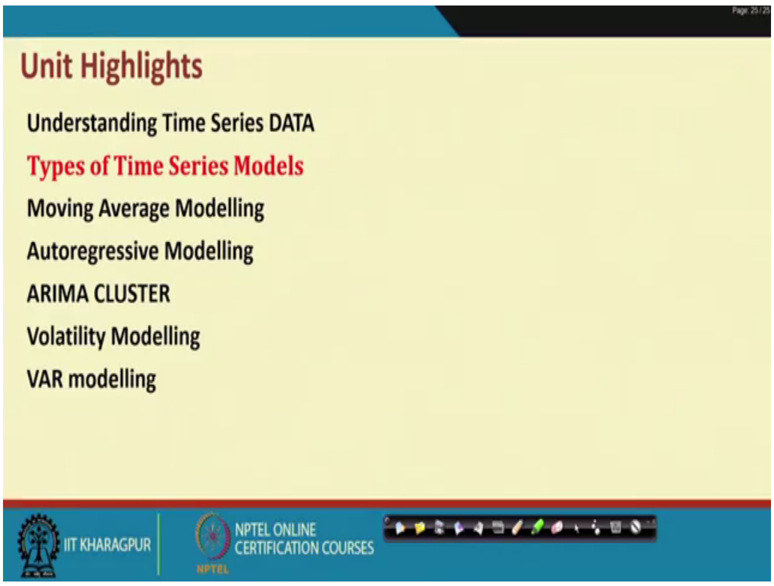
Engineering Econometrics
Prof. Rudra P. Pradhan
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

Lecture – 45
Time Series Modelling- Stationarity

Hello everybody, this is Rudra Pradhan here. Welcome to Engineering Econometrics that too the coverage on Time Series Modelling. We have already discuss various time series models and various mechanism through which you can receive the estimated model and then we check the reliability, validity and stability by using various you know time series indicators you know such as mean error, mean square error, root mean square error, mean absolute deviations, mean absolute percentage error. So that means, we have actually various alternatives you know indicator through which you can you know check the reliability, check the validity of the estimated models and then do the predictions as per the particular you know requirement.

So, now to extend this lines of you know discussion, we will see what are the other ways we can do the similar kind of you know structuring and restructuring and through which you can actually this is the best model as per the particular you know requirement and the kind of you know need industry or any kind of an organisation ok. So, by the way we have already discuss various forms of you know time series model.

(Refer Slide Time: 01:35)



And in this lecture we specifically highlight a component called as you know ARIMA Cluster. And it is also one kind of you know forecasted model through which you can do the forecasting. Against we start with a single variable and then we create a kind of you know estimated you know structure through which you can do the prediction of the same variables.

So, now ARIMA is a kind of you know structure through it you can do the similar kind of you know forecasting. So, ARIMA stands for Auto Regressive Integrated Moving Average and here we have 2 different you know structure one is the order regressive structure and the second one is the moving average structure. So, now, the issues here first you can have a actual y, then we may have a predicted y; after you know having the predicted then you will get the error component which is the difference between the actual and the estimated. Now once you get the error term, through the error terms we can create you know the kind of you know different models; there are 2 types of you know such models. So, one case you have to just connect with the y t with the error terms, then another case y t with the lag of these you know y t is for instance y t equal to yt minus 1 y t minus 2 and so on.

And that is the difference between these 2 that is moving average modelling and autoregressive modelling. And then if you know connect then that becomes you know called as you know ARIMA models. Then in between there is a component called as a integration that is the actually term called as you know order of integration; we technically in the time series angle called as you know stationarity of that particular variable. And there are couple of you know techniques are there to check the stationarity and in time series analysis again this particular component we called as you know unity root.

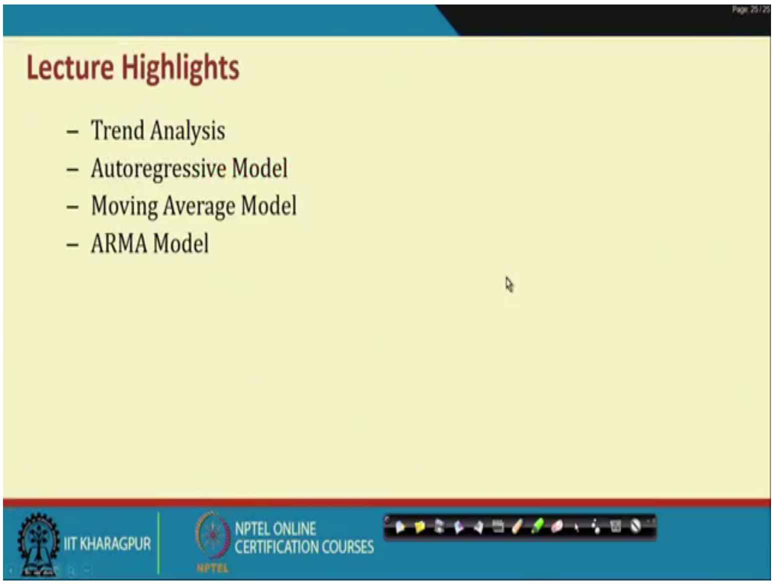
So, now you will like to check the unit root; that means, the objective is to check what is the order of integration through which the variables you know becomes you know stationarity; reaches you know stationary so; that means, stationary means it is in normal conditions, but usual behaviour of time series variable in original form is not you know is basically non stationary. That means, there is a huge volatile and imbalance, but we like to know at what level the particular variable will be in a kind of in the stable form or you know balancing form. And, we have will bring that that is what the process called as you know unit root check or you know stationarity check.

