TIME SERIES FORECASTING

WHAT WE'RE GOING TO COVER

- Time Series Forecasting
- Simple Moving average
- Exponential Smoothening
- ARIMA

EXAMPLE OF PRODUCT DEMAND DATA

Year	Demand
2004	25
2005	32
2006	24
2007	28
2008	26
2009	27
2010	??

Objective is to find a value which adequately represents the values given

Set of students were asked to forecast the demand for Year 2010. Answers are shown below:-

F1 = 27	Based on average of all the values
F2 = 26.25	Based on the average of last 4 values as earlier 2 old
F3 = 28	Consider latest values which show increasing trend (hence 28)
F4 = 30	Buying power going up so demand (based on other factors)
F5 = 26	Pattern Based (increase and decrease)
F6 = 27	Based on average of last 3
F7 = 26.833	Based on weighted average for last 3 values (lowest wt. to oldest i.e. avg = 1*28+2*26+3*27)/3)
F8 = 26	Removed extreme value i.e. 32, as an outlier and took average of the rest

WHAT IS TIME SERIES ANALYSIS

- A time series is a sequence of data points taken at successively equally spaced points in time.
- Examples of time series data: Daily/weekly/quarterly sales data, rainfall data, websites hits per minute.
- It is one of the most applied data science techniques in business and is used extensively in finance, supply chain management, production and inventory planning.
- Time series is univariate analysis where the dataset has only two dimensions: the variable itself and a time index.
- Its is different from regression analysis because the dependent variable is a function of past values of the dependent variable itself.

TIME SERIES V/S REGRESSION ANALYSIS

- **Time series analysis differ from Regression in that, dependent variable is a function of past values of variable itself ….**

Regression Analysis			
Engine Capacity (X1)	Features (X2)	Model (X3)	Car Price (Y)
1.4 CC	Air Bags	ESL	5 L
1.6 CC	No Air bags	ZSL	4 L
…	..	..	N

Time Series Analysis			
	t-1	t	t+1
Price	10000	1000	15000
Quantity	20000	2000	22000
Sales	1000	900	1100

4 COMPONENTS OF TIME SERIES ANALYSIS

1. **Seasonal Variations:** that repeat over a specific period (Short period) such as a day, week, month, season, etc.

