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

Lecture - 43
Time Series Modelling – Trend Analysis
(Least Squares Method)

Hello everybody, this is Rudra Pradhan here, welcome to Engineering Econometrics, today we will continue with Time Series Modelling. In fact, in the last lecture we have discussed about the various methods through which we can do the time series prediction that to we have discussed freehand curve, semi average methods and moving average methods and every times we will define there is a kind of you know improvement.

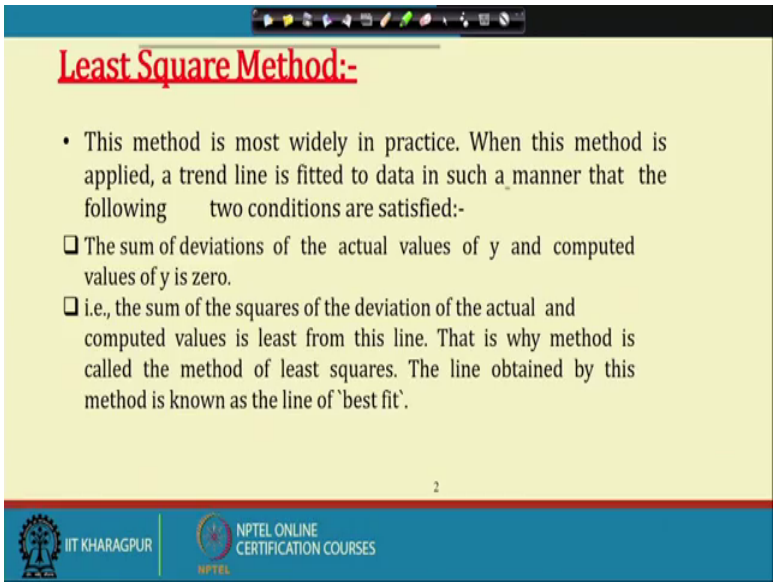
Of course, you can start with a particular problem and then you know draw the predicted line through freehand curve, then draw the predicted line with the semi average and draw the predicted line with the moving average that to may be with a 3 years maybe 5 years maybe 7 years. So that means, these are all various options again and you will find in a I mean say theoretically there is a high chance that you know getting the trend line through 7 years moving average will be much better than 5 years moving average than 3 years moving average.

And then what we can called as you know semi average and here on the 2 data points through which we can do the prediction and then I know simply you know by inspection you have to just you know draw the curve you know by checking the highest peak and lowest peak that is called as you know freehand curve. So, that means, these are all various you know mechanisms through which you do the prediction, but the last taken which you like to discuss right now is called as a least square mechanism, which is much better than the freehand curve, semi average curve. It means semi average mechanisms and then moving average mechanisms that through all forms of you know moving average starting with 3 years 5 years 7 years.

So, now will see how the least square mechanisms can help us to get the trend lines or predicted lines through which you can do the kind of you know engineering predictions, any kind of you know engineering predictions. And the trend line will be or you know least square mechanisms it will be very helpful to get the trend lines and substantially

this is very you know substantial better than other methods are compared to freehand semi average, moving average and again within a you know moving average all kinds of you know variations starting with 3 years to 7 years and so on.

(Refer Slide Time: 02:55)