Then after knowing the personality of a particular variable; that means, the order of integration through which the variable can these the stationarity, then you can use ARIMA clusters, then are VAR clusters, then volatility clusters. So, these are all actually 3 different clusters of you know time series modelling. So, in this lecture specifically we highlight the ARIMA clusters that too autoregressive structure and moving average structure. And, the procedure through which you can have this model is that you know first we have actual then the estimated and then you find out the error terms and then start with you know integrating with lag of these error terms and lag of the original variables.

So, how many lag of the actual variable and how many lag of the error term should be there, that is the kind of you know question mark and we have 2 procedure for that. In the first procedure we like to know what is the order of integration through which the variable reaching the stationarity, in the second case how many such lag variables can be finally, used so, that we can do the prediction perfectly as per the particular you know engineering requirement. So, that means, technically we have 2 different requirements here that need to be tested first or that need to be evaluated first that is the order of integrations that too the stationarity check, and the second one is the you know lag length.

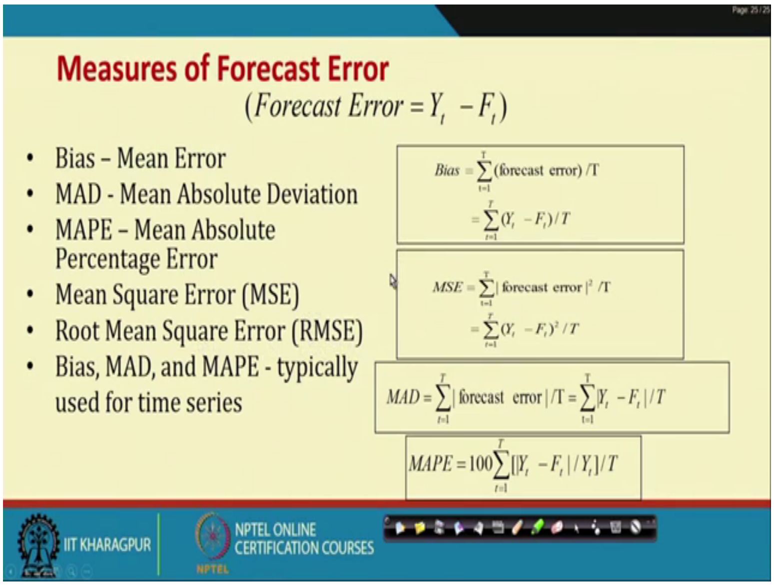
So, we need optimum lag length through which the model can be used as a indicator through which you do the forecasting. Of course, while checking the unit root test you know we also use the optimum lag concept, and then we can go for the autoregressive schemes or moving average schemes or the joint one the autoregressive and moving average scheme. In fact, the stationarity check and the kind of you know order of integrations are you know somewhat mandatory requirements for ARIMA cluster, volatility cluster and VAR clusters. So, until unless you know the order of integrations you cannot just analyse the ARIMA structure, we cannot just analyse the volatility cluster and then the VAR cluster.

(Refer Slide Time: 07:07)



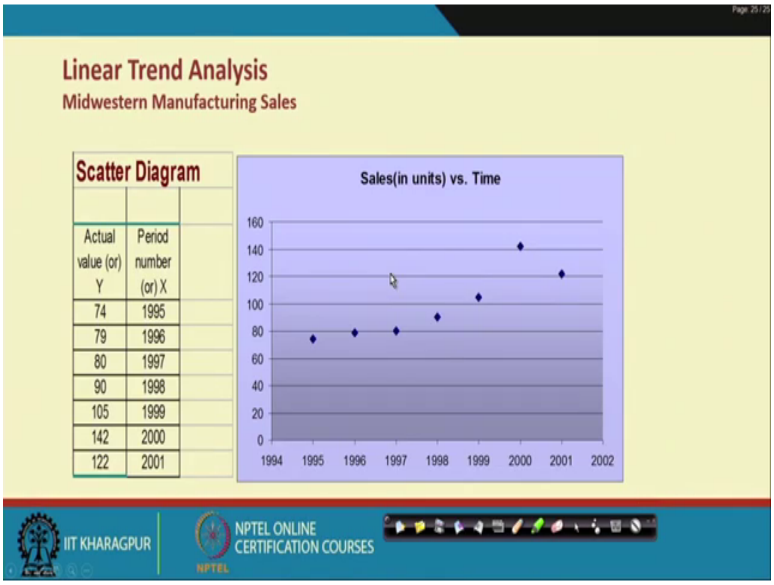
So, let us see how is this ARIMA cluster. So, basically it is the game of you know trend analysis; and again we start with the trend analysis autoregressive model, moving average models. And, then the you know kind of you know clubbing, that is the ARIMA models autoregressive moving average model.

