Data Analysis Course

Time Series Analysis & Forecasting(Version-1)

Venkat Reddy

Data Analysis Course

- Data analysis design document
- Introduction to statistical data analysis
- Descriptive statistics
- Data exploration, validation & sanitization
- Probability distributions examples and applications
- Simple correlation and regression analysis
- Multiple liner regression analysis
- Logistic regression analysis
- Testing of hypothesis
- Clustering and Decision trees
- Time series analysis and forecasting
- Credit Risk Model building-1
- Credit Risk Model building-2

Note

- This presentation is just class notes. The course notes for Data Analysis Training is by written by me, as an aid for myself.
- The best way to treat this is as a high-level summary; the actual session went more in depth and contained other information.
- Most of this material was written as informal notes, not intended for publication
- Please send questions/comments/corrections to venkat@trenwiseanalytics.com or 21.venkat@gmail.com
- Please check my website for latest version of this document

-Venkat Reddy

Contents

What is a Time Series

- Time series data is a sequence of observations collected from a process with ***equally*** spaced periods of time.
- Examples
  - Dow Jones Industrial Averages
  - Daily data on sales
  - Monthly inventory
  - Daily Customers ,
  - Monthly interest rates, costs
  - Monthly unemployment rates,
  - Weekly measures of money supply,
  - Daily closing prices of stock indices, and so on

Forecasting Methodologies

- There are many different time series techniques.
- It is usually impossible to know which technique will be best for a particular data set.
- It is customary to try out several different techniques and select the one that seems to work best.
- To be an effective time series modeler, you need to keep several time series techniques in your "tool box."
- Simple ideas
  - Moving averages
  - TSI method
- Complex statistical concepts
  - Box-Jenkins methodology

Building Model Using TSI Method

- Very Simple technique
- Spreadsheet is sufficient
- Less ambiguous
- Easy to understand & Interpret
- Serves the purpose most of the times

Components of a Time series

1. **Secular Trend(T):** Gradual long term movement(up or down). Easiest to detect
   - Eg: Population growth In India
2. **Cyclical Patterns(C):** Results from events recurrent but not periodic in nature. An up-and-down repetitive movement in demand. repeats itself over a long period of time
   - Eg. Recession in US Economy
3. **Seasonal Pattern(S):** Results from events that are periodic and recurrent in nature. An up-and-down repetitive movement within a trend occurring periodically. Often weather related but could be daily or weekly occurrence
   - Eg. Sales in festive seasons
4. **Irregular Component(I):** Disturbances or residual variation that remain after all the other behaviors have been accounted for. Erratic movements that are not predictable because they do not follow a pattern
   - Eg. Earthquake

Building Time Series Model: TCSI

O(t) = T(t) + S(t) + I(t) or O(t) = T(t) * S(t) * I(t)

Where O(t) = observed series, T(t) = Trend component, S(t) = Seasonal , I(t) = Irregular component

- **Step 1 :** Smooth the series & de trend the series

$$\frac{O_t}{\hat{T}_t} = \frac{T_t \times S_t \times I_t}{\hat{T}_t} \approx S_t \times I_t$$

- **Step 2 :** Find out Seasonal component and adjust the data for seasonality

$$\frac{O_t}{\hat{S}_t} = \frac{T_t \times S_t \times I_t}{\hat{S}_t} \approx T_t \times I_t$$

- **Step 3 :** See if there is still some trend/seasinality in the data & quantify it

$$\frac{T_t \times I_t}{\hat{T}_t} \approx I_t$$

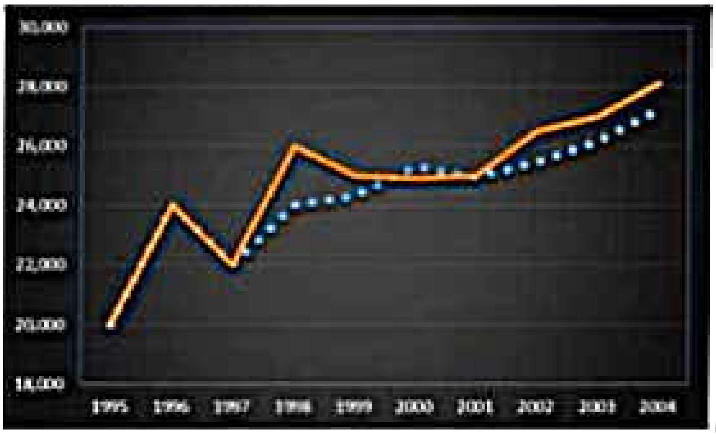
Step-1: Smoothening the series

- **Why Smoothening ? Basically to find average value and detrend the series**
- **After removing the effect of trends series is left with seasonal and irregular components**

