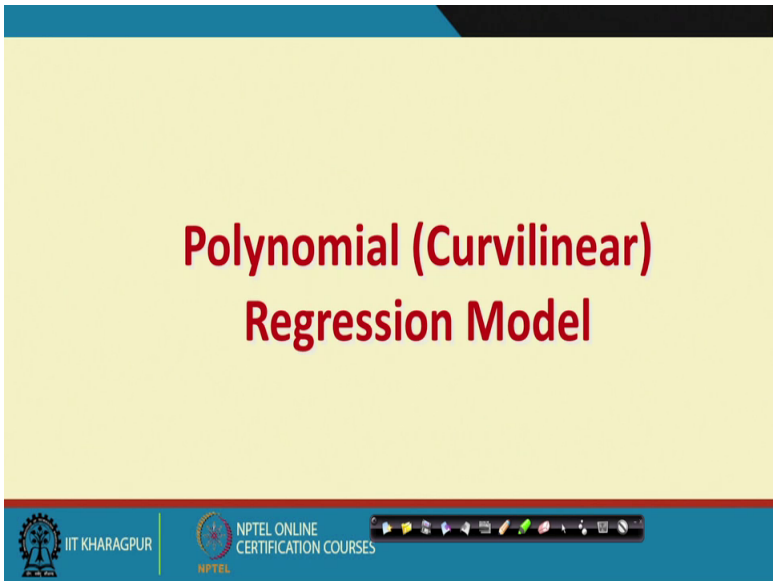
Engineering Econometrics
Prof. Rudra P. Pradhan
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

Lecture - 35
Non-Linear Regression Modelling- Polynomial (Curvilinear) Regression Model

Hello everybody. This is Rudra Pradhan here. Welcome to Engineering Econometrics and that to the discussion is on Non-Linear Regression Modelling and in the last couple of lectures, we have discussed this particular you know non-linearity and that too with the help of dummy and with the help of you know interactive effect. And, in these particular lectures, we specifically highlighted the non-linearity and that too with respect to various you know functional forms, not the kind of you know interactive effect and that to either the dot product or with the help of you know cross product.

(Refer Slide Time: 01:05)

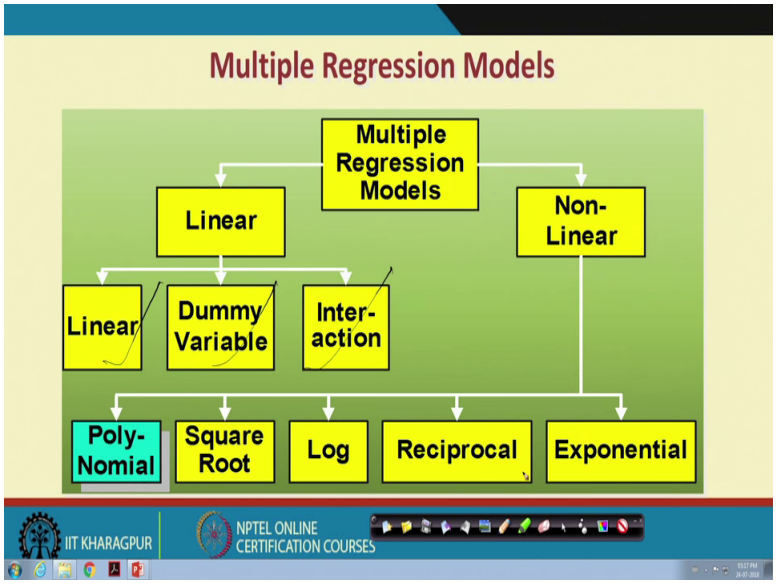


So, we will see here (Refer time: 01:00) what are the ways you can actually highlight this non-linearity with the different kind of inner functional forms and technically we can call as you know simply Polynomial Regression Models.

So, what is, what are these you know polynomial regression models, so that means if it is not linear and one way if not against interactive, it is a curvilinear or it is simply called as you know polynomial. So, the other form of you know non-linear regression modelling is nothing, but called as polynomial regression modeling or curvilinear regression

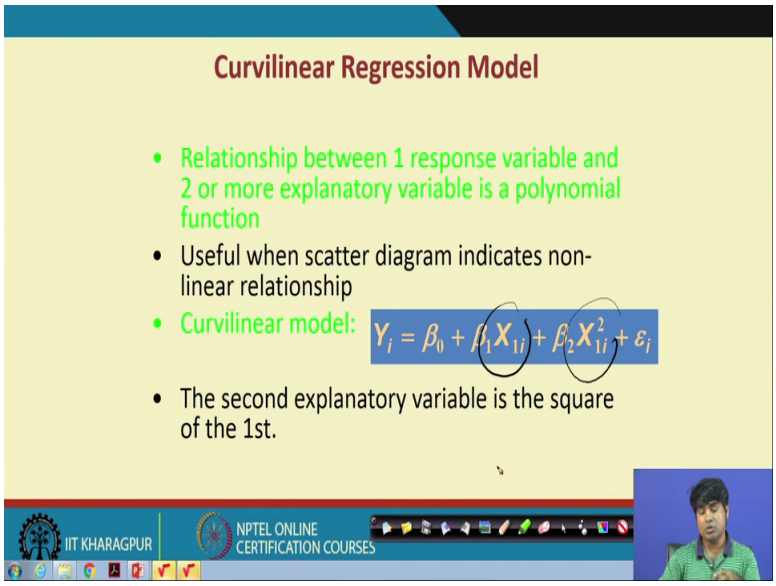
modelling. So, what are the ways you can represent the polynomial regression models or non-linear regression models and that too you know the issue of you know curvilinear and again we start with you know simple structures.

(Refer Slide Time: 01:48)



Till now in this case we have already discussed about these linear versions, dummy variable versions and the interaction versions (Refer time: 02:00) and now, we will start with actually polynomial, then we will go through other routes through which you can have different kind of you know non-linear functions and again the kind of you know non-linear regression modelling. So, let us start with the first kind of you know polynomial one and how the polynomial model can be built here to represent the situation, ok.

(Refer Slide Time: 02:25)



So, one way of you know the polynomial representation is here against between Y and X, where Y is the dependent variable and X is independent variable and here we are actually X 1 you know called as you know kind of an independent variable and then, you know we have created X 1 square. So, that is the quadratic forms which you are putting here. That is the polynomial and the degree of you know linear you know polynomial, it is 2 here. So, X 1 and X 1 squares, ok.