(Refer Slide Time: 07:27)



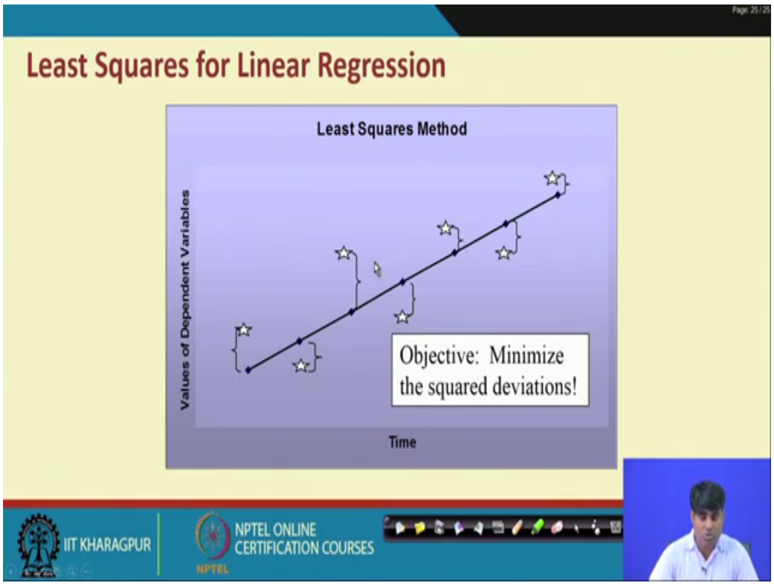
So, again these are the same forecasted errors are used like you know mean error, mean absolute deviation, mean absolute percentage error or mean square error root mean square error. So, again same indicators are used to validate the structure.

(Refer Slide Time: 07:39)



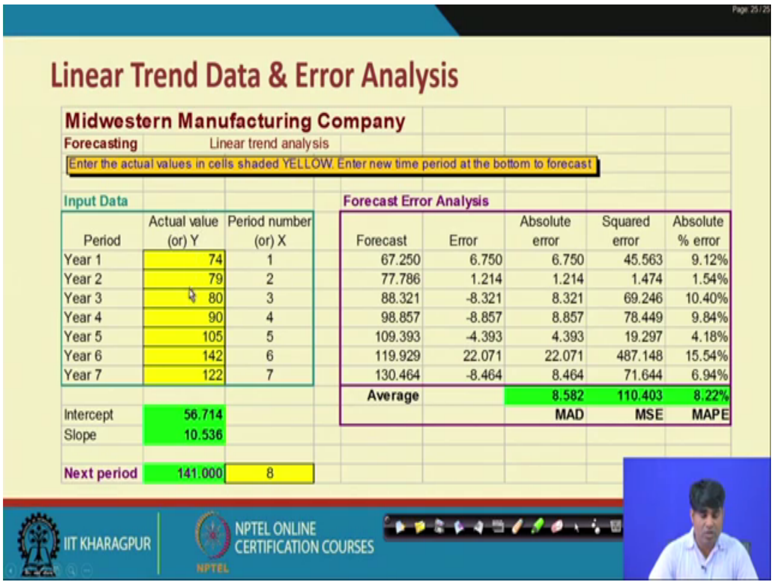
In fact, a here it is the kind of you know examples Y and X, where Y is dependent variable, X is independent variable then you find out the estimated line then find out the error terms that is the game again a one we have to check the validations. But what we are discussing here in the trend analysis, the same variables can be integrated with it's a lag variable only.

(Refer Slide Time: 08:11)



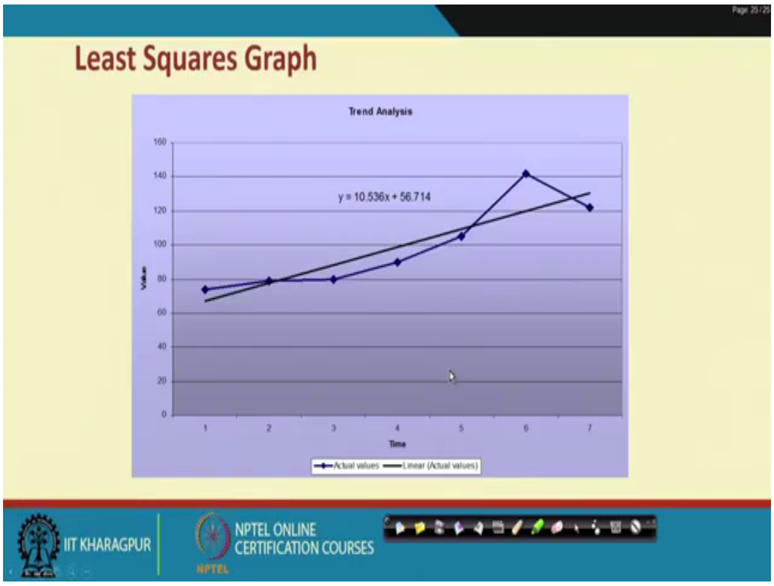
And that is that is what the beauty of this particular you know scheme. And here this square can be used as a mechanism to you know have the estimated figure.