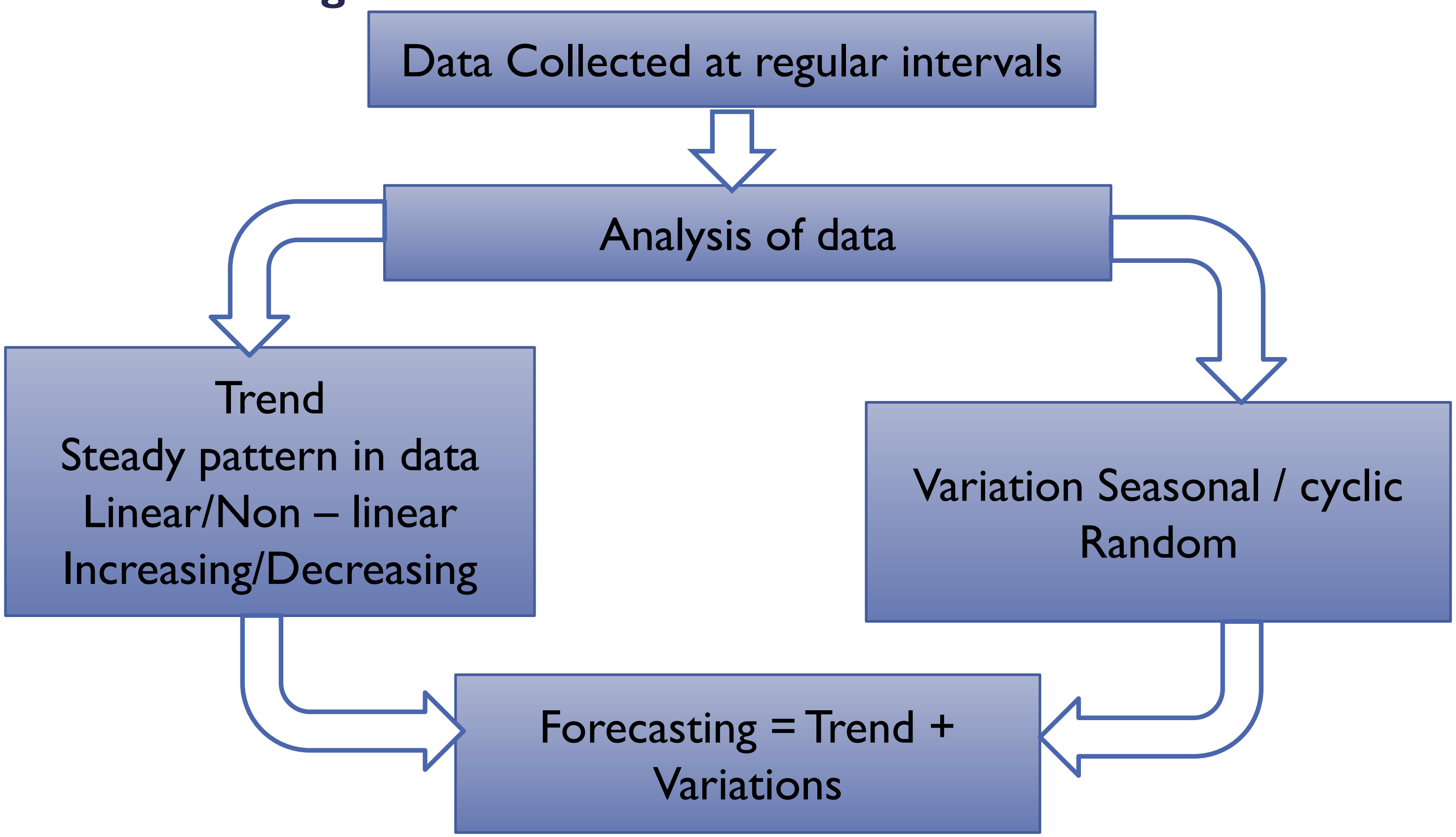
2. **Trend:** Trend is defined as long term (steady) increase or decrease in the data. It can linear or non-linear

3. **Cycle Variations:** A cyclic pattern exists when data exhibits rises & falls, that are not of fixed period (usually long period).

4. **Random Variations:** that do not fall under any of the above three classifications. Unobservable time dependent influences.

FORECASTING

- **Analysis of time series data with respect to trends and variations is called forecasting.**