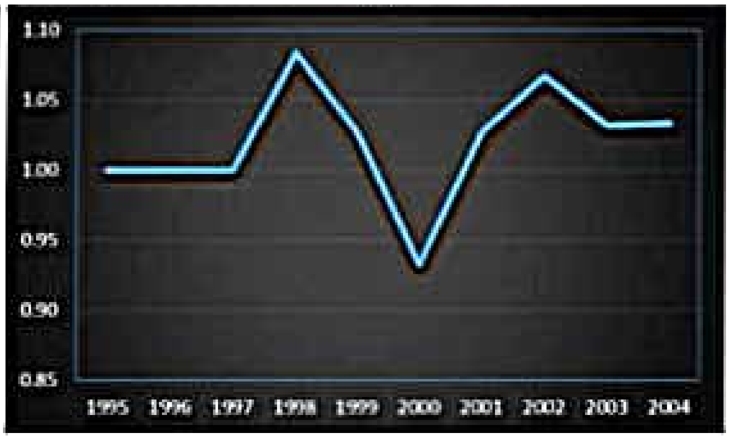
Moving Average

- **Series of arithmetic means, Used only for smoothing. Provides overall impression of data over time**
- $MA(3) = (y_t + y_{t-1} + y_{t-2})/3$

Actual Series



Series after Removing the Trend



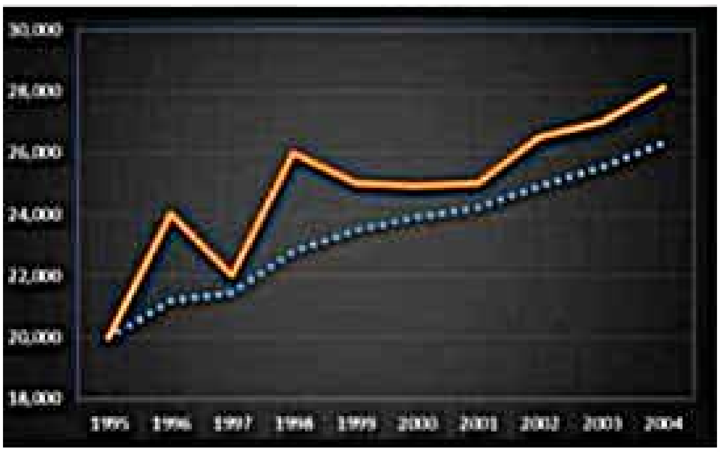
Smoothening the series

Exponential Smoothing

- **Form of weighted moving average.** Weights decline exponentially and most recent data weighted most
- **Requires smoothing constant (W).** Ranges from 0 to 1, Subjectively chosen
- Involves little record keeping of past data

$$E_i = W \cdot Y_i + (1 - W) \cdot E_{i-1}$$

Actual Series



Series after Removing the Trend

Step-2 : Capturing Seasonality - Seasonal indices

Year	Quarter	Sales
1	Q1	362
1	Q2	385
1	Q3	432
1	Q4	678
2	Q1	382
2	Q2	409
2	Q3	498
2	Q4	642
3	Q1	473
3	Q2	513
3	Q3	582
3	Q4	789
4	Q1	544
4	Q2	582
4	Q3	681
4	Q4	899
5	Q1	628
5	Q2	707
5	Q3	773
5	Q4	1008

- Quantify the effect of seasonality in each year
- Calculate overall average for each quarter for across all years
- De seasionalize the dataset with by diving with seasonal indices

Seasonal Indices & De Seasionalising

Step 1 : Reformating the Dataset

	Qtr 1	Qtr 2	Qtr 3	Qtr 4
Y1	362	385	432	678
Y2	382	409	498	642
Y3	473	513	582	789
Y4	544	582	681	899
Y5	628	707	773	1008

Step 2 : Calculation Of Seasonal Indices

	Qtr 1	Qtr 2	Qtr 3	Qtr 4
Y1	0.78	0.83	0.93	1.46
Y2	0.79	0.85	1.03	1.33
Y3	0.80	0.87	0.99	1.34
Y4	0.80	0.86	1.01	1.33
Y5	0.81	0.91	0.99	1.29
SI	0.80	0.86	0.99	1.35

Step 3: Deseanolised Dataset

	Qtr 1	Qtr 2	Qtr 3	Qtr 4
Y1	454	446	436	502
Y2	479	474	503	475
Y3	594	594	588	584
Y4	683	674	688	666
Y5	788	819	781	746

Step-3: Irregular Component

- Irregular→ No Trend → Cant be modeled
- After first level of de trending & de seasionalising, if there is still any pattern left in the time series, repeat the above steps once again

