Date	Open
May 22, 2019	6,877.20
May 21, 2019	7,105.00
May 20, 2019	6,875.00
May 17, 2019	6,497.00
May 16, 2019	6,506.00
May 15, 2019	6,590.00
May 14, 2019	6,511.00
May 13, 2019	6,626.00
May 10, 2019	6,623.00
May 09, 2019	6,637.00
May 08, 2019	6,679.00
May 07, 2019	6,744.90

Date	Open
May 22, 2019	6,877.20
May 20, 2019	6,875.00
May 13, 2019	6,626.00
May 06, 2019	6,659.00
Apr 29, 2019	6,842.85
Apr 22, 2019	7,422.00
Apr 15, 2019	7,351.05
Apr 08, 2019	7,107.00
Apr 01, 2019	6,730.80
Mar 25, 2019	6,515.00
Mar 18, 2019	7,077.45
Mar 11, 2019	6,980.00
Mar 04, 2019	6,935.15

Date	Open
May 22, 2019	6,877.20
May 01, 2019	6,666.40
Apr 01, 2019	6,730.80
Mar 01, 2019	6,876.80
Feb 01, 2019	6,679.00
Jan 01, 2019	7,449.75
Dec 01, 2018	7,688.50
Nov 01, 2018	6,648.95
Oct 01, 2018	7,354.00
Sep 01, 2018	9,100.00

Why Time Series Forecasting?

- Only one variable in Time Series forecasting – **Coninuous variable vs Time**.

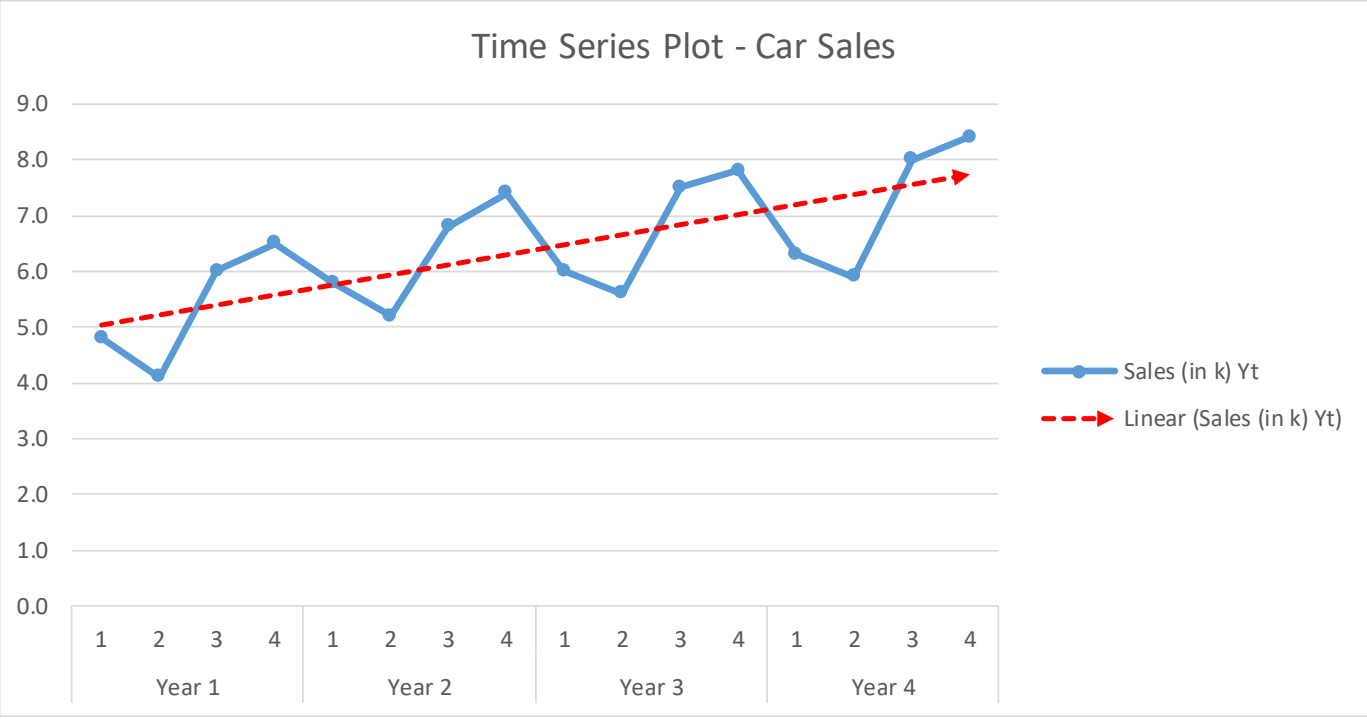
- How different from Linear regression –