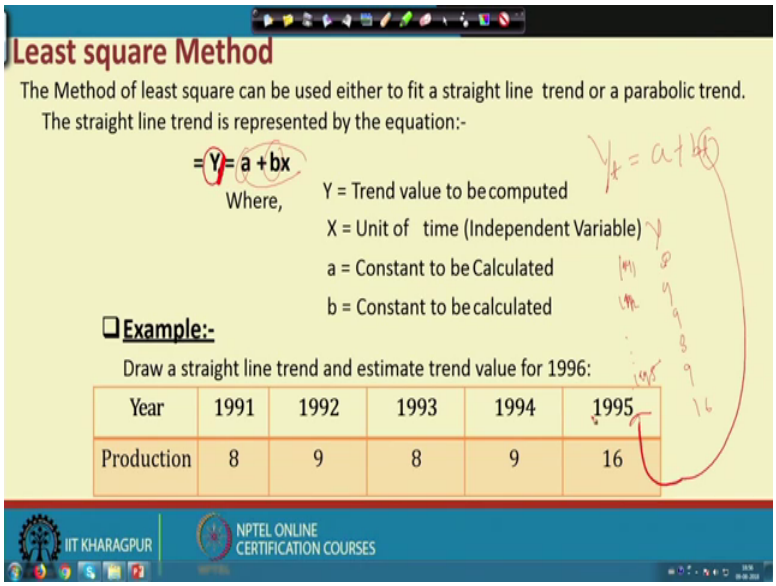


So, now what we can do technically, we will just check the least square mechanisms through which we can actually do the prediction. So, it is actually more or less seems like you know bivariate in a simply bivariate regression what is happening here is we have actually data that too let say production data that is one variables.

In fact, you know and a time in that variable is having actually a information with respect to different point of time. But, what we have discussed earlier for simple regression modelling you must have at least 2 variables let us say Y X or Y X 1 or something like that, but here is the variable is a simply Y and that too we are doing the predictions against when we use lag variables.

This simple model can be drawn into complex by you know moving towards you know trivariate into you know multivariate in nature. What is a what is actually best part of this particular process is that you know we have the you know data of a particular variable that with respect to time. And we can actually use now least square mechanism through which actually we do the you know predictions of that you know variable or subject to the you know time series structure. So, let us see how is this particular you know setup.

(Refer Slide Time: 04:27)



So, here actually it is more or less same so like you know we have discussed in the case of you know 2 variable say Y and X like this, so the predicted lines predicted line will be like this. So, Y c is equal to alpha plus beta X alpha is a parameter and beta is a parameter and then using this informations we can easily actually a predict Y. So that means, actually it is a 1991 to 1995 are to be close to we are just taking the few cluster.

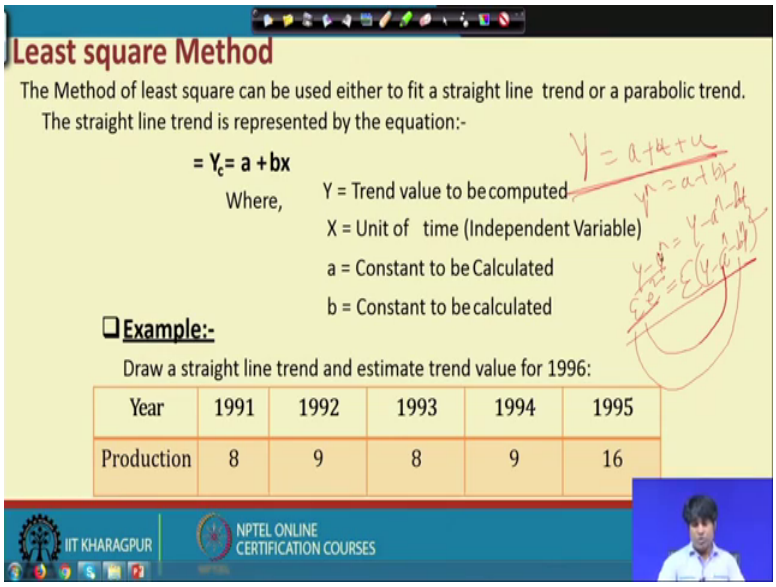
So, here what is happening so, let us say it is a 1991 then starting 1992 and you know go ahead with 1995 right. So, it is a 8 9 9 8 9 and 16 so this is a how the data information and a what will you do actually this is what actually Y and what is the predicted structure here we just take the predicted Yt equal to a plus b t. So that means, the t will be used as a you know use as a with reference to time only.

So that means, this time component by itself will be treated as a particular variable to draw the trend line, so otherwise it is actually more or less same. So, we have a Y data so now since there is you know X, so X will be represented as you know unity of time that too it is a independent variable. So that means, so this is a independent variable and t is the trend value that is what the predicted. So that means, Y is the variable that will be you know that will be predicted subject to a function of you know time that is what the a component is called as you know X here and a as usual simple regression modelling.