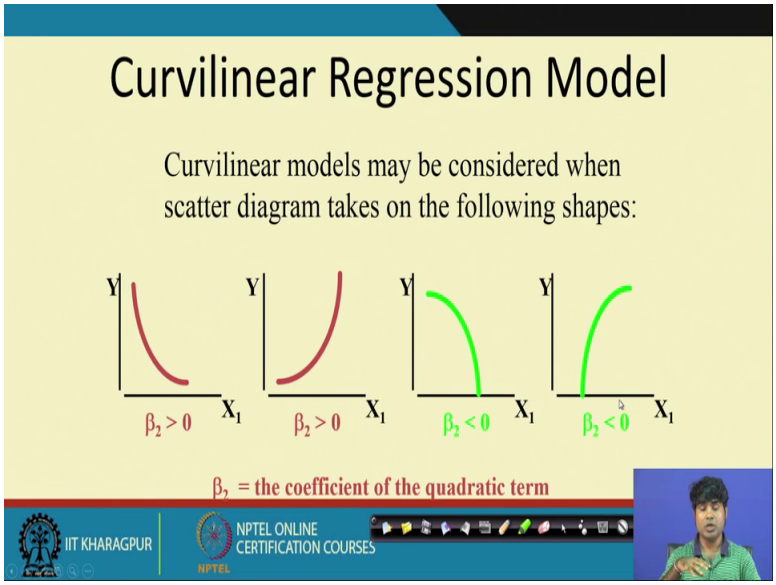
So, then by default we have 3 parameters and the model becomes you know non-linear ones and it is called as you know curvilinear regression modeling, ok. So, that means technically we start with Y and X 1. That means, you have data on Y and you have data on X 1 and in between you created a concept you know X 1 squares, but actually whether the particular you know data is fitted with you know non-linear function technically that too in a form of you know quadratic.

There are couple of you know sub-test which can directly actually give the functional forms. Particular functional form can be used corresponding to the particular you know data set. Of course, software cannot read the theory software and sometimes can understand the data and you know regress the particularly you know the equations and you know they connect kind of you know relationship with the means as per the requirement whether the linear requirement or you know non-linear requirement.

By the way we have already discussed you know models, these specification and where your actually data remains with respect to data. The model requirement is actually linear 1, but unfortunately you have used actually non-linear models and you are estimating the, establishing the relationship between these two with a kind of non-linear set up. In that case, there is a high chance you will find you know bias and that bias against we called as you know these specifications. So, that is why choosing the correct functional form or choosing the correct type of you know models, we will bring you know as efficiency or you know kind of you know good reliability and it will be by default free from all kind of you know diagnostics.

So, that is why we are actually here to know this particular concept and then, use as per the particularly you know requirement. So, this is what the, this is what actually you know curvilinear regression modeling and that too it is a kind of you know non-linear regression modeling where we use a polynomial functions that too the degree of polynomial is it too here. And, we can also extend with you know having the degree of polynomial 3 4 and so on, but before you go for you know such kind of you know polynomial functions and the kind of non-linear regression modeling, you should ensure that you know this model is fitting in this particular you know domain and after actually doing this, of course really go through the reliability check. If the functional form is wrong, by default reliability check will get affected.

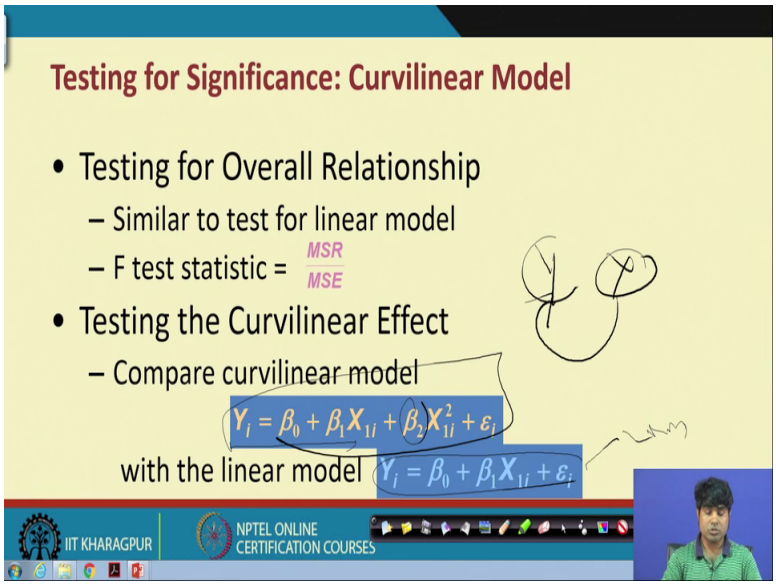
(Refer Slide Time: 05:56)



So, what is actually more important here is that you know we have here variety you know of you know kind of you know modeling that too by using various you know non-linearity structures and that too with respect to various functional forms. And, that itself will bring different kind of you know flexibility and different kind of you know flows through which you can actually establish the relationship very easily and you know very efficiently.

So, these are all various you know forms of you know curvilinear. So, in it may be increasing shapes or you know decreasing shapes which say kind of you know convict type, a convex type or you know it is kind of you know concave type. But, ultimately whatever may be the kind of you know structures or data will representing accordingly and first you can actually visualize and as per the visualization, you can get to know what kind of you know functional form you can use to estimate the models and then, to establish the relationship between the dependent variable and the kind of you know independent variables.

(Refer Slide Time: 07:03)



So, as far as a testing is concerned that too in a curvilinear situations, it is a little bit you know different because sometimes we may use you know different kind of you know functional forms like you know log functions, exponential functions and then, but the structure is more or less same. We will go through the specification test, then the goodness fit test and the kind of you know diagnostics.

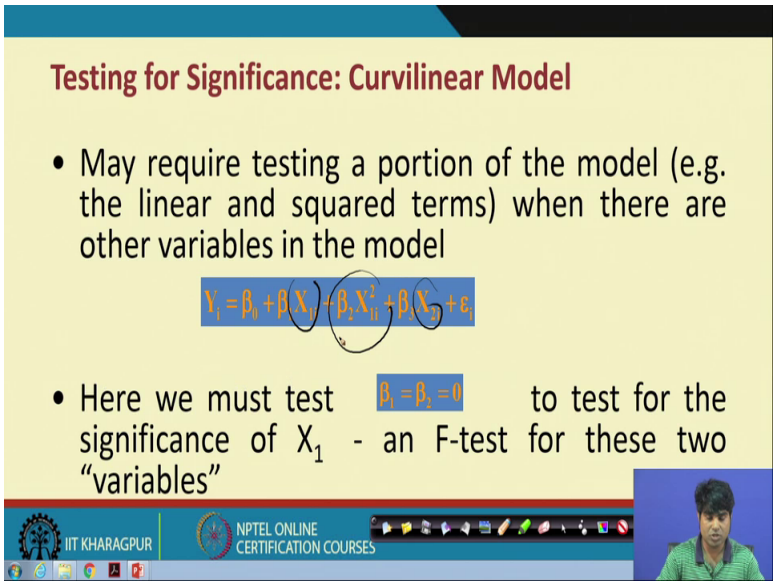
So, sometimes you know the model mis-specification is there in a particular you know functional form. So, then model diagnostic will get affected, then it will change the kind of you know functional form, then by default model reliability or model efficiency we will start you know increasing and that itself will give you better judgment and that too in the decision making process. So, comparing actually a linear with you know and non-linear models say here, there is a difference.