  - It is time dependent. Basic assumption of LR - observations are independent is ignored.

  - Presence of Seasonality – fluctuations is specific time period.

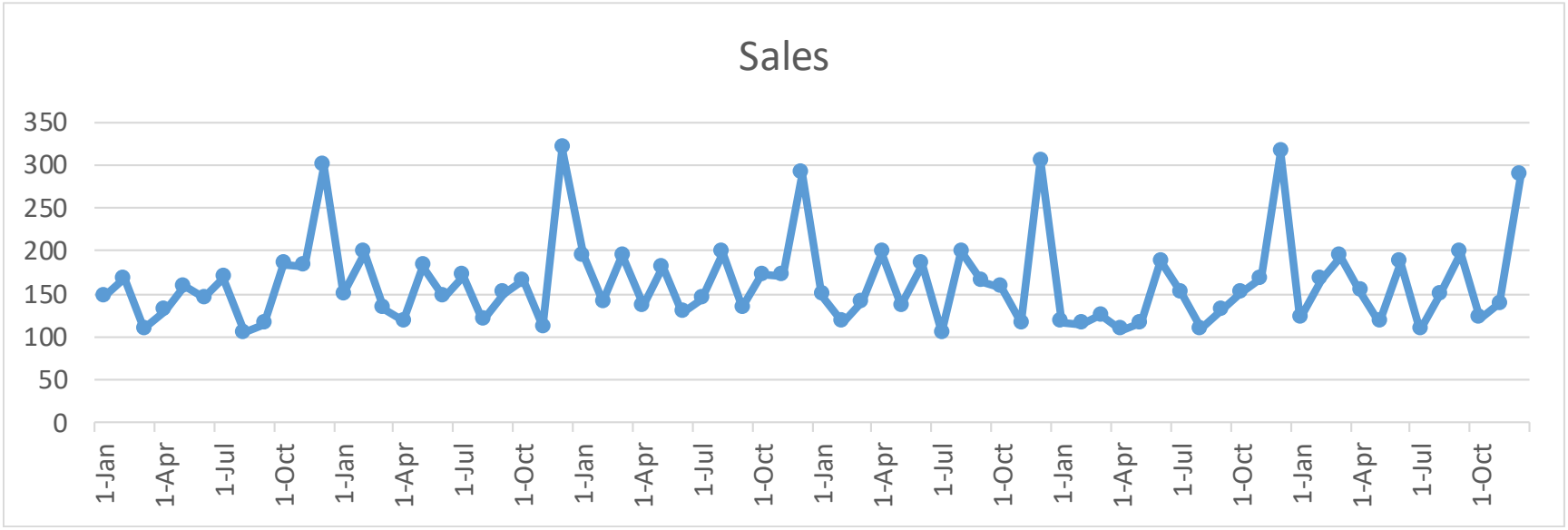
Components of Time Series

1. **Trend component** – Trend is a general direction of data series.



