(5-7 mks) Module 4
Q. Detailed ans. on: removing a trend (or steps to remove or de-trend)
→ write everything in the below order
(till [V] → applying moving window on log transformed)

PAGE NO. (.......)
DATE / /

Models with Trends and Multi-Equation Time Series

* Removing a trend:

A trend is a continued increase or decrease in a series of time.
A systematic change in a time series does not appear to be periodic is known as trend.

- Types of trends -

1. Deterministic Trend - A trend that continuously increases or decreases.

2. Stochastic Trend - A trend that increases or decreases inconsistently with lots of fluctuations.

- Identifying a trend -

1. Plot a time series data using HP Filter
2. Using the method of least squares

- Stationarity and non-stationarity in trends -

- A time series with a trend is called non-stationary.
- A trend can be modelled and removed from time series data by de-trending a time series.
- If a dataset does not have a trend or the trend was successfully removed then the data can be called as stationary.
- Following are the methods to remove or de-trend a time series:

1. De-trend by differencing - A new series is constructed where the value of the current time step is calculated as the difference between the original observation and the observation at the previous time stamp.

Q. Explain unit root in time series.
(combine 2 answers (this module + previous))
Q. Explain unit root for ϕ or θ value.
→ previous module

PAGE No.
DATE / /

observation

value(t) = OBS(t) − OBS(t−1)

ii. Log transformation - Apply log transformation on each individual value of time series data. This reduces the variance in most cases.
Example: np.log(x)

iii. Power transformation - Apply power transformation on each individual value.
Example: $x^{0.5}$ or x ** 0.5

iv. Applying moving window function - Calculate rolling mean over a period of 12 months and subtract it from original data.
values of entire data
Example: rolling(window=12).mean() − x_t

v. Applying moving window on log transformed function - We can apply more than 1 transformation for example: Applying log transformation to a time series, then take rolling mean over a period of 12 months and then subtract rolled time series from log transformed series to de-trend.

(3.5) (2.8) (~~2.8~~ 2.5)
(remember: 1%, 5%, 10%)

v.imp * Unit Roots - A testing mechanism:

write this for combined ans of unit roots.

(We cannot derive just from graphs if data is stationary or non-stationary, so we use Augmented Dicky Fuller Test)

- In statistics, a unit root test tests whether a time series variable is non-stationary and contains a unit root
- The null hypothesis, generally defined as the

presence of unit root and the alternate hypothesis states either the data is stationary or not.

- A dicky-fuller (DF) test tests the null hypothesis in an AR time series model. However, an augmented dicky fuller (ADF) test is an extension of DF test which removes autocorrelations from the series.
- ADF test is a statistically significant test which means that the test will give a result in hypothesis, i.e. Ho and HA.
- ADF function (adfuller) provides the following functionality:

i. p-value
ii. test-statistic
iii. no. of lag observations used
iv. critical value such as 1%, 5%, 10%.

1% → −3.435
5% → −2.863
10% → −2.567

- Condition to reject null hypothesis - If test statistic is less than critical value AND p-value is less than 0.05, then reject null hypothesis that means the time series does not have a unit root, that is the data is stationary.

Example 1 : Determine from the below observation if the time series data is stationary or not using ADF test and give proper justification.

Test statistic : −1.3380

p-value : 0.6115

no. of observation used : 1258

critical value (1%) : −3.435

* { once write full form of ADF and then we can write ADF only.

PAGE No.
DATE / /

critical value (5%) : -2.683
critical value (10%) : - 2.567
condition (Justification)

→ Test statistic < critical value AND p-value < 0.05
As per the above observation,
- 1.3380 > critical value AND 0.6115 > 0.05
which fails to reject the null hypothesis that means the data is non stationary and it contains unit roots.

(5-7 mk) Example2 : Determine from the below observation and find the values for the following questions:
i. complete the missing values of critical values given.
ii. state if the time series data contains unit roots or no with respect to null hypothesis.
iii. Mention the statistical test used to check the unit roots in (ii).
iv. Provide proper justification.