So, if we see these two models, that means technically we have here actually two variables. So, that too the game between Y and X and then, here the issue is you know whether X is influencing the Y and for that actually we have two different models. One model specification, it will be like this and that is the curvilinear regression modeling and another model will be like this that is called as you know linear regression modeling.

So, now if we will compare these two models, then the coefficient beta 2 is the instrumentals like we you know the case which you have discussed in the in the context of you know interactive effect. So, now if bit of actually the model you know adequate model requirement is you know be X 1 square and by default beta 2 will not equal to 0, so in that case beta 2 will be non-zero and it may be positive, it may be negative. That is not the issue because it is the data will or the kind of you know estimation process will give you the is size of the beta or the nature of the beta, but whatever may be the reasons. So, most of the instance, it will not actually be 0. So, it will give some value either in a positive or negative. Whether it is a positive negative, that will make the difference between the linear and non-linear like you know the case between you know interactive effect where we have an issue between yes and no.

So, if it is yes, then this particularly component will be added. If no, then the particular component will be out in the system and sometimes it is not a question of adding it, may be also subtracted depending upon the nature of the coefficient. If it is actually plus, then this will be added to the intercept and in fact will be high and if it is a negative interest you know coefficient, then that will be actually subtracted from the intercept value and as a result there is a decline. So, that is how the kind of you know differential refined while comparing the two different groups or you know two different structures, right. Here the same models we are actually presenting with the two different you know functional forms and then, checking the reliability of this model.

(Refer Slide Time: 10:20)



This is what actually you know kind of new models, where we are using actually two independent variables and that to what we can say here X 1 and X 2 and again we are allowing X 1 squares. Sometimes you know if that particular variables you know as per the data, the fitting is showing actually cortex.

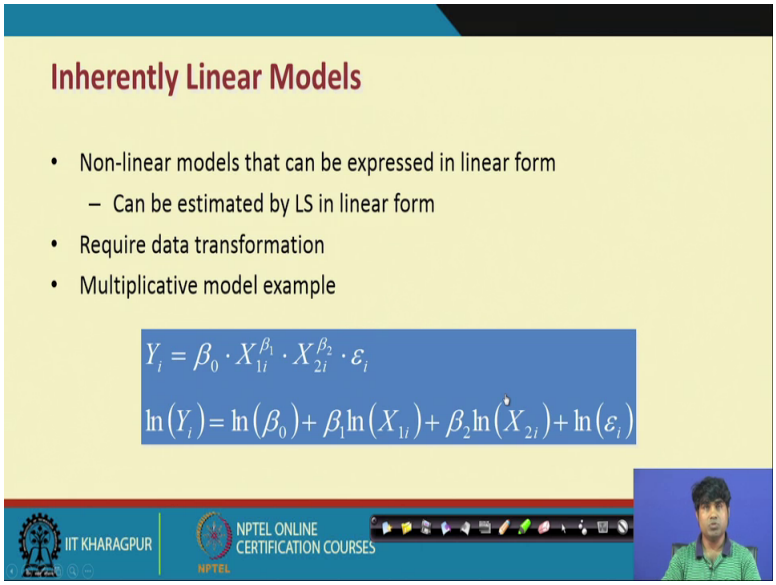
So, you have to represent you know you know X 1 and X 1 square together, otherwise the model by default will get mis-specify mis-specified. So, if you know in order to avoid model mis-specification, it is better to actually you know use this model and then, check the reliability. So, that means technically if against if X 1 square is not required, X 2 is not required and by default the model will be restricted to y equal to beta 0 plus beta 1 X 1 which is you know actually you know linear 1.

So, in fact actually whether the functional form is correct or not, you can check the model specification test and a again first start estimating and then, check the reference whether the whether there is a difference. If actually there is a difference, then the particular coefficient cannot be equal to 0 if the particular you know very you know variables in part is there.

So, the coefficients will be non-zero and it will give you the kind of you know weight. The weight may be you know negative one or the weight may be the positive one, but ultimately whether it will be positive or negative whether it means high increase or high decrease, so that depends upon the data and the kind of you know estimation process. So,

ultimately this is another form of you know model through which actually you understand the non-linear regression modeling.

(Refer Slide Time: 12:15)



Again another form of in a linear regression modeling, this actually in business scenario is called a c cd type you know function form COB Douglas production functions, where we have actually Y dependent variable and the other independent variables are exponentially connected. So, then what will we do? We can you know we cannot directly apply where this to this particular you know models. So, we can go for you know dark or you know model transformation. So, just you know transfer the model with the help of you know log and then, that means it becomes a log transformation, then the variables can be removed from the non-linear to linear one and just you know apply log, but beside and simplify, then by default you will get you know this model and this model becomes actually the kind of you know linear once.

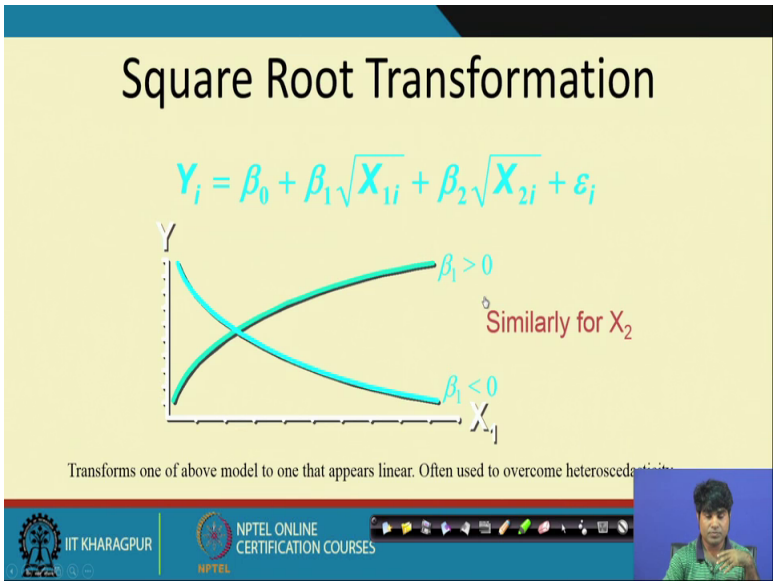
So, that means technically you have actually non-linear structure check and then, it can check if not feasible. So, you know requirement is not there, then weight of that particular you know inclusions will give you the message. If that is the case, then it will add value to the systems. If not, it will just you know give the 0 value as a result, that particular item will be removed in the estimation process. That is the kind of you know game which you can have here.