So, we look for the alpha parameter and beta parameter and here the a values of you know a parameter and b parameter through which you can you know draw the line. So,

now what we can do here, so we will check the particular you know issue and then we actually you know draw the conclusion. So, what how we can obtain this one is very interesting.

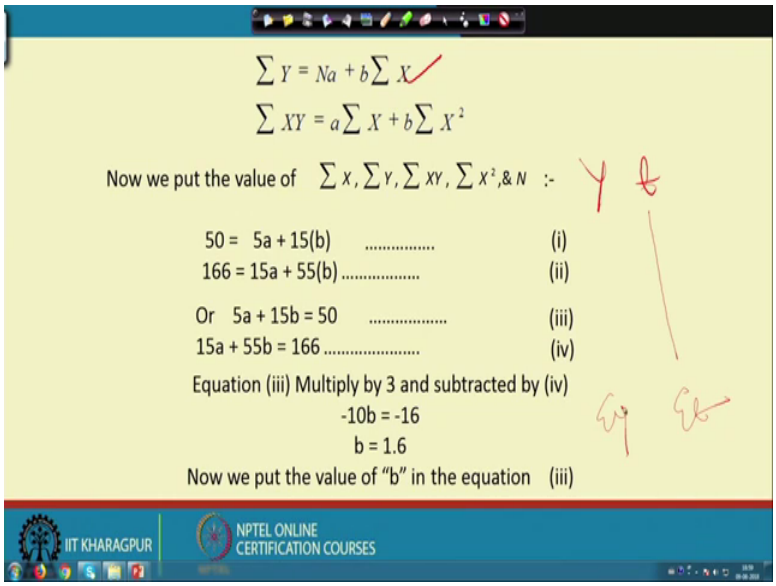
(Refer Slide Time: 06:43)



Same thing actually we just you know had a component simply you know let us say there will be error term so, that means, starting will be like this. So, let us you know the model will be Y equal to simply a plus b t and then error term and then what will you do here we minimize the error term and that to where Y hat equal to simply a plus b t and then Y minus Y hat equal to you know say error term that to Y minus a hat minus b hat t and a this is what the error term.

So, we will do the a minimization error sums sum of squares which is nothing but summation Y minus a cap minus b cap t and that 2 square. So, now you have to minimise a this error sum scale with respect to a cap and b cap like you know as usual simple regression modelling. And then after doing the simplification so, you will have actually only two different equations the 1st question will be like this summation Y equal to N a plus b summation X.

(Refer Slide Time: 07:58)

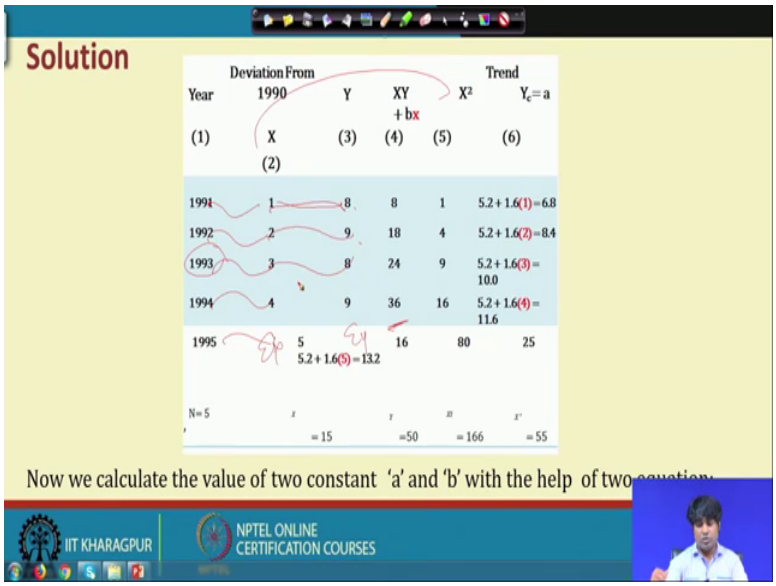


So, here X is nothing but called as you know t and Y is nothing but the original variables and by default here you know error sums some of square you know so for as a optimisation is concerned. So, you have to minimize error sum of square with respect to a, then if we simplify you are have the first equation that is and then we have the second equation when we actually differentiate with respect to b.