Test statistic : -36.114
p-value : 0.01
no. of observations used : 1639
critical value (5%) : -2.683
critical value (1%) - ~~-3.435~~ ___?
critical value (10%) - ~~-2.567~~ ?

→ i. The missing values for critical values in the question are :
critical value (1%) - -3.435
critical value (10%) - -2.567

ii. The condition rejects the null hypothesis, that is the data is stationary and does not contain unit roots.

iii. The Augmented Dicky Fuller (ADF) test is used to check the presence of unit roots.

iv. Justification:
The condition states:
Test-statistic < critical value AND p-value < 0.05
so as per the observation given,
$-36.114 < 1\%, 5\%, 10\%$ AND $0.01 < 0.05$
Therefore, the data is stationary and has no unit roots as stated in ADF test.

* Intervention Analysis -
Intervention means a change to a procedure that is intended to change the value of the series ~~X(t)~~ X_t.
Suppose that at time t there has been an intervention to a time series, then we need to estimate how much the intervention has changed the series (if at all).
Example - suppose that a region has instructed a new maximum speed limit on its highway and want to learn or analyze how much the new limit has affected the accident rates.
Intervention analysis in time series refers to the analysis of how the mean level of a series changes after an intervention when it is assumed that the same model structure for the series X_t holds before and after the intervention.
The intervention model, suppose that to be ARIMA model for X_t (the observed series) with no intervention is as follows:

PAGE No.
DATE / /

$$X_t - \mu = \frac{\Theta B}{\Phi B} w_t \text{ (noise/whitenoise/seasonal noise)}$$

(ΘB → MA, ΦB → AR)

where ΘB is an MA model
ΦB is AR model
w_t is error term.

Let z be the amount of change at time t that is an intervention.

$$X_t - \mu = Z_t + \left\{ \frac{\Theta B}{\Phi B} w_t \right\} \text{ same}\uparrow$$

which is the same model including the intervention effect.

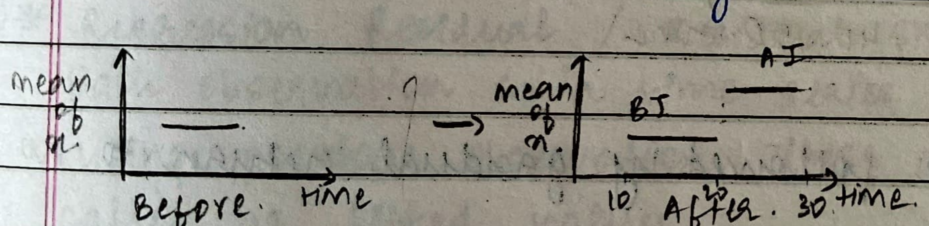
- Possible patterns for Intervention Analysis:

(7mks) Q. Explain IA with appropriate examples. { like speed camera example } (~~one para of IA~~ one para ~~to~~ of example)

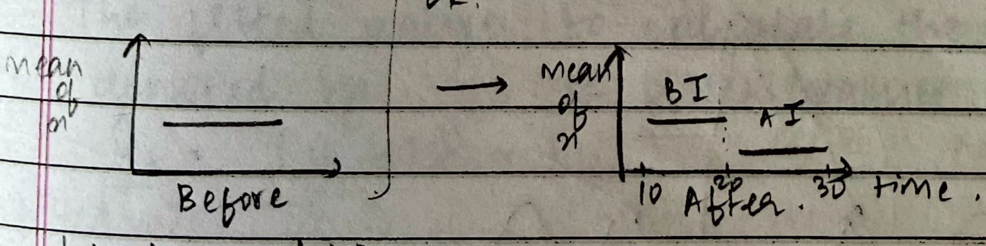
Q. only possible patterns (with ex. for each)