(Refer Slide Time: 13:51)

So, here the issue is actually or the requirement is you know transformations. So, that means technically whatever may be the form of you know model representation or the kind of you know model requirement, we can do the transformations and then, the degree of non-linearity can reduce the linearity. Sometimes you know you start with the linearity and if not fit, then you go for you know non-linear functions and again you can start with in non-linear functions, then you end up on the process of you know testing and the estimations. Ultimately it will change the form. That means, finally you need a kind of you know optimum scenario whether you start with you know linear to non-linear or non-linear to linear.

Ultimately, we need your models which will represent the immensely which will establish the relationship between the dependent variable and independent variable very perfectly, very efficiently and that will, it will be actually well fitted with the data and you know theory. That is what actual requirement and for that we are getting to know all these you know process you know linear to non-linear and again non-linear to linear depending upon the kind of you know requirement.

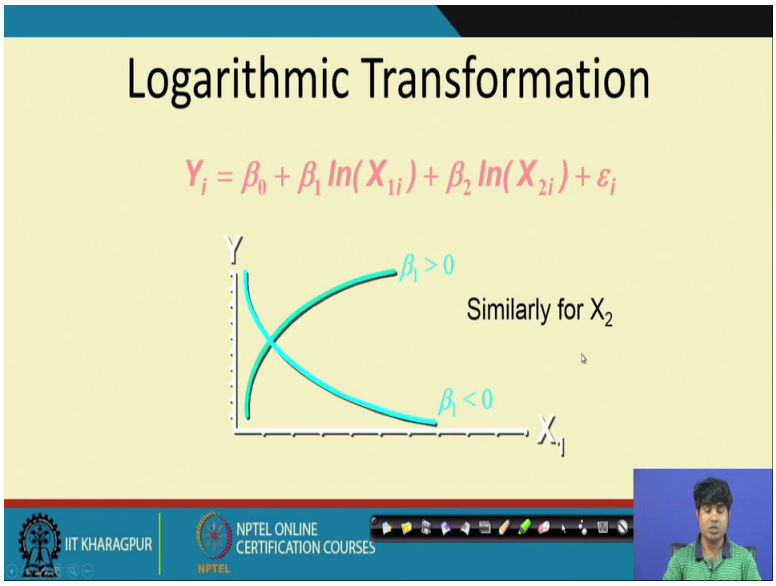
(Refer Slide Time: 15:09)



So, this is another form of you know a non-linear functions. We have the same variables means we are just connecting Y with you know independent variables say one point of time say only X once, then X1 or X2 or X1 X2 X3 and so on. Sometimes we are bringing interactive effect, sometimes you know going for simple log transformation simple.

And then, again you will standardize the particular you know variables and here instead of you know directly using X1, you can use actually square root of X1 or square root of X2, then that becomes you know brings the non-linear structure. It is not that every time you list using this square of that particular variable. You can go for in a square root of that particular variables. By default the volatility only get affected and the model relatively get actually improved.

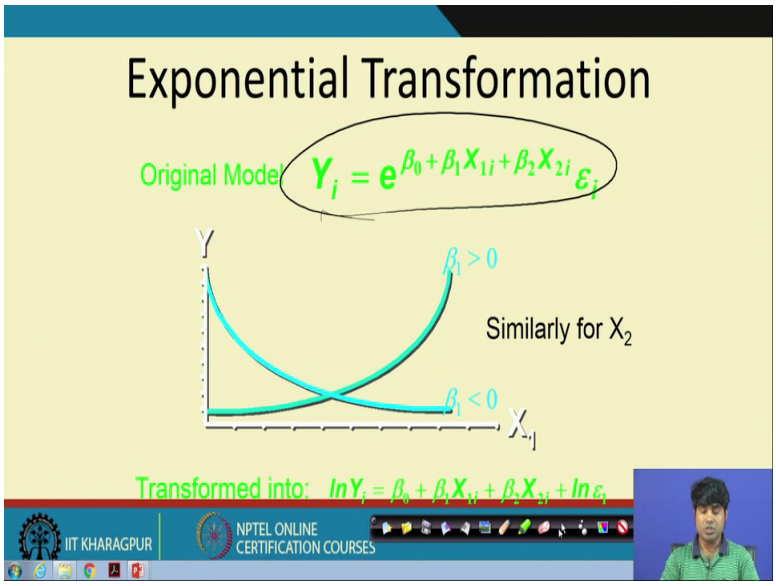
(Refer Slide Time: 16:06)



So, these are the ways actually you can again and see whether the particularly you know modeling process really bring some kind of you know reliable output and that too as per the requirement of you know solving some of the engineering problems. So, again instead of you know square root of X1 and square root of X2, now we will go for you know simple log X1 and log X2.

So, this is simply called as you know a single log functions and then, we like to just regress actually Y with the log X1 and log X2, then we check whether the model is really fitted to this data and the theory. If that is the case, then you can go ahead with this particular model. If not, then again change the functional form and check the model reliability. I am very sure in one case it really it will give you some kind of you know message that yes this is the model through which you can actually proceed and do the forecasting and do the my decision making as per the requirement of any engineering problem.

(Refer Slide Time: 17:02)

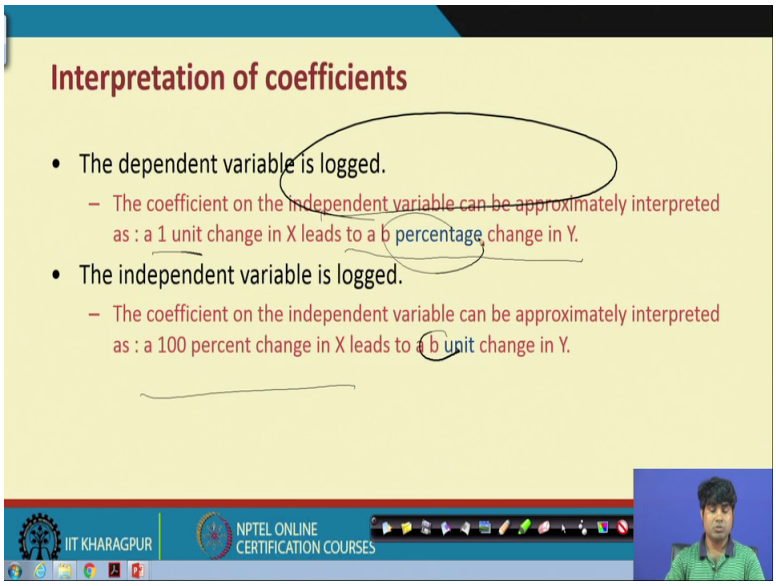


So, another form of you know and non-linear structure is the exponential transformation. So, here the model as such actually behave like this. It is a very complicated Y i e to the power you know all these parameters starting with the beta 0 beta 1 and beta 2. So, we cannot actually directly estimate this kind of you know models. Again we will go for you know transformation and simple transformation can work here. Just apply the log in both the sides, then instead of Y, we will use log Y and then, by default it will be beta 0 beta 1 x 1 and beta 2 x 2.