So, there is a extra component in b so that will be multiply and then you will have a 2nd equation, summation X Y equal a a summation X plus b summation X square. So that means, here spreadsheet will be having data like this Y, then t for simplicity point of view since data is very consistent over the time the first data point can be consider a that is with respect to time can be considered as a 1 and second time product can be 2 3 and so on up to you know last unit.

Then this will be treated as a X then the rest you can actually simplify. So, now we need actually summation X and summation Y that means, summation X into the summation X this is summation t, then we have a summation Y and then you can multiply you can have summation Y t and then we can have summation X square. And after the simplification we will have a actually these two equations and you know you put these value like this here. And then after simplification you will get the a value and b value which we can actually analyse more details here and that too with the help of this you know tables.

(Refer Slide Time: 09:57)

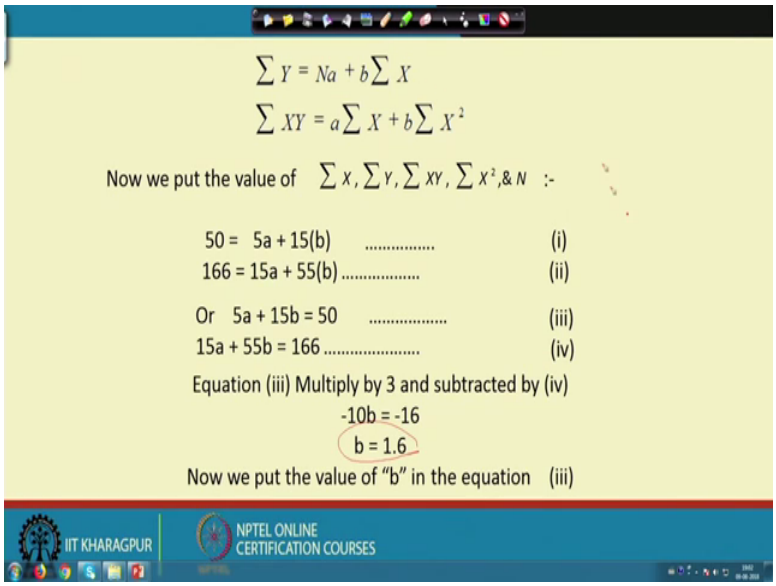


This is what the kind of you know table here and that means, a here the data you know it is big data we are just taking you know a 5 data points. And here Y is the in means this is here and this is actually a X and then we have to actually predict Y that to with respect to X. So that means, this year is transferred into you know X what I have already I told you here, for instance 1991 can be written as 1 1992 can be written as 2 then this is can be written as 3 then 4 and 5 like this whatever data we have. So, that means, the first data instead of writing 1991 so you simply write you know 1.

Then next time next data here will be 2 and so on up to large data point and then every times here actual Y will be a you know generated there. So, that is in that means for year 1991 data is there, so the data transformation will be instead of 1991 8 we can write 1 and the 8 then again instead of saying 1992 9 so, we can say 2 and 9. Then again in place of 1993 you can say 3 and 8 like this we have a different you know pull data and then finally you can do the processing.