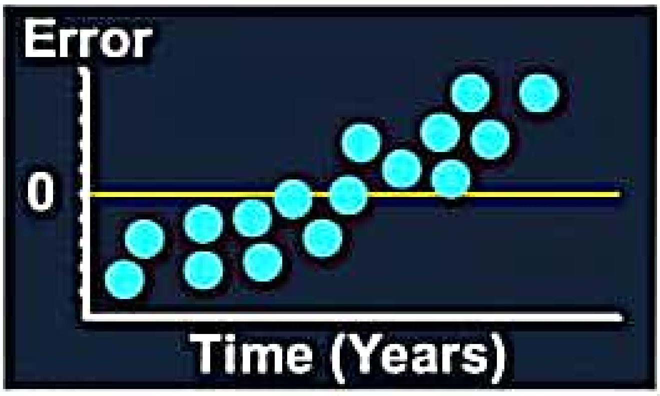
Lab

- Download the data from here Time Series Data
- Draw the trend chart, can you see trend
- Calculate MA2 and MA3
- De-trend the series and create the graph again
- Is there any trend now?
- Can you judge the seasonality in de trended data?
- Find seasonality indices for each quarter
- De-seasinalise the series by diving with seasonality indices
- Forecast sales for next four months, one by one
    - First assume I as 1, multiply with S, multiply with MA

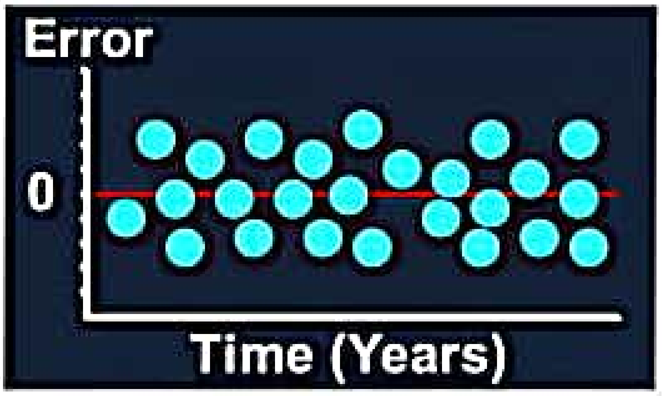
Residual Analysis

- Residual analysis is done after the model is built
- There should be no Pattern (information) left in the residuals.
- Residual should be totally random