(Refer Slide Time: 08:21)



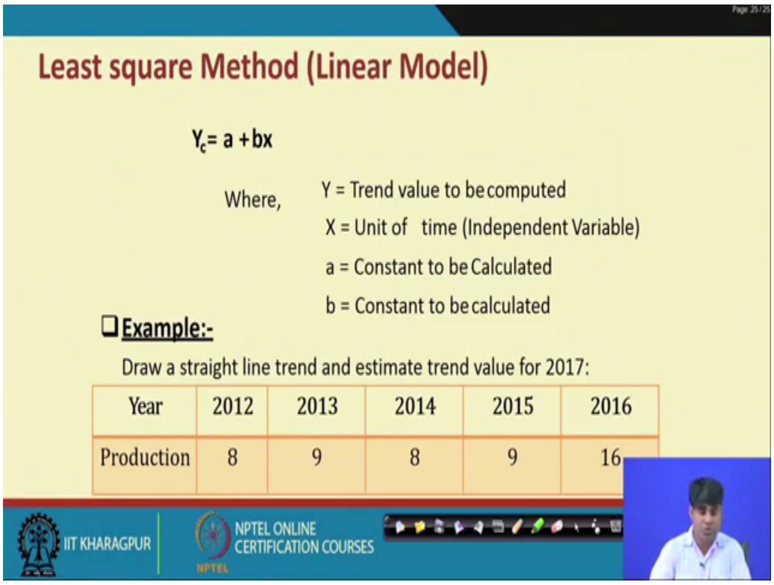
And the again this is the actual and this is the what independent variable. So, we can have the a forecasted value and then error value, and again we go for you know checking the reliability and stability of these you know forecasted figures; by checking the a mean error, mean absolute error and mean square errors, root mean square error and so on.

(Refer Slide Time: 08:45)



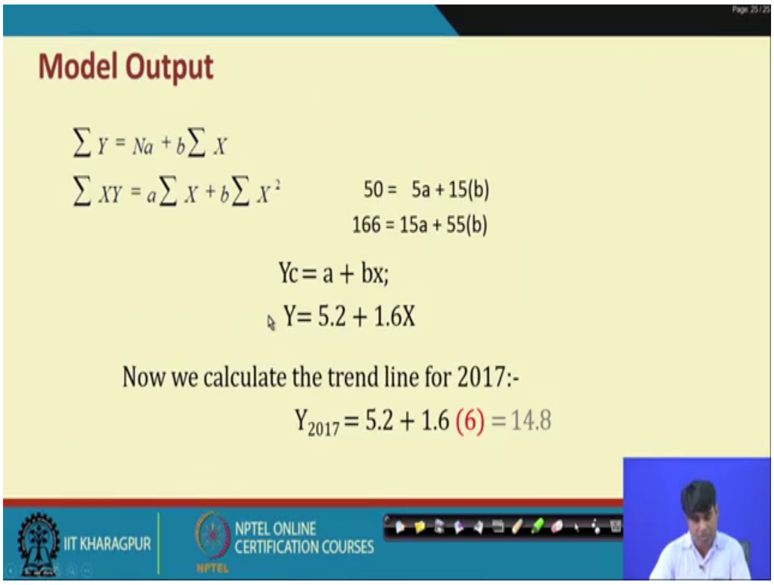
So, here in the trend analysis, by default you may you will have actually a actually friction actual observations that is what the blue one and then the predicted one that is the black one, which is slightly a linear kind of you know things. And the actually the slightly non-linear kind of you know things; that means, it is not straight. And, the mechanisms through which you use trend and the kind of you know forecasting procedure that, we need actually a trend line through which actually you can visualise and do the prediction perfectly as per the particular you know requirement.

(Refer Slide Time: 09:29)