So, that means if this is the model or this is you know functional form and that too linear, a linear regression modeling is these kind of you know structure. So, what he is supposed to do actually first you go for you know Y transformation that to bring the log Y and then, as usual X1 and X2 will be there. So, then we started regressing and by default the estimated output will follow these particular functions and that is the output of you know simply a non-linear regression modeling.

What is actually more important or what is a significant requirement? So, you need to check the parameters and the models that too with the help of a safety test and diagnostic test and the kind of you know sample prediction test and finally, the declaration is that you know model is absolutely as per the particular you know requirement and the kind of you know management decision making process. So, ultimately this is another way to represent the non-linear regression modeling.

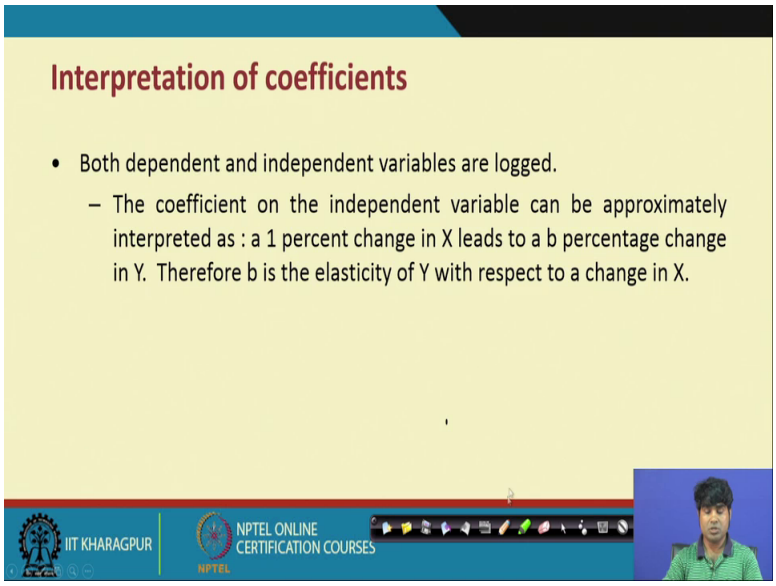
(Refer Slide Time: 18:54)



So, here the interpretation is actually slightly different you know because it is a log transformation. So, it is the simple representation. Here is the percentage change of you know X to Y and as a simple 1 unit of change in X is to percentage change in Y. So, that is the beauty of this you know exponential type of you know models that will represent the non-linearity structure and the independent variable is the logged. Then, you know a 100 percent change in external is to be a unit change in Y. So, you see the difference yes.

So, if the dependent variable is logged one and the others are actually a normal than the interpretations. 1 unit change of the X1 lead to 1 unit change of X1 leads to the kind of you know a percentage change in Y. Then, as an independent the independent variable is a logged 1, then 100 percent change in X that means, there 1 percentage change, here 100 percent change in X raised to be unit change in Y. So, that means you see here the percentage terms and here b unit and again here 1 unit and then, b percentage change in Y. So, that is the difference between the case. We are dependent variable is logged one and in comparison with the independent variable, they logged one; so, that is the difference we will find here and actually you know do the kind of inner requirement.

(Refer Slide Time: 20:52)



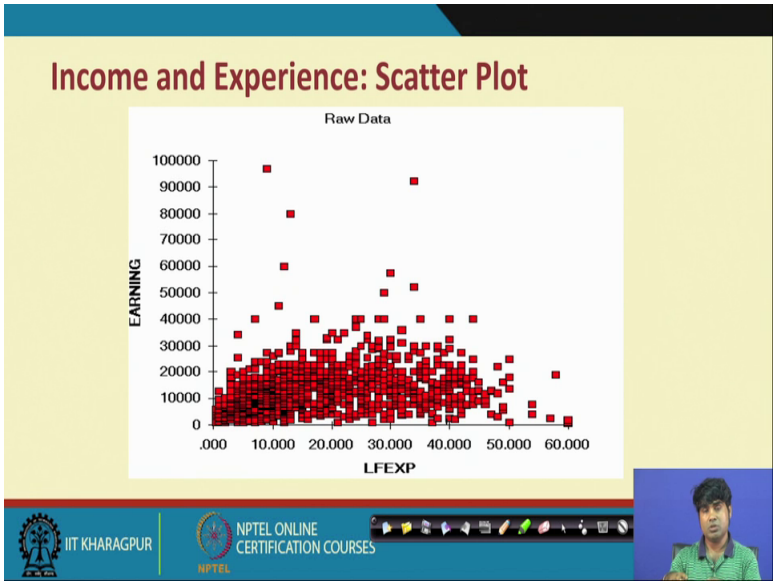
So, another way of you know interpreting coefficients that you know both dependent variable and independent variables are logged. So, that means it is actually a single log. Well, you know dependent variable log independent variables are not logged normal and again independent variables are log 1 and a dependent variable is the normal and against what is happening here the 3rd case a, both the variables dependent as well as independent variable will be the log transformation. So, that means we are just establishing the relationship between Y and X.

The problem is the same, but here what we are doing actually we are checking the robustness and with the intention that you know will being you know kind of inefficient models or you know efficient structure through which you can actually do the prediction very effectively and commit your decision making process as per the particular you know engineering problems requirement. So, ultimately in the case of you know in the 3rd case where we use actually log in on the both the sides, in that case the coefficient on the independent variable can be approximately interpreted as you know 1 percentage change in X rate to a b percentage change in Y.

Therefore, b is the elasticity of Y with respect to change in X elasticity is nothing, but actually percentage change of you know dependent variable to independent variables. So, now with the help of you know logarithmic transformation, the particular change will be well established and you know interpret within a different ways and that itself will give

you lots of you know flexibility to understand and to apply in that too as per the particular unit requirement.