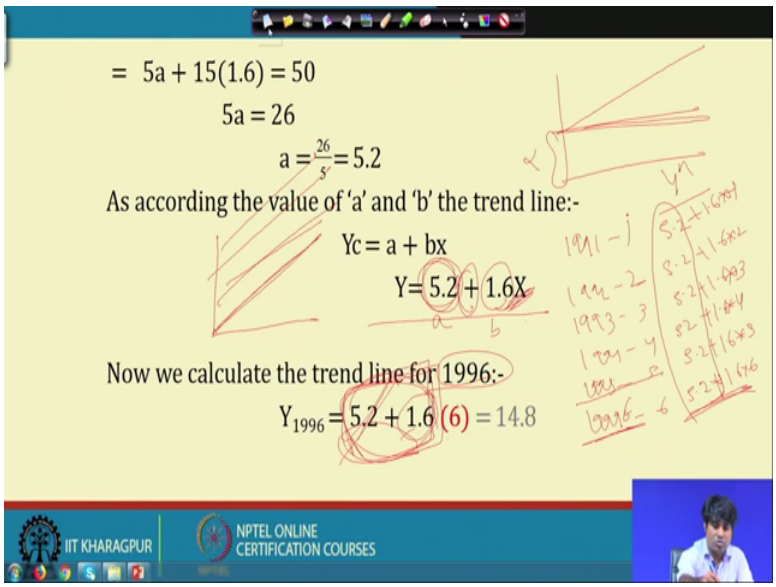
So, now, we will have here summation X then we will have summation Y here then again you can have X Y and that is what you know X Y and then X square and then after that we can actually get sum X sum Y summation X Y and summation X square. After having all these information then you can actually move to this you know equation and this equation will give you for instance now come to this you know equation here so this is the equation.

(Refer Slide Time: 11:57)



So, now summation Y so you see here summation Y is nothing but actually 50 and summation X is nothing but actually here a 5. So, likewise you can actually so, here actually summation X summation Y summation X Y and summation X square. So, now we put all these things and simplify after the you know after the process you will find the value of b equal to this much, then now put this value in a you know equation 1 you will get a value. So, ultimately the process will give you the value of you know a and value of b to do the you know trading that is what here actually.

(Refer Slide Time: 12:45)



So, now this is what the predicted line now 5.2 plus 1.6 X, so that means, technically 5.2 is the a and 1.6 is a b not necessarily every time the coefficient will be positive. But, the maximum times the coefficient will be positive and of course the data would be a you know I tell you the real message what is the exact you know structure. So that means, technically what is happening X is here actually used as you know time series a time series structure.

So, now you know say 1991 that we can say 1 1992 that becomes 2 and so on. So, you have up to 1995 data then when you look for you know let us say 1996, then that means, the here is a 1990. So, we have actually let us say we can fill up here 1993 then 1994, so this becomes 3 this becomes 4 this becomes 5 and this becomes 6. So, now after having this one so what will you do so, we can have a predicted lines and the here I just you know let us say 5.2 plus 1.6 into 1. Similarly, 5.2 plus 1.6 into 2 plus 5.2 plus 1.6 into 3, similarly 5.2 plus 1.6 into 4 and then 5.2 plus 1.6 into 5 and then finally 5.2 plus 1.6 into 6 that is what the a you know immediate predictions next year.

So, we have data up to this and then next year prediction will be 1996 will be this much, so that means if you look very carefully here you know this is very common actually. So, like you know we have intercept so you know the curve the structural will be like this actually finally, it will be like this 5.52 means this is the weight of you know intercept and then with respective you know time moment. So, the curve will maybe starting like this of, of course you know there is no need to draw this line. So, ultimately this can be movement will be like this also it may a means this will be also clearly indicates that you know there is a alpha since alpha is coming some positive value.

So that means, it will be above than origins if it is negative then you know below than origins, if alpha is not there that means intercept is not there then the prediction will be start from origin actually a this can be also true. Otherwise, it can be like this it can be like this it can be like this, obviously whether the movement will be up and down or you know negative. So, it is depends exclusively on you know a coefficients and that too it is derived through the data only. So, after getting this you know alpha you know a value, so we can get to know what is the position of this you know line starting you know beginning of that you know kind of you know structure.