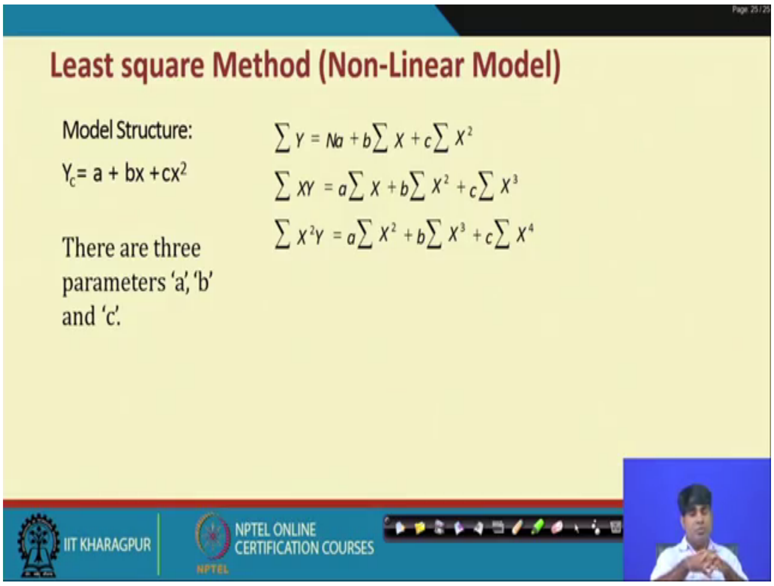
So, now know the kind of you know least square mechanisms can be used to do the forecasting so, that is what with respect to 2 variable.

(Refer Slide Time: 09:35)



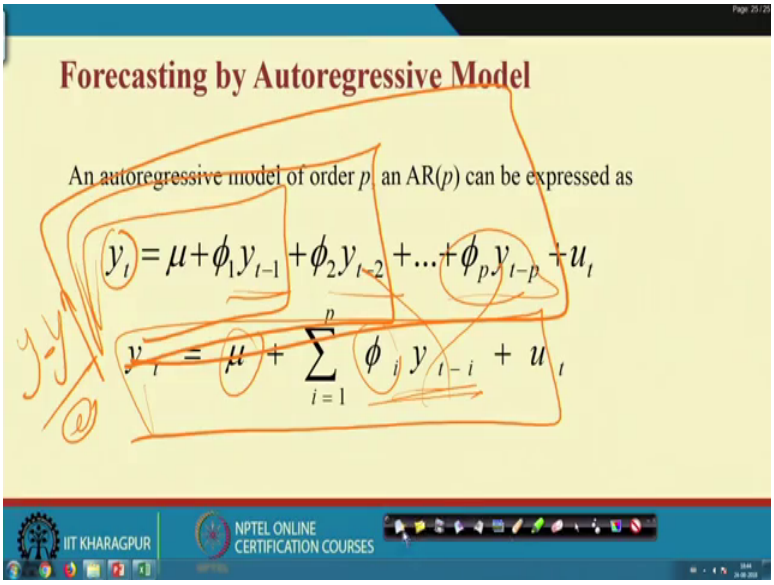
And same we know Euler's mechanisms can be used to minimise the structure and then we have these parameters, and then we have the estimated model. Now what is happening here we are just doing Y and x, and here the this will just interchange while doing the time series instead of x we can put Y Y t minus 1, then you have the similar kind of you know structure and get the trend line.

(Refer Slide Time: 10:03)



Again if you are connecting with the you know say X and X square, now in the time series set ups you can have Y Y t minus 1 and Y t minus 2. So, in that case you have 3 parameters and as a result you have a 3 different a equation and then if you simplify you will by default you will get the parameters value a b c, and that will used as a you know indicator through which you can do the forecasting.

(Refer Slide Time: 10:29)



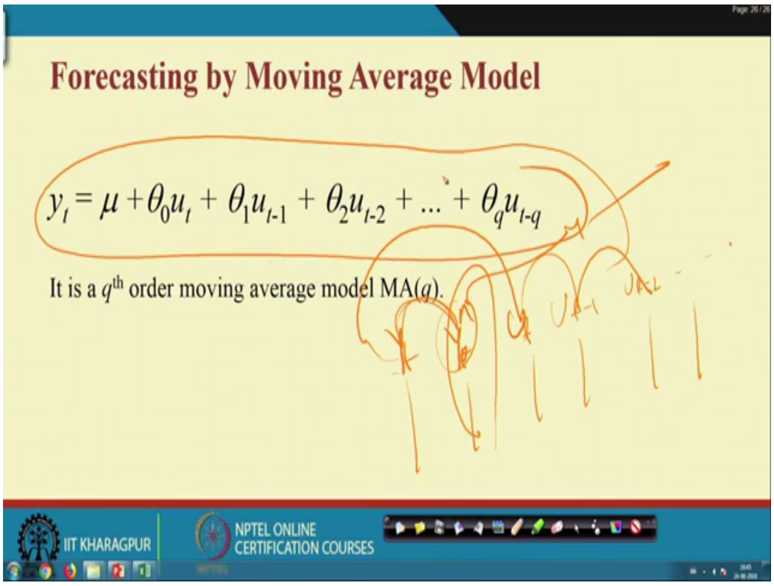
Now, after knowing this kind of you know trend structure, we can move into the autoregressive scheme here means first instance you should have estimated you know estimated equations, that is that is what the structure we called as you know y and y head right. So, that that see here y head and then we can have y a means y and y head, and technically y head is the estimated one and then we have y; so, the difference by default to the give you the error component.