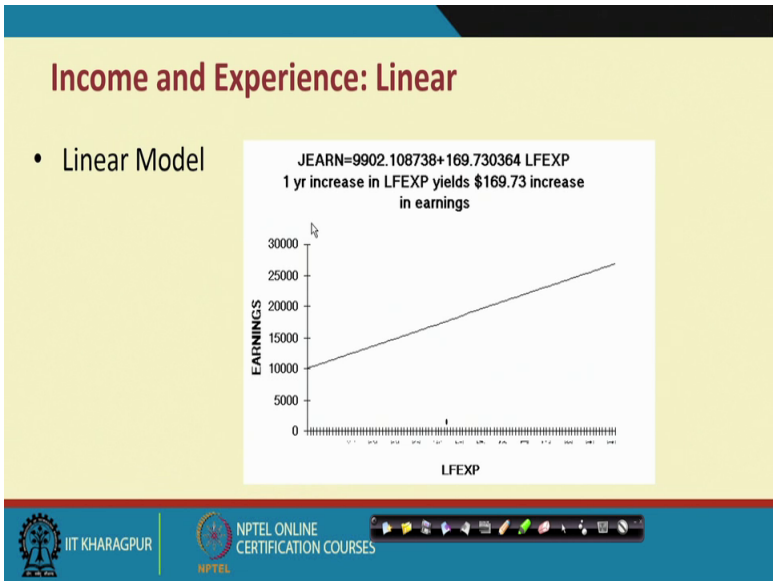
(Refer Slide Time: 22:50)



So, ultimately we have various functional forms and this is another kind of no plotting means actually use a particular functional form for the model building, the model estimation. So, this simple summation is your simple suggestion is that you know you just check the functional contrast.

If that is ok, then you can go ahead then finally of course finally it will be passed through all the diagnostics checks, but try you know it is better always to go through visualization process and get to know the function, actual functional form and then, fit the data or estimate the model with that particular functional form, so that the model output or model signal will be very effective to strengthen these systems as per the requirement. So, these are all various you know means, technically it is a income experience case.

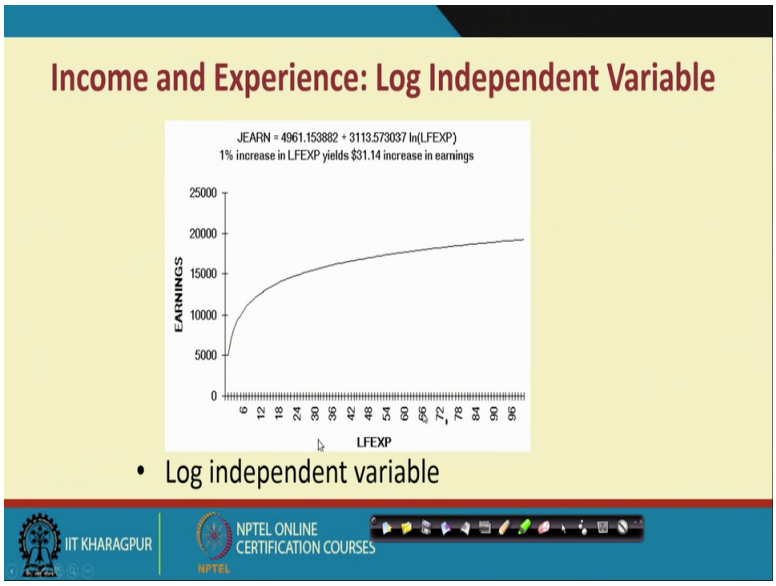
(Refer Slide Time: 23:41)



It means here the game is that you know if you are more and more experienced, then your income level will start increasing, but it is not always the case having you know more and more experience. So, there is no not means. It is not necessarily that flow will be increasing in a kind of linear structure. It may or it may not. So, with the help of the data, you can just check whether the particular flow is actually linear one or non-linear one.

Again if it is a non-linear one, what kind you know shape is available. That means, the particular change happening between you know let us say investment in actually or you know increasing the level of experience and the kind of you know impact on you know income. So, these are the things which you must be very careful before you do the processing and then, go as per the particularly you know modeling requirement.

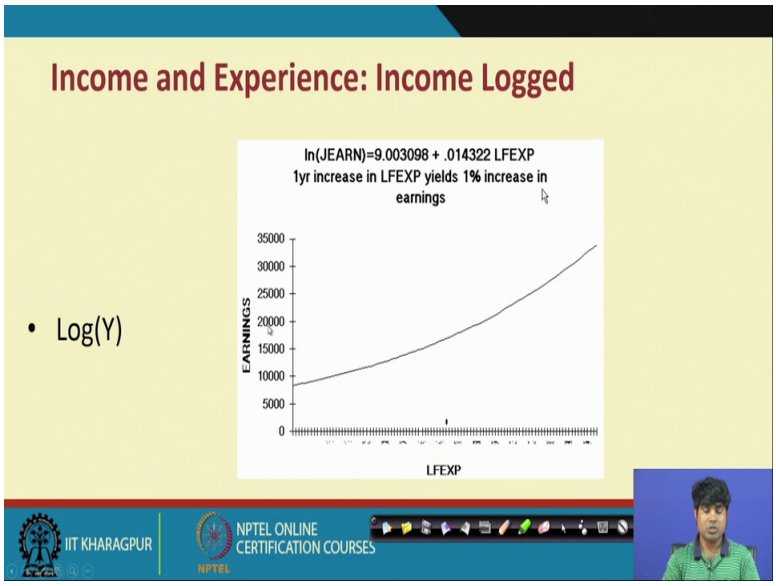
(Refer Slide Time: 24:41)



So, this is another kind of you know you know flow where we are going for you know log independent variable. That means, just to you know check here the kind of you know movement of this curve to establish the relationship between two variables dependent and independent. So, here the plot linear one just moving actually very straight; it may not be back you know perfectly like you know 45 degree angle, but this is the change, but the change is actually very linear.