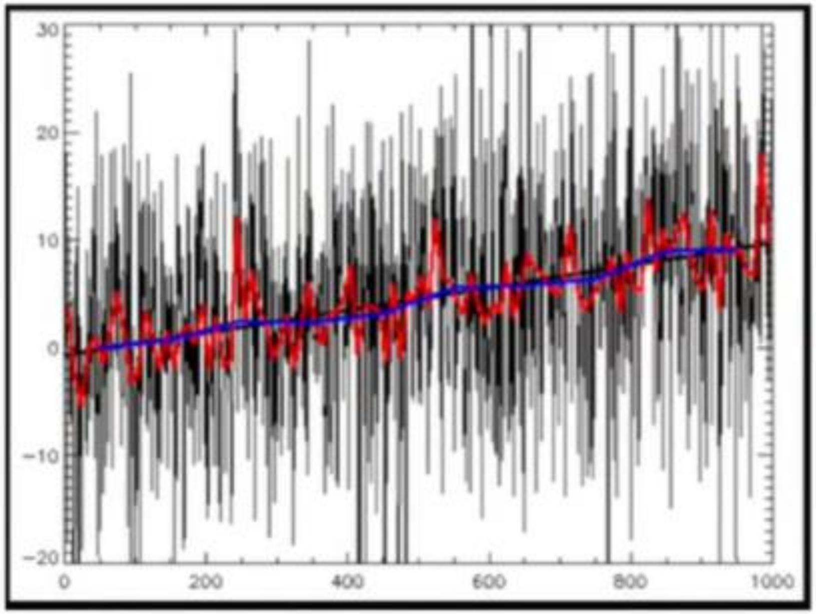


COMPLEX NATURE OF DATA IN TIME SERIES

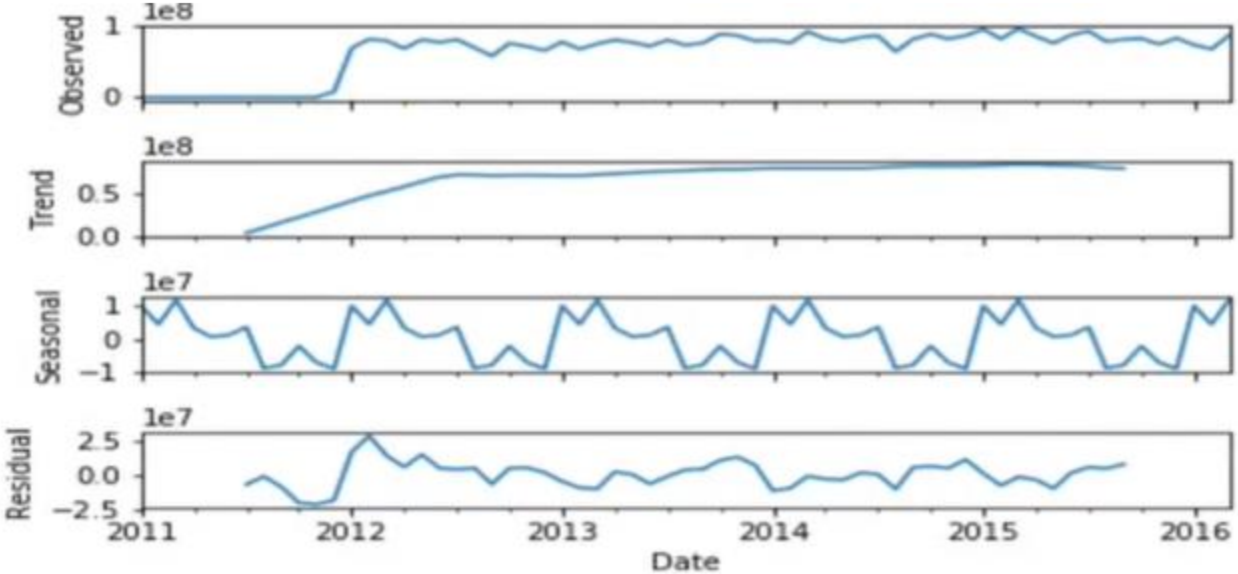
- The complex nature of data in a time series, such as seasonality, trend, and level, may bring numerous challenges to produce accurate forecasts.
- For forecasting purposes, it is important to identify and work with parts of the series which are more systematically driven and hence can be forecasted.
- This is an example of noisy series
- Black (Daily) and Red (weekly):