Components of Time Series

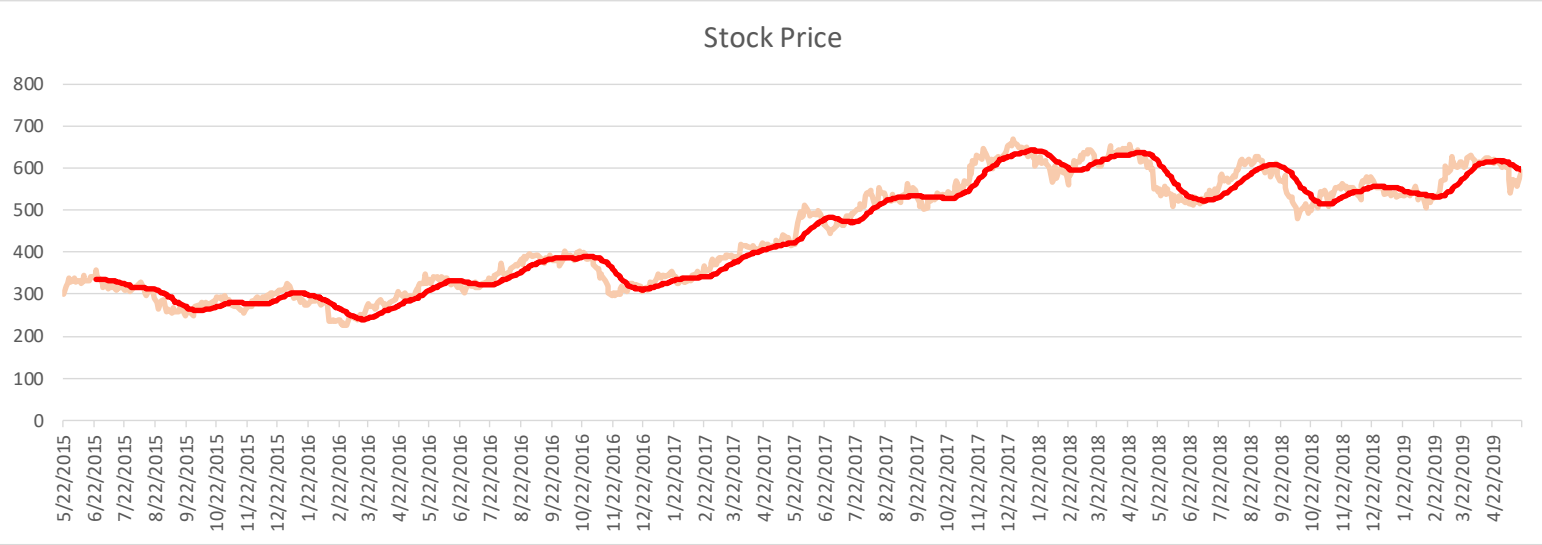
2. **Seasonality component** – Pattern that repeats at regular time interval which is known as the seasonality.



- Exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week).

Components of Time Series

3. **Cyclical component** – A cyclic pattern exists when data exhibit rises and falls that are not of fixed period.



Components of Time Series

4. **Irregular component** – The irregular component of a time series is the residual time series after the trend-cycle and the seasonal components (including calendar effects) have been removed.

- It corresponds to the high frequency fluctuations of the series.

Stationarity

- Time Series requires the data need to be Stationary.
- Most of the Time Series models work assuming the data is Stationary.
- Stationarity means –
  - Conatant mean according to the time.
  - Constant variance (at different time intervals),
  - Covariance doesn’t depend on time. Should be constant over time.