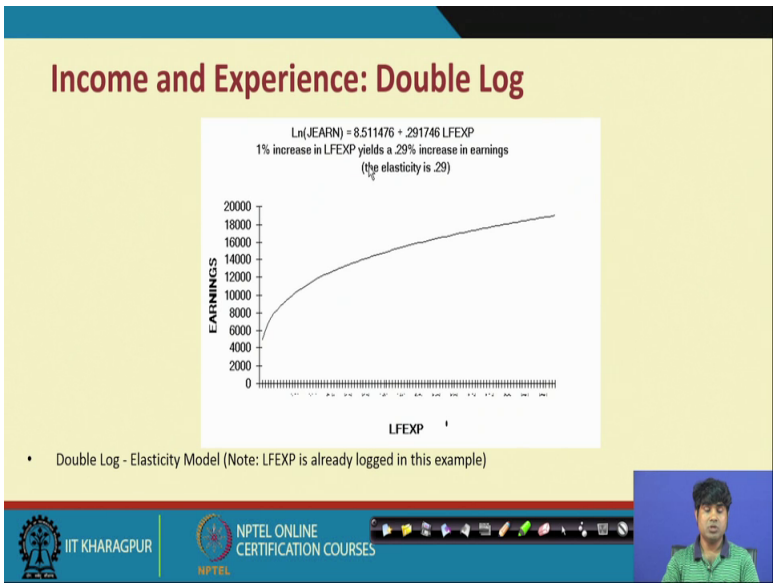
If not, then if it change the functional form, if that is this change and this collage you know, then by default the modeling process or the estimation process would be a log in the log you know transformation that with the independent variables against you can go for you know a log dependent variable if just it is opposite, ok. So, if you check the previous one, this is the flow like this and here the flow is it in slightly like this, and slightly like this.

(Refer Slide Time: 25:46)



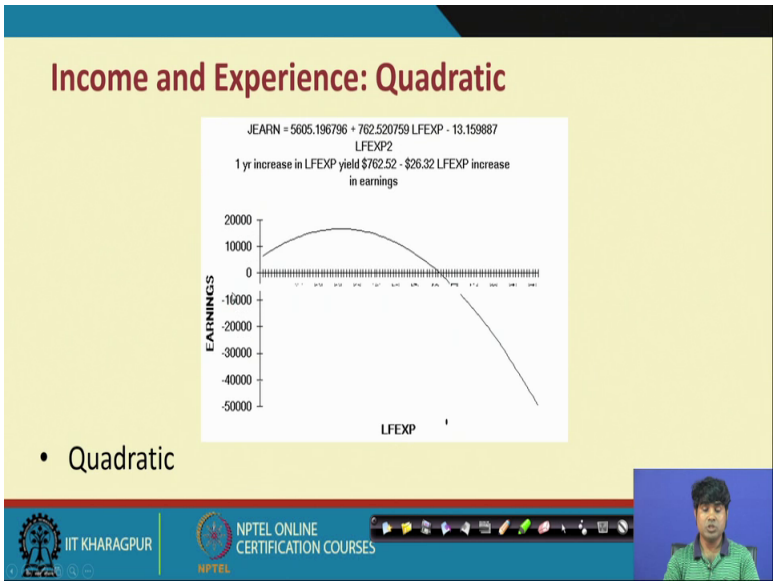
So, it is little bit you know yes you know slightly convex type of things. So, here we take actually a lot transformation that too in the dependent variables and then, connect with the independent variable where the game is between an experience and you know earnings against, this is the double log case.

(Refer Slide Time: 26:09)



If the model is slightly like this, then you can actually use double log functions than to bring the relationship between income and experience.

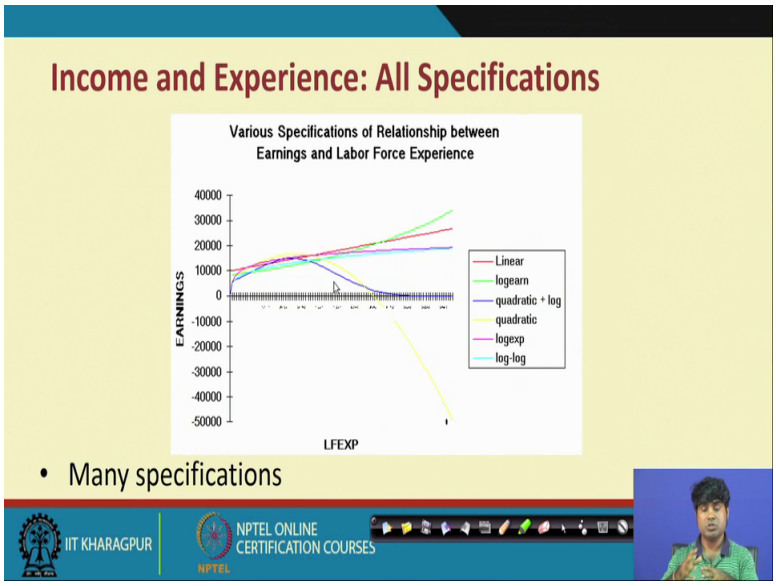
(Refer Slide Time: 26:17)



Why do we actually we are trying to establish the relationship between these two variables in command experience with the intention that you know highly experienced, highest income; low the experience, low the income. So, it is a direct professionally related, but the thing is that you know whether it is a particular merit professional relationship is a very in a kind of you know linear structures or some kind of ups and downs are there.

So, I mean it is maybe core type maybe like you know the kind of in a single log structure, double log structures. So, whatever may be the kind of answer, no harm to you know run the model with you know differ changing the different functional opponent and then, finally a picture which is actually have a better kind of an indication in the decision making process. So, that is the things are there so many ways.

(Refer Slide Time: 27:09)

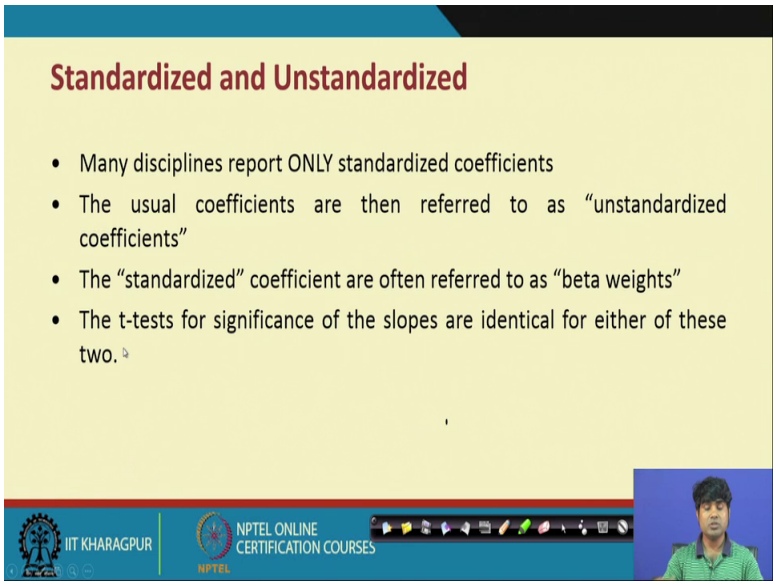


This is actually combined you know I mean see in this figure. If you look here, the same variables you know in the relationship between you know earnings and experience and same size of the data. So, we are just you know adding one after another with you know different function of a start with linear, then single log with you know log dependent variable, then again single log with the log independent variable let is say you know log, but the, but this side with respect to dependent independent, yeah.