  seasonal Variations

  Blue: Trend

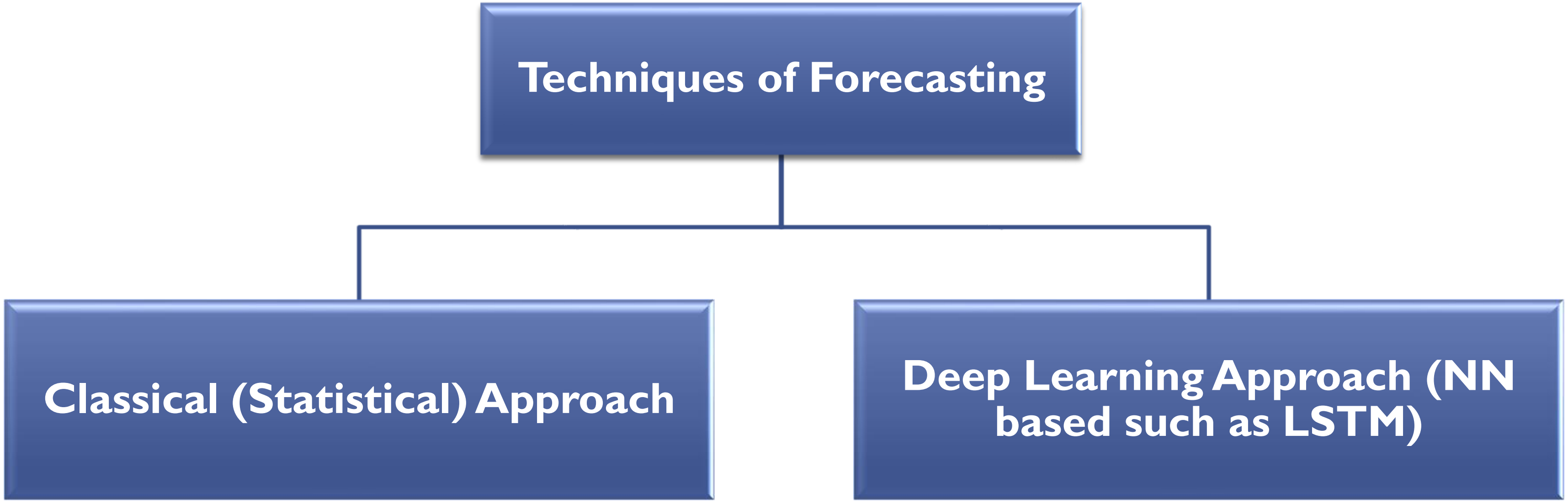


DECOMPOSED TIME SERIES



- <u>Stationary Time Series</u> – Mean and Variance are constant in time.
- <u>Trend</u> – Visible
- <u>Seasonal</u> – Seasonal variations are roughly constant across all years
- <u>Residual</u> – Changes with time – Indicates time series

TECHNIQUES OF FORECASTING



TECHNIQUES OF FORECASTING - CLASSICAL APPROACH

Classical (Statistical) Approach

Average
- Simple Average
- Weighted Average
- Simple Moving Average (SMA)

Exponential Smoothing
- Simple Exponential Smoothing (SES = SMA + Weighted Average)
- Holt's linear exponential smoothing (Double)
- Holt's winter exponential smoothing (Triple)

It captures average and trend both

Auto Regressed Integrated moving Average models
- AR Models
- ARMA Models
- ARIMA Models

It captures average, trend & seasonality

Regression on its own lagged values

Regression on its own lagged values + Moving Average

Regression on its own lagged values + I[The I (for "integrated") indicates that the data values have been replaced with the difference between their values and the previous values] + Moving Avg.

SIMPLE AVERAGE AND WEIGHTED AVERAGE

- **Weighted average** (less weight to older data & higher weightage to latest data). These weights are assigned by domain experts.
- **Simple average** is the simplest way of determining the constant under the conditions that:
  - All data points are considered
  - Equal weight to all data points

MOVING AVERAGE

Each moving average is for a consecutive block of 5 years

Year	Sales
1	23
2	40
3	25
4	27
5	32
6	48
7	33
8	37
9	37
10	50

etc…

Average Year	5-Year Moving Average
3	29.4
4	34.4
5	33.0
6	35.4
7	37.4
8	41.0
9	39.4
…	…

$$3 = \frac{1+2+3+4+5}{5}$$

$$29.4 = \frac{23+40+25+27+32}{5}$$

SIMPLE MOVING AVERAGE (K PERIOD MA) …

- Moving average based on window of K = 2,3,4,………
- Drawback:
  - We are not considering few values
  - The larger the k, more the values omitted.

ADVANTAGES OF SMA

- The SMA is a simple method and easy to understand.
- It gives a good visual of the trend and smoothens out short term fluctuations.
- It also reduces out the effects of extreme values.
- On the con side the method does not have a statistical methodology to determine the forecasting period.