Tests to check Stationarity

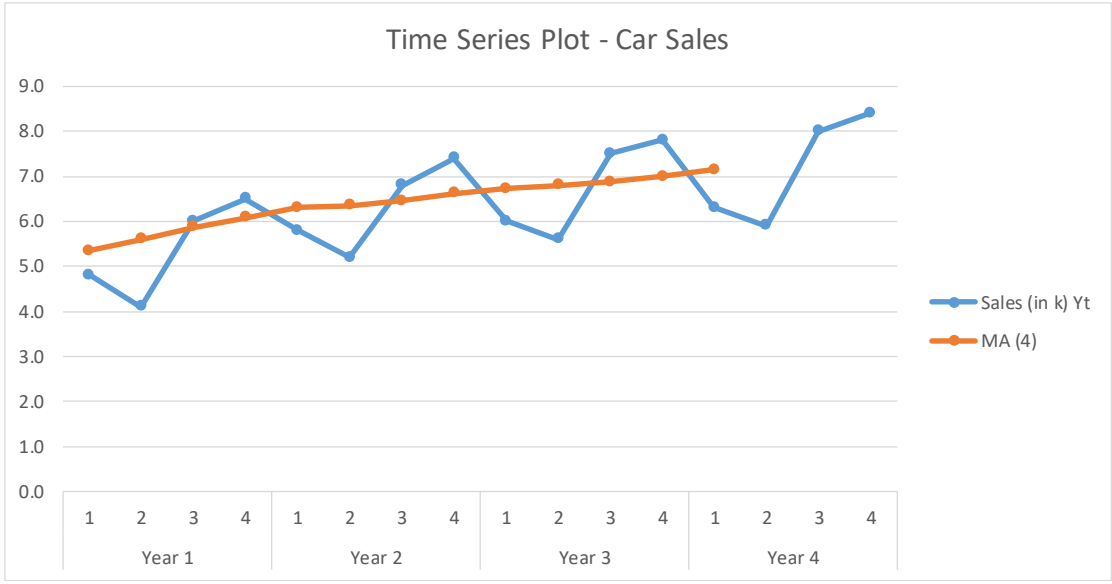
- Rolling test
  - It is a visual technique.
  - Plot the moving average or variance and see if it varies with time.
- Dickey Fuller Test
  - Null Hypothesis is Time Series is non-stationary.
  - Test results include test statistic and critical values.

Auto Regressive (AR) model

- An autoregressive (AR) model predicts future behavior based on past behavior.
- Used when there is some correlation between values .
- Basically a linear regression between current data vs previous.
- **AR(p) Model:** AR(p) model is an autoregressive model where specific lagged values of y are used as predictor variables.
- Lags are where results from one time period affect following periods.
- The value for “p” is called the order.

Moving Average (MA) model

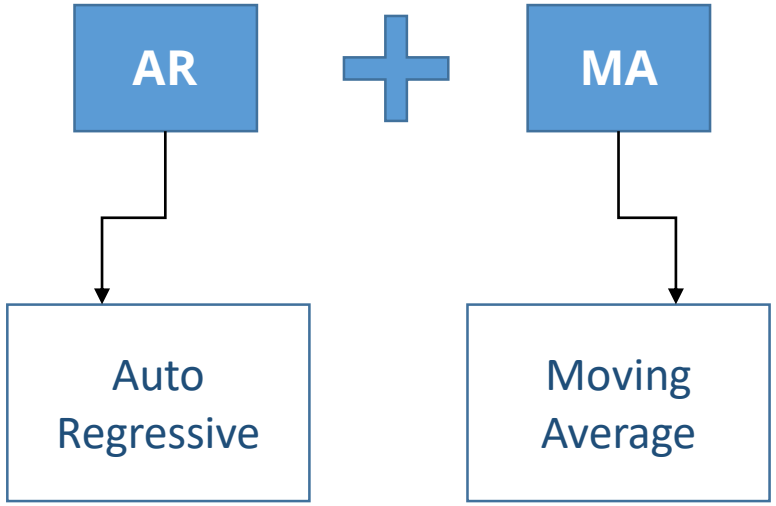
- A moving average is a technique to get an overall idea of the trends in a data set.
- It is an average of subset of data points at specific periods .
- The moving average is extremely useful for forecasting long-term trends.