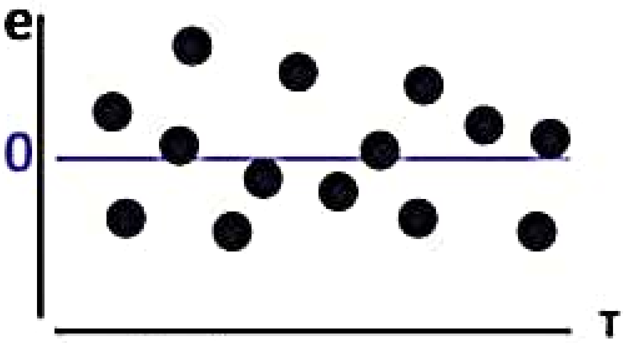
Trend Not Fully Accounted for



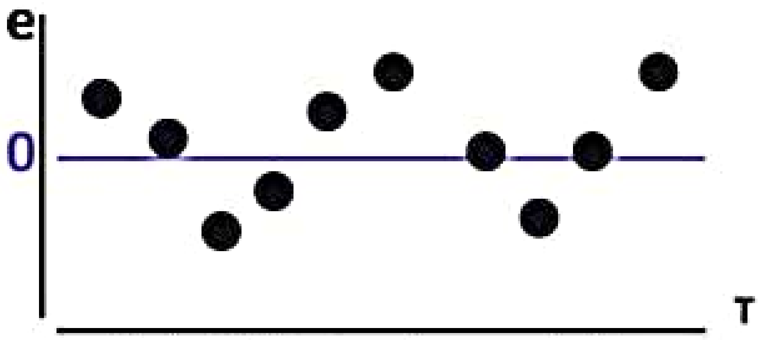
Desired Pattern



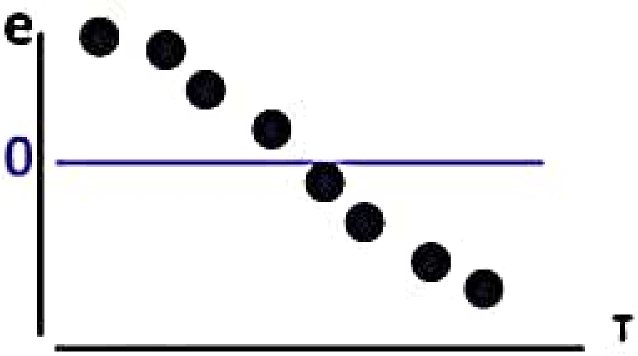
Residual Analysis



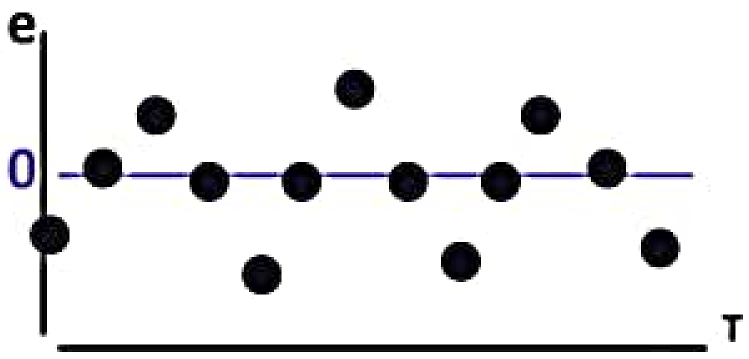
Random errors



Cyclical effects not accounted for



Trend not accounted for



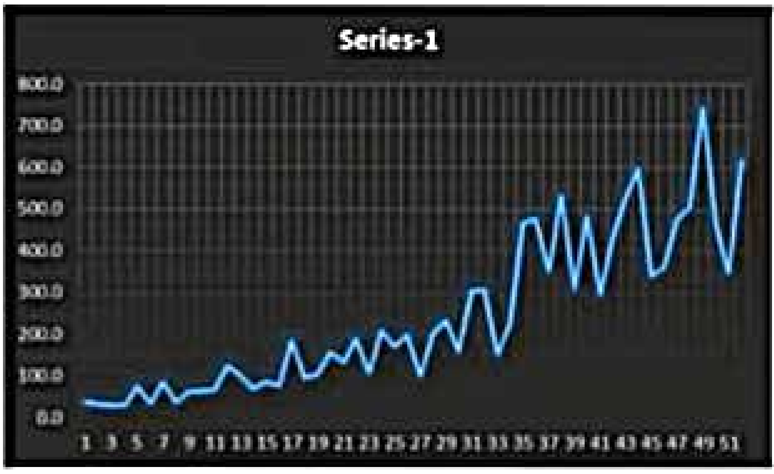
Seasonal effects not accounted for

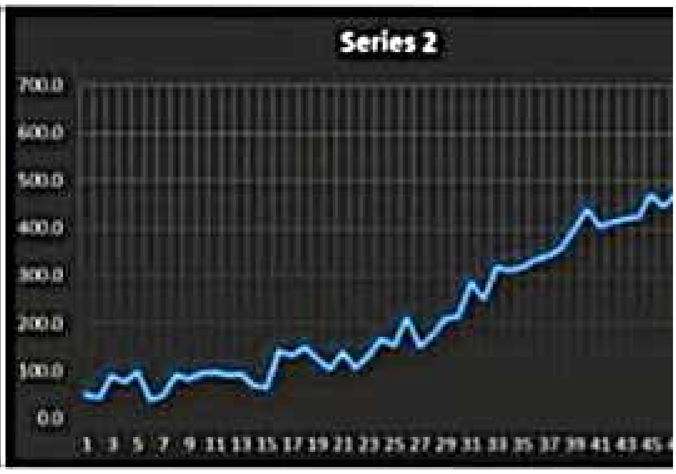
Goodness of fit

- We need a way to compare different time series techniques for a given data set.
- Four common techniques are the:
  - Mean absolute deviation, $\text{MAD} = \sum_{i=1}^{n} \frac{\left|Y_i - \hat{Y}_i\right|}{n}$
  - Mean absolute percent error $\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \frac{\left|Y_i - \hat{Y}_i\right|}{Y_i}$
  - Mean square error, $\text{MSE} = \sum_{i=1}^{n} \frac{\left(Y_i - \hat{Y}_i\right)^2}{n}$
  - Root mean square error. $\text{RMSE} = \sqrt{\text{MSE}}$

Additive and Multiplicative. What model to use?

- Is there any striking difference between these two time series?





Additive and Multiplicative. What model to use?

- In many time series, the amplitude of both the seasonal and irregular variations increase as the level of the trend rises. In this situation, a multiplicative model is usually appropriate.

- In some time series, the amplitude of both the seasonal and irregular variations do not change as the level of the trend rises or falls. In such cases, an additive model is appropriate.