Q. Patterns will be given, find which pattern.

1. Permanant constant change to the mean level -

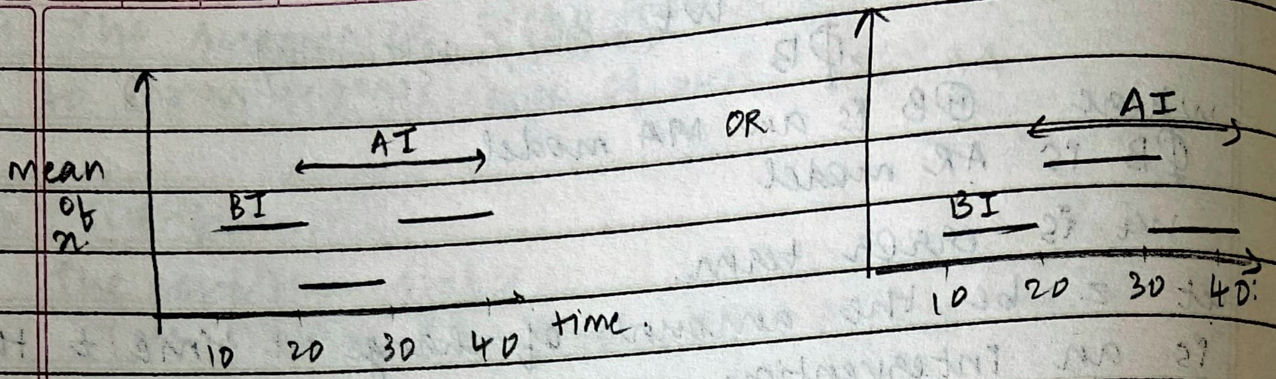


OR

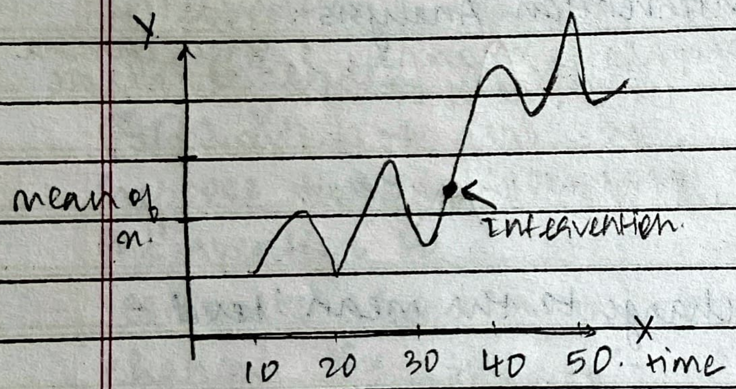


(no need to draw this)
~~just for understanding~~

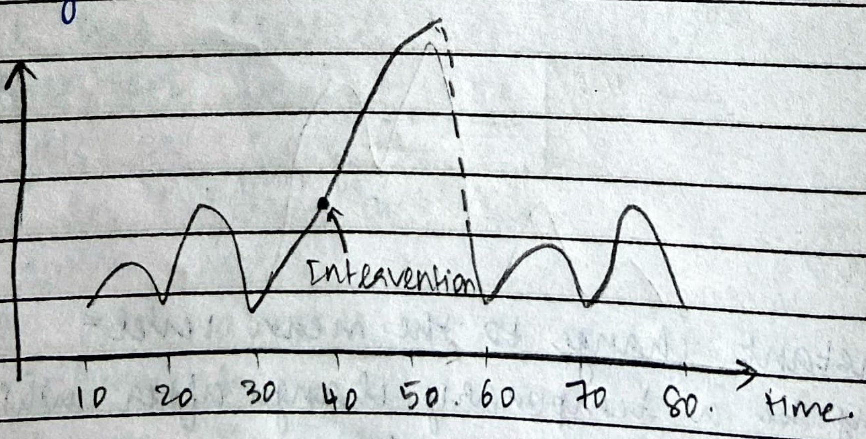
2. Brief constant change to the mean level -
There may be a temporary change after which there is no change in the intervention.



3. Gradual increase or decrease to a new mean level.
There may be gradual increase or decrease in the amount, that means the levels eventually increase or decrease to a new level compared to before intervention.



4. Initial change followed by gradual return to no change -



Q. Obs. that are forecasted using prev. values: {IA - Intervention Analysis}
method - Regression Residual / Fitted values
model - AR.