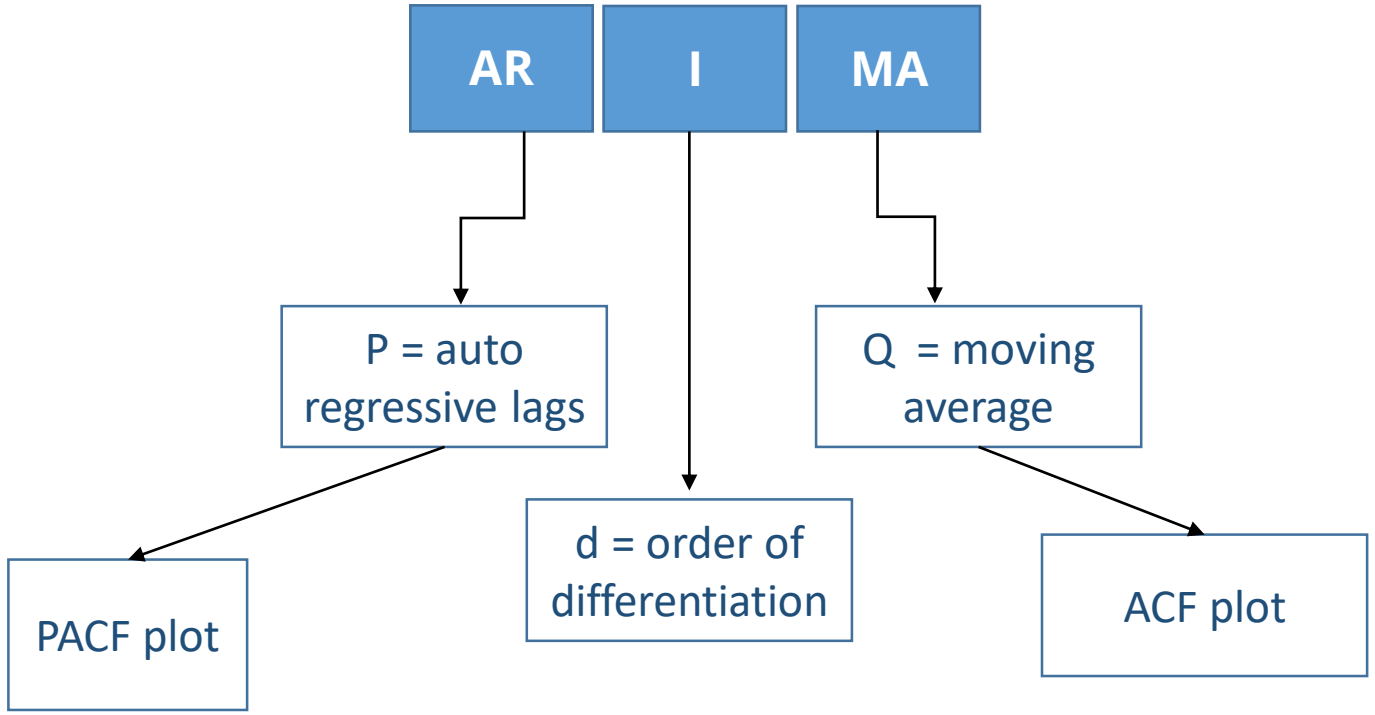


ARIMA

- ARIMA stands for **Autoregressive Integrated Moving average**.
- It is specified by three ordered parameters (p,d,q).
  - **p** is the order of the autoregressive model (number of time lags)
  - **d** is the degree of differencing(number of times the data have had past values subtracted)
  - **q** is the order of moving average model.
- Applied when data show evidence of non-stationarity, where a differencing step can be applied to eliminate the non-stationarity.

ARIMA


AR
MA
Auto Regressive
Moving Average


AR
I
MA
P = auto regressive lags
Q = moving average
PACF plot
d = order of differentiation
ACF plot

SARIMA

- This is an extension to ARIMA model, applied to ARIMA time series that shows seasonal patterns.
- SARIMA models are ARIMA models with a seasonal component. Per the formula SARIMA(p,d,q)x(P,D,Q,s), the parameters for these types of models are as follows:
  - p and seasonal P: indicate number of autoregressive terms (lags of the stationarized series)
  - d and seasonal D: indicate differencing that must be done to stationarize series
  - q and seasonal Q: indicate number of moving average terms (lags of the forecast errors)
  - s: indicates seasonal length in the data