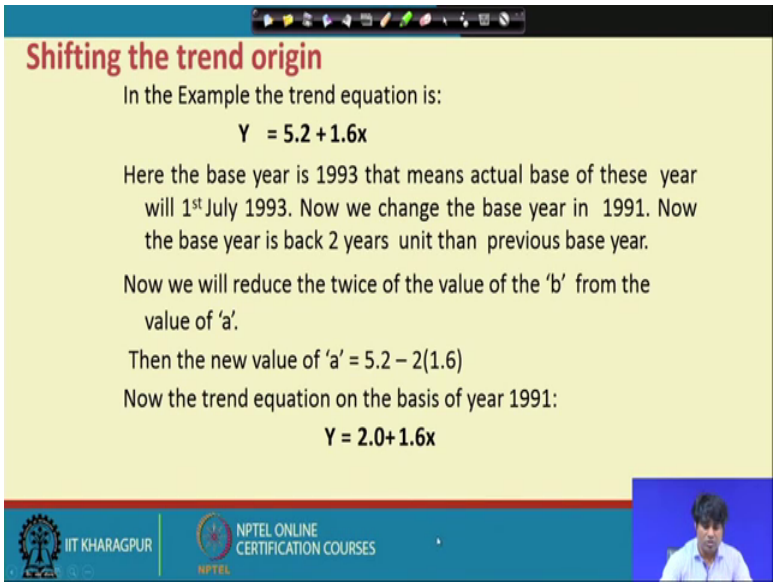
So the that means, the first and you know moment and then over the time the kind of you know predicted movement. So, now you just put you know 6 here then that will give you the prediction about 1996, then say we have already established this is actually common. So, the a you know intercept and the kind of you know the slope is the common, now you just you know replace here in place of 6 if you put 7 then the prediction will be for 1997. Again if you put you know 8 then the you know the prediction is for you know 1998.

So, that means, now having the past data or you know using the past data we have actually you know build a kind of you know trend, where the a and you know you know the b this slope coefficients are you know more or less constants that is what the average. Actually that means, that is the you know minimum which we would be carry through which you know do the kind of prediction.

That is what the basic understanding of the this you know trend lines and the kind of a predicted line and the kind of you know average lines and that too you know it clearly see it means it clearly highlights that in every time. So, time series is a book big boost to how you can actually easily get the you know trend lines or the predicted lines, means simply without actually doing like this you know just enter the data in excel and you know scroll it then you will find a you know a generated data. And that is actually you know excel software by default will read the kind of you know trend, understand the particular structure then you know create a you know future kind of you know requirement.

That is how the beauty of this you know trend you know least square mechanism and that too by using actually linear equation only. But not necessarily we can every time use linear equations that depends you know exclusively the kind of you know data 0n the basis of you know diagrams.

(Refer Slide Time: 18:09)



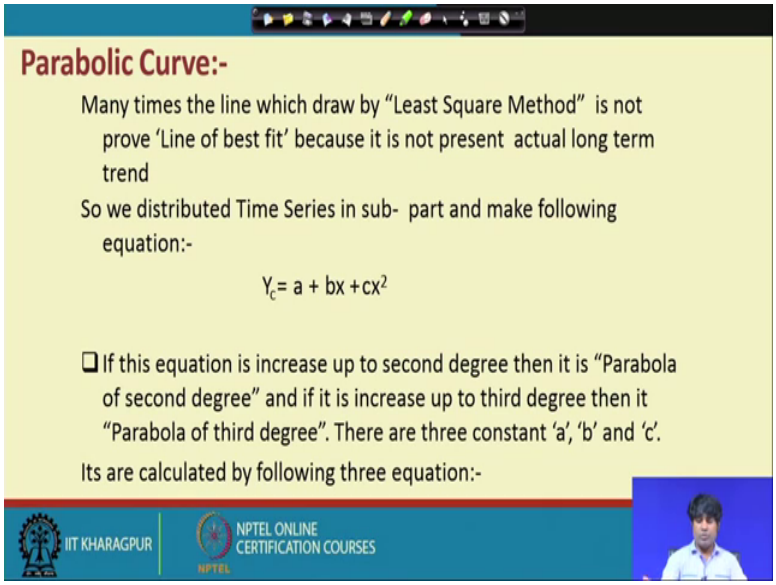
So, that means we should first you know go for you know graphical visualisation and get to know whether repeat will be with respectively linear ones or with respect to you know non-linear one. So, now, after simplification so this is what the a final predicted line, so Y equal to 20.0 plus 1.6 X and that too here actually the kind of you know predicted equations.