CASE No.
DATE / /

• ~~5.~~ Estimating the intervention effect -
1. Two parts of the overall model need to be estimated:
i. The basic time series model for the series needs to be analyzed.
ii. The analysis of the intervention effect
~~iii.~~

2. Steps to be followed for IA (Assuming the model to be ARIMA):
i. Use the data before the intervention point to determine the ARIMA for the series.
ii. Use the ARIMA model to forecast the values for the period after intervention.
iii. Calculate the difference between actual values after the intervention of the forecasted values.
iv. Check for the difference, if any, to determine whether there was a change after the intervention effect.

(4 mks) * Regression Residual / Residual Statistics -
Each observation in a time series can be forecasted using previous observation. These observations are called as fitted values.
The fitted value to calculate the residual is denoted by: $\hat{y}_t$ is fitted values, y_t is real value.

$$\underset{\text{fitted val}}{\hat{y}_t} = \underbrace{(y_t + 1)}_{\text{forecasted value}} - \underset{\text{real}}{(y_t)}$$ {error = next - prev.}

The residuals are equal to the difference between the observed and the corresponding fitted value which can be denoted as:

$$\underset{\text{error term}}{\varepsilon_t} = \underset{\text{real}}{y_t} - \underset{\text{residual / fitted value}}{\hat{y}_t}$$

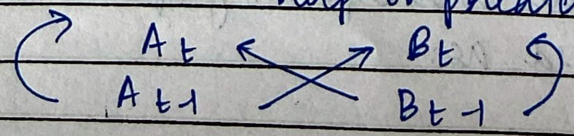
A good forecasting method will provide residuals with following properties -
- Residuals are uncorrelated
- Residuals have 0 mean
- Residuals have constant variance
- Residuals are normally distributed.

* Monte carlo Method :—
It is used to analyse complex system that involves randomness and uncertainty.

Steps
1. Data generation
2. Model selection (will include stochastic methods as well)
3. Parameter Selection
4. Monte Carlo stimulation process - generate multiple random stimulation using the estimated model and parameter.
5. Analyse

- Model 1 : → current is dependent on previous time
Vector Autoregressive Model (Multivariate model)

Example : How many apples and bananas are we going to sell in a given month ?
Two variables A_{t-1} and B_{t-1} are interactive which will help to predict the values of A_t and B_t



2 variables: $A_t = \mu + C_{11} A_{t-1} + C_{12} B_{t-1} + \varepsilon_t$

$B_t = \underset{\text{mean}}{\underline{\mu}} + C_{21} A_{t-1} + C_{22} B_{t-1} + \varepsilon_t$

$$F_t = \mu + \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} A_{t-1} \\ B_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon A_t \\ \varepsilon B_t \end{bmatrix}$$

Generalized equation:

$F_t = \mu + C F_{t-1} + \varepsilon_t$

components

- The vector method is used for a compact model, If the time period increases then the no. of equations will be equal to no. of variables.

- VAR is a multivariate forecasting algorithm that is used when two or more time series influence each other.

- The basic requirement in order to use VAR model are -

i. Need at least 2 time series variable

ii. The time series should influence each other.

Q. Why is it Autoregressive?

→ Each variable is modelled as a function of past values which are lags of the series.

Q. How is VAR model different from AR, ARMA and ARIMA?

→ • The primary difference in these models are that they are unidirectional where the predictors influence the value of y and not vice-versa whereas VAR is bidirectional and the variables influence each other.

PAGE No.
DATE / /

Q1. Apply a bidirectional model considering n variables influencing each other.
→ (1) var → definition (assume n=3)
(2) diagram
(3) ~~vector~~ 3 equations
(4) vector equations
(5) generalized equation
(6) explain each component

Q2 Apply the VAR(2) model which influence each other. (2 variables)

VIMP ★ Q3. Given a dataset containing time series data for multiple economic indicators such as GDP and inflation rates, how would you use a VAR model to forecast future values of these indicators and demonstrate the steps you would take including data preparation, model specification, model setup, estimation and result.