EXPONENTIAL SMOOTHING

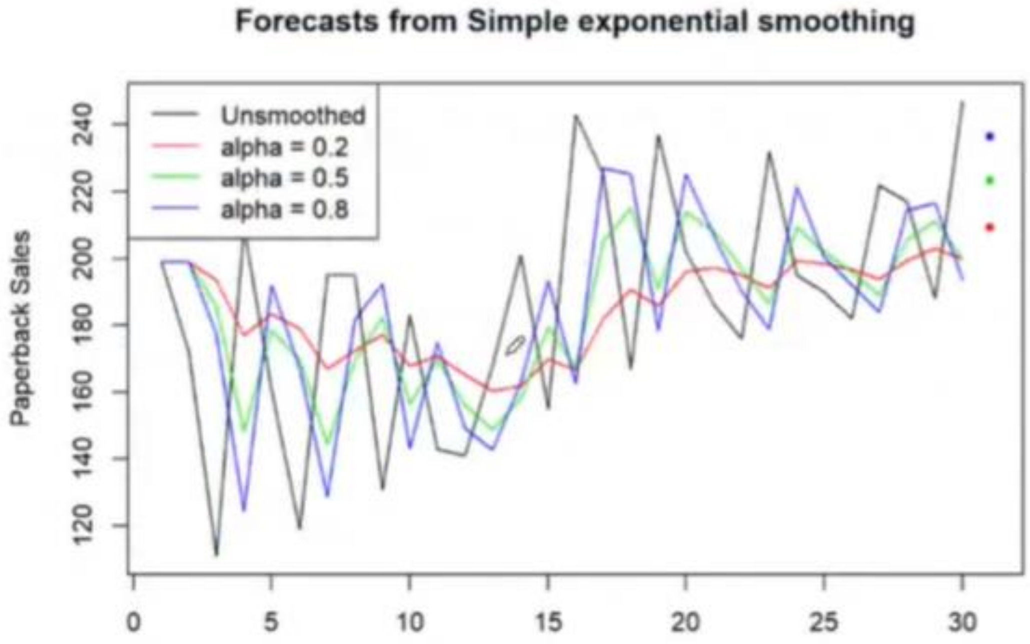
- Simple Exponential Smoothing model = Progressive weights + moving average
- Its model where:
  - We consider all the data
  - Give progressively increasing weights (less to older)
  - Moving average is taken
  - Weights are taken logically and acceptable to everyone.
- Data → 25,32,24,28,26,27

SIMPLE EXPONENTIAL SMOOTHING

- Data → 25,32,24,28,26,27
- Model
- F(t+1) = alpha*D(t) +(1 – alpha)* F(t) (F = Forecast, D = Demand, alpha = smoothing constant (value between 0 & 1, 0<alpha<1)
- If alpha towards 1, contribution of D(t) more otherwise F(t) more
- Alpha is taken smaller normally
- F(7) = alpha*(D6)+1-alpha*F(6)
- Assuming F1 = 27 (say simple average) and alpha = 0.2 → F2,F3,F4,F5,F6,F7 can be calculated

SIMPLE EXPONENTIAL SMOOTHING CTD …

- In the figure below where we see that alpha = 0.2 is (red line) is much smoother than the alpha = 0.8 (blue line).



- Values of alpha close to one have less of a smoothing effect and give greater weight to recent changes in the data, while the value of alpha closer to zero have a greater smoothing effect and are less responsive to recent changes.

VARIANTS AND EXTENSIONS OF SES MODEL

- The key limitation of SES forecasting is that the SES models assume that there is no trend of any kind in the data.
- This acts as a limitation for anything except very short-term forecasting, particularly if there is a short term or long-term trend in the data.
- Various extensions of the SES models have been developed to get around this problem. The main ones are:

Variants of SES Model

- Holt's Linear Exponential Smoothing Model (Double exponential smoothing) SES + trend
- Holt's Winter Model (Triple exponential smoothing) SES + trend + seasonality

HOLT'S LINEAR EXPONENTIAL SMOOTHING (DOUBLE)

- Holt's Linear Exponential smoothing (LES) also called Trend Models
- 26
- 28
- 29
- 31
- 32
- 35
- ?

Forecast based on trend for this data → Value more than 35

- SES + TREND
- Uses two coefficients Alpha & Beta

HOLT'S WINTER MODEL /TRIPLE EXPONENTIAL SMOOTHENING

- Holt (1957) and Winters (1960) extended Holt's method to capture seasonality.
- SES + Trend + Seasonality
- This is also called the Multiplicative Holt – Winters, is usually more reliable for data that shows trends and seasonality.
- The Holt – Winters seasonal method comprises the forecast equation and 3 smoothing equations – one for the level, Lt, one for the trend Tt, and one for the seasonal component, St, with corresponding smoothing parameters α, β and gamma.
- We use m to denote the frequency of seasonality, i.e., the number of seasons in a year. For e.g., for quarterly data m = 4, and for monthly data m = 12.
- There are 2 variations to this method that differ in the nature of the seasonal component.
- The **additive** method is preferred when the seasonal variations are roughly constant through the series, while the **multiplicative** method is preferred when the seasonal variations are changing proportional to the level of series.