So, now what is happening actually, before you go for you know estimated want to get the error terms again integrated with the actual y, and you have check what is the estimated you know y for that it depends upon you know some any lag variables y t minus 1 y t minus 2 and up to y t minus p and again if you simplify the model can be transferred into this form. So, where mu is the intercept and then these are the coefficients against with the lag variables, when i equal to 1 then this is actually y t minus 1, when i equal 2 this will be y t minus 2 and so, on up to y t minus p.

Now, the question is a whether you will rested the model up to this or whether you will rested the model up to this or whether you will be rested the model up to this so; that means, it can be a simple bi variate, it can be simple trivariate or it can be simple you know kind of you know multivariate in nature. So, issue is actually how many lag variables ultimately we can use, so, that the model will be very perfect for the prediction and forecasting; that is the big deal in the case of you know ARIMA models and we are we are discussing here autoregressive scheme. So, the difference between autoregressive scheme and that moving averages that. So, the actual y will be connected with a lag of that y, and that is the form called in auto regressive structure.

When you are connecting y t with you know the error terms, then this becomes a moving average scheme. In the case of you know moving average scheme so, first you estimate the model, get the error term then against the new model will be y t versus the is you know error terms, then you check how best how good you know best we can in a pickup the things. So, now ultimately so, this is the general framework of you know autoregressive model, and then the counterpart will be moving average models. So, in that case you have the scheme here y t equal to u t u t minus 1 u t minus 2 u t minus q.

(Refer Slide Time: 13:19)



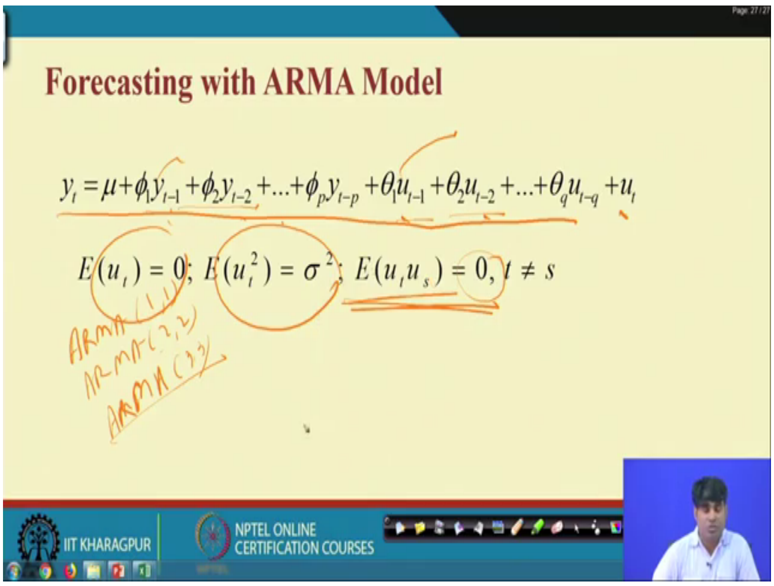
So, here you see y t as a function of you know y t minus i technically and here y y equal to y t equal to means y t as a function of you know u i for u t minus i. So, when i equal to 0, then it becomes u t and when i equal to 1 then becomes a u t minus 1, but it is the game called as you know error terms, but technically the framework is like this.

So, you have Y t, then Y t you know head then error terms which is the difference between Y t and Y t head that is the estimated and how to get the estimated that depends upon the kind of you know last structures and then you have U t and you can create U t minus 1, U t minus 2 and so, on so; that means, we have a couple of you know derive variables now after getting error terms. So, you can omit this one. So, then the new estimation process will be Y t upon U t again U t minus 1 and U t minus 2 that is what we have here actually and in this case the structure is called as you know moving average.

So, in the case of you know autoregressive, it is the actual variable with its lag integration and in the case of moving average the actual variable that is Y t with a the error component error component and the error lags. So, that is what the kind of you know structuring. So, now, against the question is the how many such lags you can use for this you know moving average. So, now, in one case we have here the kind of you know what you can call as you know the autoregressive, and then here the issue is actually optimum lag length and again in the case of moving average, there is also question of you know optimum lag.

Then now what will do we will just clog this 2 autoregressive scheme and moving average scheme, and we will have a single component and that single component is called as you know ARMA.

(Refer Slide Time: 15:39)



Of course, this is what the ARMA, ARMA clusters this is what the ARMA cluster and what is happening here? What is happening here. So, you will just you know integrate with the y t with the lag of the y t and then lag of the error term so; that means, we have scheme called as you know ARMA 11, then ARMA 2 2 and so, on like this ARMA, ARMA, ARMA 3 3 like this right so; that means, technically see here if it is 1 1 then;

that means, y t minus 1 and then u t minus 1. So, then 2 2 means y t minus 1 y t minus 2 then u t minus 1 u t minus 2; so, when ARMA 3 3 then y t minus 1 y t minus 2 y t minus 3, then u t minus 1 u t minus 2 and u t minus 3. So, that is what the linkage is all about.