→ (1 variable will change)
To use a vector Autoregressive Model (VAR) to forecast values of multiple economic indicators such as GDP and inflation rates follows the following steps -

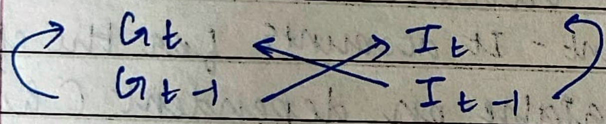
1. Data preparation - check stationarity, VAR model assumes that the time series are stationary i.e. the statistical properties do not change over time.
2. You can check for stationarity using unit root test such as ADF test. If the series are non-stationary, you need to differentiate to make it stationary.

3. Model Specification - Determine the appropriate no of lags to include in VAR model using AIC criteria.

4. Model Setup - Specify the VAR model where each variable is regressed on its own lag values and the lag values of all other variables in the system.

For example - If we have two variables: GDP and inflation rate, the model will look like below:

i. Definition - VAR is a multivariate forecasting algorithm is used when two or more time series influence each other.

ii. 

iii. $G_t = \mu + c_{11} G_{t-1} + c_{12} I_{t-1} + \varepsilon_{G_t}$

$I_t = \mu + c_{21} G_{t-1} + c_{22} I_{t-1} + \varepsilon_{I_t}$

$$F_t = \mu + \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} G_{t-1} \\ I_{t-1} \end{bmatrix} + \begin{matrix} \varepsilon_{Gt} \\ \varepsilon_{Bt} \end{matrix}$$

$F_t = \mu + C F_{t-1} + \varepsilon_t$

c_{11} ... ⇒ coefficients of the lags of respective variables

μ ⇒ mean

G_{t-1}, I_{t-1} ⇒ previous lag

5. Estimation - After fitting the model, check for the residual with respect to forecasted or predicted values.

6. Results - Access the accuracy ~~criteria~~ of the forecast using metrics like MAE and RMES.

* Autoregressive Distributed lags (ADL):
The ADL model in TSA is used to analyse the long run relationship and short run dynamics between variables.
It is particularly useful for examining relationship between dependent variables and one or more independent variables.
The ADL model ~~follows~~ combines the elements of Autoregressive (AR) and distributed lags (DL) as follows:

i. Autoregressive component - It captures short term dynamics of dependent variable (current with past).
ii. Distributed lags component - It accounts for the effect of the independent variable on dependent (how independent variable influences the dependent variable).

Independent (AR)

$$Y_t = (\alpha + \beta_0 x_t + \beta_1 x_{t-1} + \ldots + \beta_p X_{t-p}) + (\delta Y_t + \delta Y_{t-1} + \delta Y_{t-2} + \ldots \delta_q Y_{t-q}) + \varepsilon_t$$

dependent (DL)

* Transfer functions:
A transfer function is a mathematical model used to determine the relationship between two or more time series variable.
It is common tool to analyse the model on how one time series variable affects or influences other.

Monte - Cario Method :

used to analyze complex systems that involve randomness and uncertainty

- Monte cario simulation /method / model is used to predict the probability of a variety of outcomes when the potential for random variables is present
- It is also referred as a multiple probability simulation
- A monte cario method requires assigning multiple value to an uncertain variable to achieve results to obtain an estimate
- The probability of varying outcomes cannot be firmly pin-pointed because of random variable inference. therefore, a monte-cario method focuses on constantly repeating random samples.
- step by step process of how monte-cario simulation is applied in time-series analysis :

(i) Data generation : collect or obtain historical TS data that needs to be used as a basis for simulation.

(ii) Model selection : choose an appropriate TS model such as ARIMA, GARCH, any other stochastic model.

(iii) Parameter selection: choose appropriate model paramet
such as p, d, q, etc
(iv) Monte-Carlo simulation process:
a. Generate multiple random simulation using the estimated model and parameter
b. vary the key factor or assumptions in each simulation as needed to explore different scenarios.
c. Repeat the simulation process multiple times to create a distribution of ~~pro~~possible outcomes.
(v) Analysis :- Draw conclusion, make forecasts, access the risks or evaluate the impact of different scenario.