So, what we can do actually you know we can actually check the difference and the actual and the kind of you know a predicted. Of course, the predicted it is derived from the least square and which is actually derived through a kind of you know process what we called as you know OLS mechanism. And now having the same data you go for you know freehand curve semi average moving average and then again to you know least square mechanism and every times to check the reliability of the model.

Of course, you can you know in this particularly least square mechanisms you have to see that you know the statistical significant goodness fit is ok, all kinds of you know diagnostic check then by the a you know end of the day. So, you will have actually forecasted line and the actual line, so that means a you know starting with you know freehand curve then semi average moving average and least square every times. So, you have two different you know you know positions one is the actual position and the predicted positions. Of course, actual position is a very much common for all these you know mechanisms the predicted position is it slightly deviating.

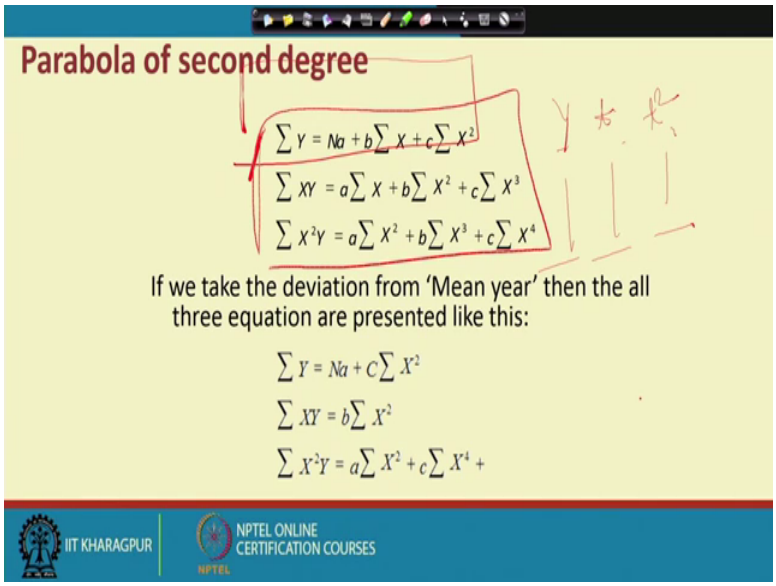
Maybe slightly deviating because, you know we are using different mechanism and different process through which we will get this you know line. But, most of the instances you know keeping other things you know remain constant more or less it would be actually better way of you know predicting the structure and that too as per the particular you know engineering requirement.

(Refer Slide Time: 20:23)



So, now so not necessarily that you know the kind of you know trend line always in a linear structure, if not then you can start with you know parabolitic parabolic kind of you know structure that is what called as you know linear kind of you know a structure through which we can do the predictions. But of course, here this is a x becomes you know t here so now instead of using Y equal to a plus b x. So, now we are using Y equal to a plus b x plus c x square, so that means some t will be there Y will be there. So, now we are getting you know additional one parameter in that with the help of you know dot product only.

(Refer Slide Time: 21:07)



So, now with you know two variable case the structural form will be like this a here that is what the kind of you know games. So, this is two variables so we have 2 structural equation, of course which is derived from the OLS mechanism. And if you go through actually the kind of you know 3 variable case and this is how the you know kind of you know structural equations through which we can do the predictions. So, now what is happening here so we have a Y and then t and then you know t square.

So, now you can start the game so some Y some t some t square then some Y t some Y t squares the X square. So, that is how you will calculate and then finally you simplify and once you simplify. So, now in the previous case that too for you know bivariate and that too for linear structure we have only a and b, so now we have a b c.

So, alternately the equation will be a plus b X plus you know c X squares not necessarily positive, so data will tell you whether the coefficient is positive or negative and where your data will tell you what is a value b value and c value. Once you get to know what is a value b value c value and it is nature, when we can actually simply you know means very easy way you can you know predict the kind of you know requirement. That is what actually basic structure of you know prediction the with the help of you know least square mechanisms.