WHEN TO USE DIFFERENT TYPES OF EXPONENTIAL SMOOTHING & WHY

SES

- SES is usually used to make short term forecasts. It is more effective than SMA because it gives a higher weight to more recent data points v/s equal weightage given by SMA.
- SES cannot do longer term forecasts reliable primarily because this method does not consider any trend in the data.
- If we believe that the trend in the past are important for determining current trends, we can use the LES models. If we are not sure of whether there is a trend or not we can use SES model

Holt's Double Smoothening

- If we know that there is a trend in the data, then this method can be used.
- Extrapolating trends over very long periods may not make sense given that trends change as product life cycle changes, increased competition in a market or a cyclical downturn.
- Hence, we find frequently SES performs better often despite its naïve assumption of a horizontal trend

Holt's Winter Model

- Use this when there is seasonality in the data.

WHAT IS ARIMA

- **Auto Regressive Integrated Moving Average (ARIMA)** model is among one of the more popular and widely used statistical methods for time – series forecasting. It is a class of statistical algorithms that capture the standard temporal dependencies that is unique to a time series data.
- **<u>Let us define the 3 components of AR Models:</u>**

AR (Autoregressive)

- The AR part of ARIMA indicates that the evolving variable of interest is regressed in its own lagged (i.e. prior) values p is the order (number of time lags) of the autoregressive model

I (Integrated Differencing

- The 'I' (for "Integrated") indicates that the data have been replaced with the difference between their values & the previous values (this differencing process may be performed more than once.
- The step is carried out to make the series stationary d is the degree of the differencing (the number of times the data have had past values subtracted)