So, there are videos you know phones. So, you specify this permanence and see the signal that it is actually prefers. Ultimately you may not actually be sure whether you know the particular functional form is you correct one. Of course, you know it may be passes through a kind of you know specification test and goodness fit test, diagnosis test state. Actually it should be passing through all other you know requirement like you know misspecifications and then, the kind of you know efficiency check through out of sample prediction test.

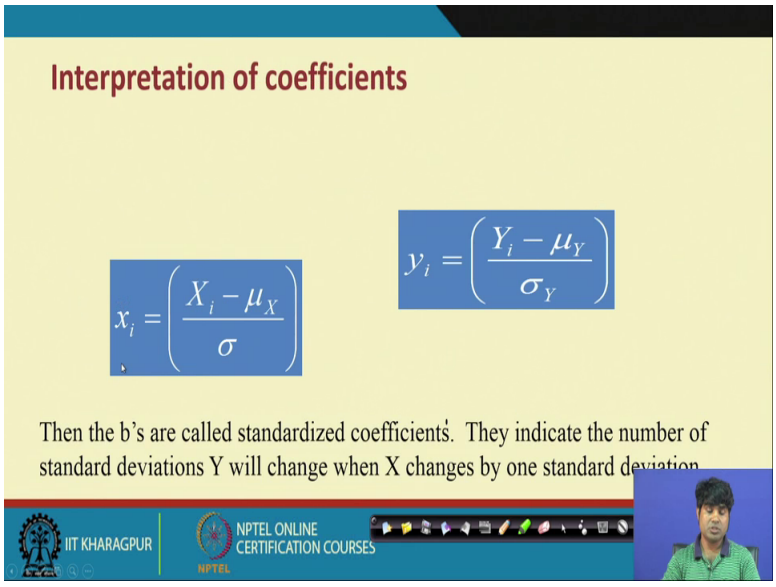
So, it will give you some kind of you know confidence whether the particular specification is very correct one to establish the relationship between the dependent variable and the independent variable.

(Refer Slide Time: 28:30)



Another way you know bringing the non-linearity is equally a unstandardized and standardized process which we have already discussed slightly earlier and this is where the process of standardization.

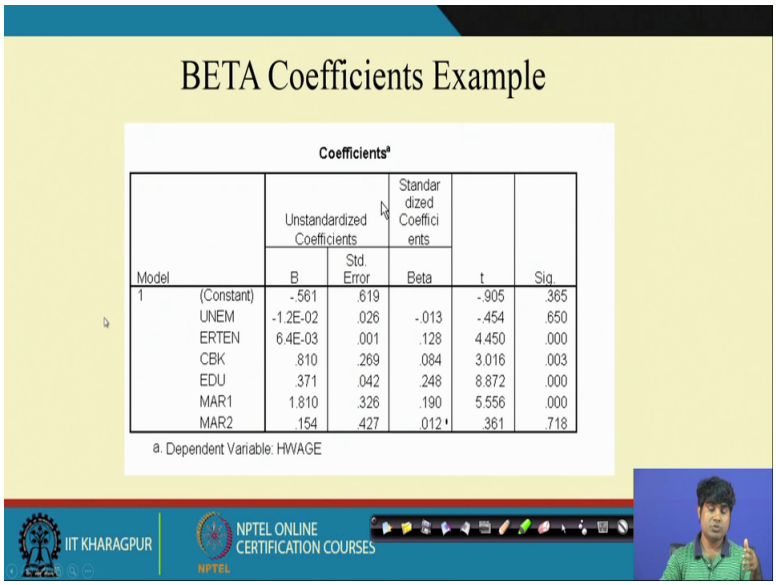
(Refer Slide Time: 28:40)



So, what he can do instead of actually using simple X, you can actually transport data into X by standard deviation. So, that means one way to reduce the volatility and the way you will reduce the volatility, then the non-linearity degree of non-linearity will start actually reducing. So, the balance you know or the outcome is that you know the

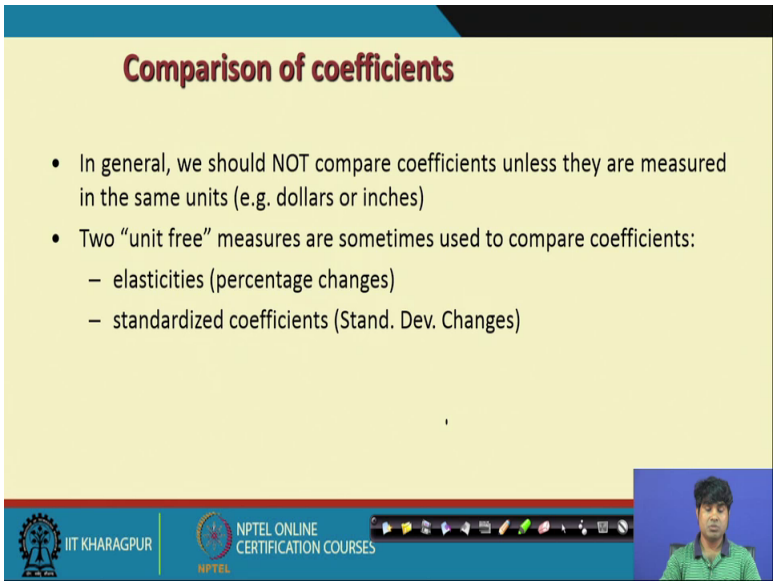
estimated model will be very perfect as per the particularly you know requirement. So, with this you know kind of you know a you know income the experience. So, we have actually some standard output through spaces while actually integrating in this kind of you know functional form.

(Refer Slide Time: 29:26)



So, ultimately the usual procedures or the estimation process is more or less same. It is up to you how you capture or rotate the kind of things and then, check whether there is a significant difference in the estimated output and finally, to pick up a particular functional form that with the indication of estimated output. And then, finally go for the choice which is actually good for as per the theory, as per the reliability and as per the kind of you know requirement in the decision making process.

(Refer Slide Time: 30:13)



So, these are all things are you know various ways to understand these situations and then, represent as well as per the particular in our engineering requirement and the kind of you know decision making process to understand the situation and to bring the kind of environment whether actually the particular outcome will give you a solid kind of an indication and that means technically it will give you better management decisions as per the particularly you know engineering problem requirement.

But, we will stop here and will continue the similar kind of our discussion in the next lectures by taking a particular you know problems, then connect with you know different forms and like to check whether there is a significant difference and how we can pick up a better one you know in response to various alternatives and that too with the help of you know various functional forms. So, with this we stop here.

Thank you very much.