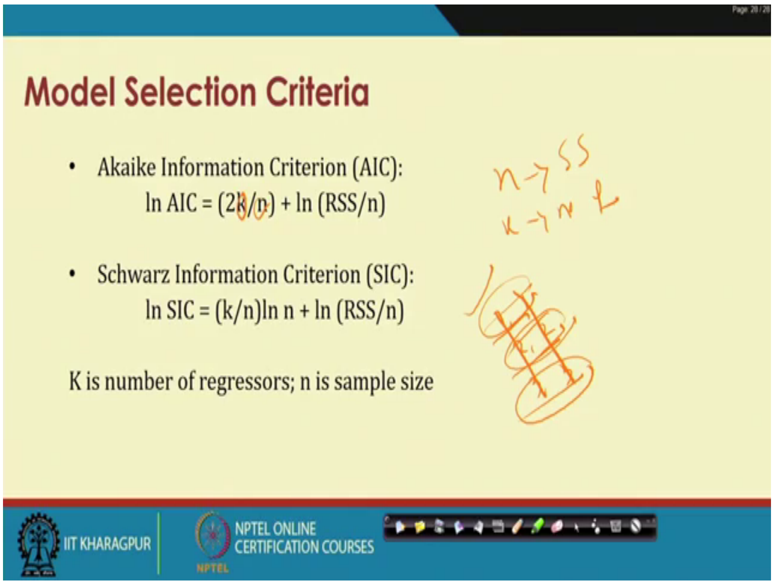
So, likewise you can go you know go on adding you know one after another y lag and the error lag. So, ultimately you start with a single variable y t, and then we have a plenty of you know lag y t again we derive error component and create n number of you know lag of the error term and then you are integrating y t with you know lag of these y t and then the lag of the error terms.

Then finally, you have the system through which actually do the predictions and do the forecasting, and you know and you know check the reliability and stability by using this indicators like you know again mean error, mean absolute percentage error, mean square error, root mean square error. So; that means, various ways of you know getting the estimated equation and then check the reliability stability before you use this model for any kind of you know engineering predictions, and the kind of you know the kind of you know requirement.

Of course, we can check the reliability by taking the clue of you know mean of the term should be equal to 0 like this here, then variance of the error term should be equal to sigma square. And if variance of the error term is not sigma square that is what the term technically called as you know how much scedasticity, and if there are variance is not send then we called as you know heteroskedasticity and then cross correlation that is what the term called as you know autocorrelation, that should be also not there that is called as you know 0. Of course, once you get the error term this should we satisfy before you integrate the y t with you know lag y t and the lag u t.

So, once it is clear then you can proceed for the kind of you know extension and then use the extended model for the you know kind of you know prediction and the kind of you know forecasting.

(Refer Slide Time: 18:55)



Of course, we have discuss specifically autoregressive scheme, then moving average schemes, then ARMA scheme. So; that means, all together we have 3 different clusters all together that is the under the ARIMA cluster. Autoregressive cluster one, then moving average the second, and then autoregressive moving average the third.

Every time the I n instrument that to check the order of integration through which the particular variable reaches the stationarity. Here the stationarity can we check the y variables, then y t minus 1 is then error component. So, whatever variables will be finally, use a you know in this particular destination process all the variables are individually you know tested the order of integration through which it reaches the stationarity. And of course, one of the most important issues is the a lag specifications, we need a optimum lag length through which you actually get the a best estimated model through which you do the prediction and forecasting.

So; that means, optimum lag length is the criteria, through which you can you know reach the target and reach the destiny. And we have a couple of you again criteria through which actually we can fix the optimum lag length, that is the AIC statistics and then the SIC statistics, a Akaike information criteria and Schwarz information criteria and basically this criteria exclusively depends upon you know 3 components; the item called as you know k and the n, n stands for sample size and then k stands for number of variables. If it is moving average and y t with you how many lag and then if it is moving average, then y t with the how many a lag u t again under ARMA, y t with the how many lag y t and how many lag u t ultimately; that means, we want to know whether structure is 1 1 or 2 2 or 3 3 like this.

What 1 1 particular cluster will be you know uniquely descending factor to do the forecasting you know occurs mathematically we have n number of clusters, but technically only one cluster which can give you better result of in every times with again use these indicators are the tool you know to find out the best one you know whether it is 1 1 or 2 2 or 3 3, and then we will you know use that model for the kind of you know kind of you know forecasting and the kind of you know requirement.

So, here you know issues that you know like mean error, root mean square error, mean absolute percentage error or you know kind of you know mean absolute deviation. So, every times the comparison is that you know where these parameters are you know having lag value compared to the previous one, that model will be used for the kind of you know final prediction. So, the same criteria will be used here while choosing the optimum lag length. So, you know you start with 1 1 then you have a model, and then again 2 2 we have a model whether it is a autoregressive scheme separately or moving average scheme separately or together that is what called as you know ARMA. So, you know you can start with the if it is a autoregressive, then this structure will be restricted with you 1 only, that times y t y t minus 1 and so, on.