Moving Average

- It's the moving average for last defined interval

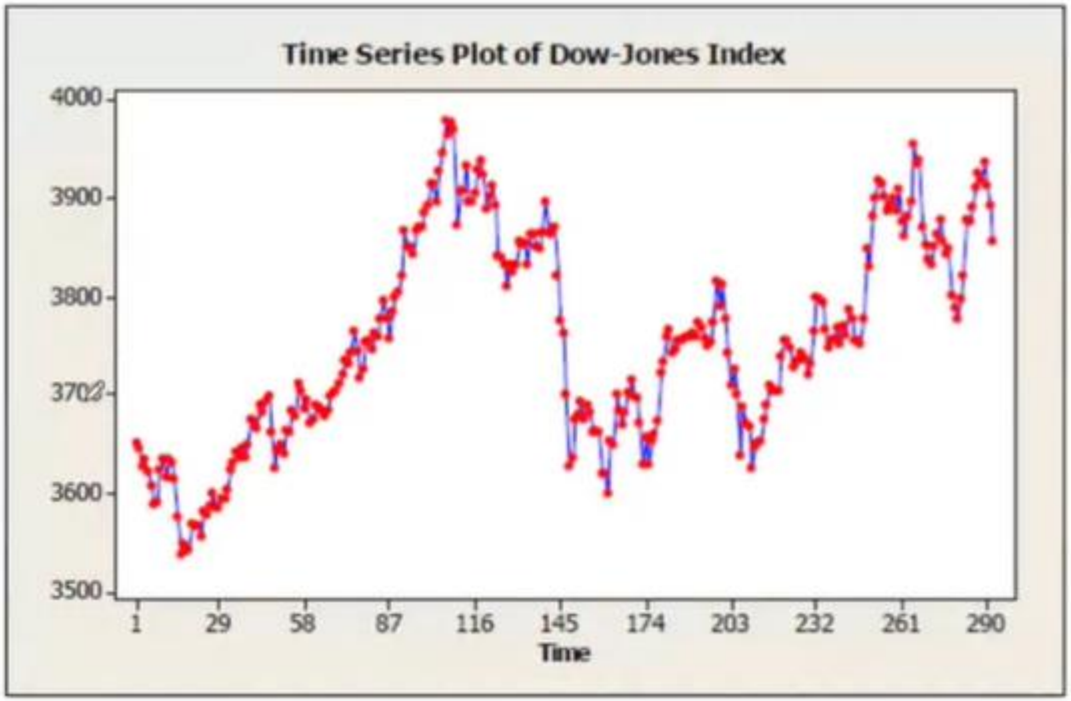
CORRELATION AND AUTOCORRELATION

- Another important concept to understanding ARIMA is autocorrelation.
- How is it different from the typical correlation?
- First, correlation relates two different sets of observations (e.g. between the housing prices and the number of available public amenities) while autocorrelation relates the same set of observations but across different timings (e.g. between rainfall in the summer versus that in the winters).

STATIONARITY

- ARIMA model assumes that time series which we are working on is stationary.
- For forecasting we need to find the constant (time variant) component in the series, otherwise it's impossible to forecast. Stationarity is essentially invariance of the series over time.
- This would require the parameters of the distribution i.e. mean and variance to be constant.
- For e.g. if the series is consistently increasing over time, the sample mean and variance will grow with the size of the sample & they will always underestimate the mean and variance in future periods, hence it would be difficult to forecast.
- For a time, series to be stationary it should satisfy the following conditions:
  - Mean (μ) is constant for all t
  - SD () is constant for all t
  - No seasonality (Seasonality is absent) for all t
- Technique of differencing is used to make the non – stationary series as stationary.

EXAMPLE OF NON-STATIONARY SERIES



Mean = Not Constant
SD = Not Constant
Seasonality = Yes
Not Stationary
Check for Stationarity
→ Augmented Dicky Fuller Test (ADF)

THE AR IN [AR]IMA: AUTO REGRESSIVE

- The Auto Regressive (AR) regression model is built on top of the autocorrelation concept, where the dependent variable depends on the past values of itself (e.g. rainfall today may depend upon rainfall yesterday, and so on).
- The general equation is:

$$Y_t = \beta_1 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_p Y_{t-p}$$

- Auto Regressive model in ARIMA:
- As illustrated, an observation Y at time t, Yt, depends on Yt-1, Yt-2, …. . Yt-P.
- The p here is called the lag order which indicates the number of prior lag observations we include in the model

THE I IN AR[I]MA: INTEGRATED

- The Integrated part of ARIMA attempts to convert the non – stationarity nature of the time – series data to something a little bit more stationary.
- Differencing in ARIMA: it is done by performing prediction on the difference between any two pair of observation rather than directly on the data itself.

$$Z_t = Y_{t+1} - Y_t \qquad \ldots d = 1$$
$$Q_t = Z_{t+1} - Z_t \qquad \ldots d = 2$$

THE MA IN ARI[MA]: INTEGRATED

- Final piece of ARIMA is the MA or moving average. It attempts to reduce the noise in out time series data by performing some sort of aggregation operation to your past observations in terms of the residual error **ε.**

$$Y_t = \beta_2 + \omega_1 \varepsilon_{t-1} + \omega_2 \varepsilon_{t-2} + \ldots + \omega \, \varepsilon_{t-q} + \varepsilon_t$$

- Moving Average in ARIMA
- The **ε** terms represent the residual errors from the aggregation function and q here is another hyperparameter that is identical to p.
- q specifies the time window for the moving average's residual error.

ARIMA MODELS

- Univariate ARIMA (p,d,q) is a forecasting technique that projects the future values of a series based entirely on its own inertia or lagged values (p,d,q are parameters & non – negative integers)
- Its main application is in the area of short – term forecasting and requires at least 40 historical data points.
- It works nest when the data exhibits a consistent pattern over time with a number of outliers.
- Sometimes called BOX – Jenkins (after the original authors)
  - ARIMA is usually superior to exponential smoothing techniques when the data is reasonably long and the correlation between past observations is stable.
  - If the data is short or highly volatile, then some smoothing ,method may perform better.
  - If you don't have at least 38 data points, you should consider some other method than ARIMA.