So that means, least square is one of the standard mechanism through which actually a you know we can derive a trend lines or predicted lines or the forecasted lines, through

which actually we can do some kind of you know engineering predictions or engineering forecasting. So that means, technically after knowing or after understanding the time series data that too you know day wise structure, week wise structure, monthly structure, quarterly structure, annually structure, decadal structure. So, means these are all various forms of you know times series data and when we start with you know day wise then week wise then month monthly quarterly and annually semi annually.

So, these are all various levels actually means a it means what we can called as you know time series reporting and when you start with you know that means, let us say have you know data with respect to day, then data with respect to week data with respect to say month data with respect to quarterly and data with respect to annually. Of course you know the data with respect to you know day have a huge a sample.

Then this sample will be slightly low and instead of you k now day wise data we may have weekly data. Again this sample will be very small instead of week wise data, if we use monthly data, again the sample size will be very low and instead of using monthly data if you use annual data because for a annual data. So, we would like to follow you know moving average concept like you know 12 years 12 data plans to become a you know single and so that means when you will do some kind of you know normalisations like you know moving vaporization something like that.

So, you should actually logically connect every time, so they you know once you know logically within establish a this can be done. So, when there is you know such a big error so you can proceed, of course error may be there in the system which we can actually test later stage, once you get the predicted line and try to compare and contrast with you know actual line actual line and then check the difference.

Whether the difference is actually coming very very minimum, so if that is the case then by default you can as you can you know declare that you know this is the predicted line through which you can do the predictions and that too with the help of a you know time series data. So, we have you know means a first and we have a time series data and there are various mechanisms and now using this time series data on the different mechanisms we can have a different kind of you know predictions and different kind of you know forecasting.

So that means, by default it gives some kind of you know robustness here the same data you know with respect to same time period. So, we are you know having different kind of you know prediction structure, then you know graphically means all this possible through a graphically just you know have the predicted structure and you know draw the predicted lines. So, now ultimately a means we have already discussed various mechanism through which you can you know declare which one is the best and by the way you know starting with you know freehand curve to least square mechanism.

These mathematical you know require you know we are improving actually, so that means in the first case freehand curves that is you know that mathematics, say simply you just check the highest point lowest point. And blindly just draw a line in between it will not cross you know or perhaps you know peak and you know it is not you know again across the peak at the bottom side, so it will be parallely in between.

So, if that is the case you know you cannot do the kind of you know a better prediction. So, all this things are you know very to do the forecasting and you can have written check and compare. So, one way of you know robustness check to justify that you know the model is very good or you know can be used for any kind of you know engineering forecasting is that you know. You check you know means draw the line with respect to actual then you slightly inspect that there is a which one is the much better, that is you know called as you know graphical visualization, but parallely you can do some kind of you know quantification by calculating the a forecasting error.

Which is actually difference between actual Y and predicted Y and usually a the sum of you know actual sum of the difference between actual and predicted lines should be equal to 0. If that is the case in the first instance you can say that you know the particular predicted line will be the best for the forecasting or the kind of you know prediction, if not what will we do you have to find out you know forecasting error actually. So, then after that you will you will compare with the all these you know methods differently, then you check where the forecasting error will be actually minimum.

If it is 0 then by the default that will be the best if not 0, then how much error is coming and how you know which methods the error is actually very minimum and that will be by default will be used as a kind of you know engineering predictions or you know kind of you know engineering forecasting. We have lots of you know forecasting tools

through which you can declare that you know the particular model is very good to do the predictions, like you know mean error mean percentage error or mean absolute percentage error or root mean square error you know like.

We have a you know various you know indicators through which you can declare that you know this model is actually good for the prediction and forecasting. Through which in the detail about this you know indicator through which you can check which one is the best model for the prediction and forecasting, we will discuss the details in the next lecture and a in the meantime which we will stop here.

Thank you very much have a nice day.