So, every time it will go with one then again moving average. So, it is again you know means the moving average case, this will be silent and this will be changing and then in the case of autoregressive, this will be remain silent and this will be changing and again ARMA cluster both with this change simultaneously. And technically whether to stop here in the case of 1 1 whether to stop 2 2 or whether to stop 3 3 so, that is exclusively you know depends upon the AIC value and SIC value. So that means, technically you may have 3 different models of all together, in one case we have 1 1 scheme then has the estimated models and then you find out the AIC statistics then 2 2 have the estimated model outcome and then AIC statistics, then again 3 3 most have estimated model and then have the AIC statistic.

Now, you have a 3 options and out of these 3 options, where the AIC statistic will be lower one and that will be the final choice. So, the way we will choose the best one through AIC, SIC also having the similar kind of you know testing procedure and similar kind of you know evidence. If a for instance out of this 3 1 1, 2 2, 3 3 if AIC is a through the means; that means, on the basis of AIC if we find 2 2 scheme is the best one out of 1 1 2 2 then 3 3. So, this SIC statistic will also give similar kind of you know inference where 2 2 is the best ones out of 1 1 2 2 and 3 3.

So, we must you know the model selection criteria that too the choice of optimum lag length and then again you should know the model validation indicator or stability indicator starting with mean error to mean error absolute percentage error and root mean square error. And, every time is you know the criteria is actually to you know pick up the models which you have actually you know where of these you know means where values of these indicators that is you know again that is you know either AIC statistic or SIC statistic or any other time series validation indicator starting with you know mean error and then root mean root mean square error or something like that.

So, ultimately we have discussed you know various forms of you know models through which actually we can actually go for the kind of you know forecasting and then the kind of you know engineering requirement. So, ultimately so, we have various time series set ups, through which you can you know have the estimated models and then use the estimated for the prediction and forecasting. Means in the software you know you just enter the data and then there are couple of softwares by default if you connect with the ARIMA, and then you have to just fix the order you know 1 1 or 2 2 or 3 3, and by default you will to get you know various estimated output that is you know in a kind of you know simulation process, you just change the structure then you know run the models you will get different levels of you know output.

And if the you know standard software is not there like you know you use or you know start or something like that, then still you can do in the excel sheet just to have y t, then create y t minus 1 first. And, then again link with y t y t minus once and have the estimated model, and get error terms then after that you can create y t minus 2, y t minus 3, y t minus 4 just to extend this you know existing y t y t minus from the you know yt minus 1. And, then you know you just get the error terms which is actually difference between y t and y t minus 1 and then again if create u t and lag y u t minus 1 lag u t minus 2 u t minus 3 and so on.

Then finally, if you say 1 1, then definitely the dependent variable will be every times y t, then if it is 1 1 then you have to give you know means in the data analysis have to give indication about the y t minus 1 and u t minus 1 then ran the model. So, then you will have the estimated model with respect to y t, y t minus 1 and u t minus 1 then again if for 2 2 scheme, the same y t will be the dependent variable and then y t minus 1, y t minus 2 and u t minus 1, u t minus 2 are the independent variables and again run the models you will find the results under the scheme of you know 2 2.

And then finally, a under the scheme 3 3, same y t will be the dependent variables then you start with the a independent variables like u y t minus 1, y t minus 2, y t minus 3, then u t minus 1 u t minus 2 u t minus 3 and then run models you will get the a you know estimated outcomes and that too for the scheme of you know 3 into 3. So; that means, actually finally, when you will start you know moving from 1 1 2 2 2 and then 3 3 so, on then every time the model will start you know simple moving from simple to complex and where the parameters value will be start you know increasing.

So, when the parameters value will be start increasing, then definitely the significance of these parameters maybe getting affected. Because the degree of freedom will be less and less when the parameters will start increase; that means, if you start with the 1 1 2 2 moment from 1 1 2 2 2 and 2 2 2 3 3 and so on; so, every time in the number of parameters in the estimation process will be you know very high and when number of parameters are high to test and then the sample size is remain constant so; that means, in a given format. So, if you start with 1 1 2 2 3 3, the sample every time is constant for the particular you know problem. Then what is actually major problem that you know the parameters are different from 1 1 2 3 3. So, the mean different means it will be having increasing numbers. So, giving sample size and the increasing number of parameter to test by default the degree of freedom is less, and there is high chance that you know the particular variables may not be statistically significant.

So, there is need of you know justification here to pick up the optimum lag length. So; that means, choosing optimum lag lengths and lag length they technically means, you are fixing the you know variables number of variables to be used or to be included in the system, through which you know the particular engineering problem can be analysed properly without any sure or without any conflicts or without any kind of you know defects.

So, we have discuss various types of you know time series models to do this kind of you know forecasting's and after this ARIMA clusters we have 2 different more schemes, that the volatility modelling and then the VAR modelling. So, in the next lectures so, we will be start with the volatility modelling first, and then with copper VAR modelling and then we will go to other forms of you know modelling like you know panel data modelling and the count data modelling and so on. With this we will stop here today.

Thank you very much have a nice day.