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

Lecture – 32
Model Specification- Choosing the Independent Variables (Contd.)

Hello everybody. This is Rudra Pradhan here. Welcome to Engineering Econometrics. Today we will continue with Model Specification and in the last lecture we have discussed this issue. In fact, when you talk about model specification it may be either the involvement of number of variables this issue is either less number of variables involved in the process as per the requirement which is called as under identification case. And in the second instance more number of variables involve as per the particular requirement that is the over identification case.

And we have discussed both these you know issues and we have gone through the discussion and come to a conclusion that we need actually optimum number of variables and that to be in the system through which you know model will be free from, all kind of you know bias and then it can be used for the decision making process. In the second instance a particular models may be free from all the tests starting with specification goodness speed test and all kinds of you know diagnostics stills the model may not be actually good to you know use for the decision making.

If again it is miss specified that to not with the issue of you know variables involvement over the identification case or under identification case here it is the issue of you know functional form. Sometimes when you when you estimate a particular you know equation, so be careful: what is the functional form we have actually specified there. So, if the specification is not good then by default the estimated output may have some kind of you know a bias again. So, as a result this model cannot be used for decision making process.

Again whatever, maybe the engineering problem, so you so, ultimately we have to again check whether the model is correctly specified or not. So, there are against lots of standard checks are there we means the usual procedure is the as per the classical linear regression modelling that you start with the linear structure, and by using the linear structure if the estimated model or you know estimated outcomes are passing through all the kind of you know diagnostics. Then by default the model will be declared as you know correct specification.

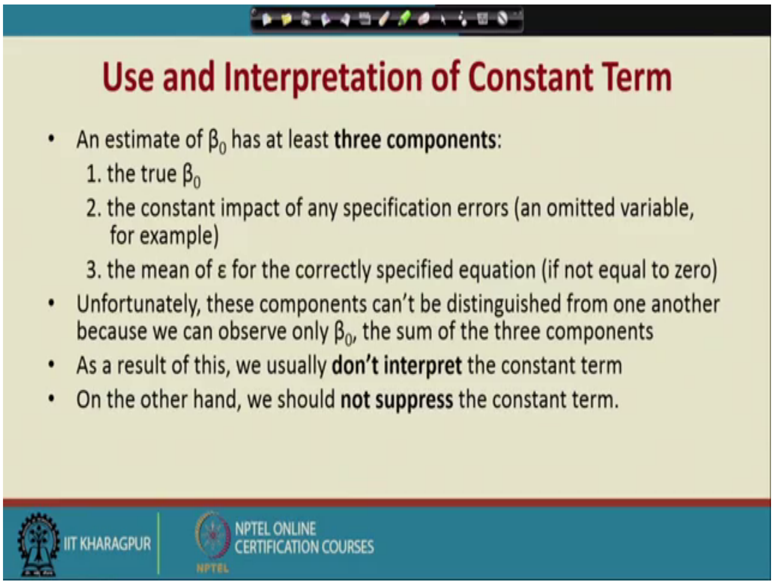
But if there is a some blocks and there is a high chance it may be due to miss specifications and that too because of you know you know un even functional form. So that means, the requirement is let us say exponential type, but the use ultimate use is the linear type. So, now, if you change functional form for and re estimate then by default the model will be free from all errors or all diagnostics and then that can be used for the decision making process. So, what is the requirement?

So, the same problem you have to change different kind of you know functional form then look for the estimations differently and by default you will have different alternatives. Even if we the model actually correctly specified no harm to re-estimate the particular you know model means in this case what is happening here the variable pools are you know constant which we have actually taken care by the previous discussions. Ultimately we have to declare that you know it is not the under identification case not the over identification case. So, it is the exactly identify.

So, if that is the case next check of this you know miss specification is the functional form. So, now, the previous discussion will give you the kind of you know idea that you know this is what the exact number of variables should be in the system. And now the question is how these variables against will be in a particular you know setup, is it in a linear setup or is it in a non-linear setup, and if it is a non-linear setup is it in a exponential type, is it in a logarithmic type, or is it in a kind of a you know quadratic forms or polynomial form or whatever may be the kind of you know ways we can actually represent.

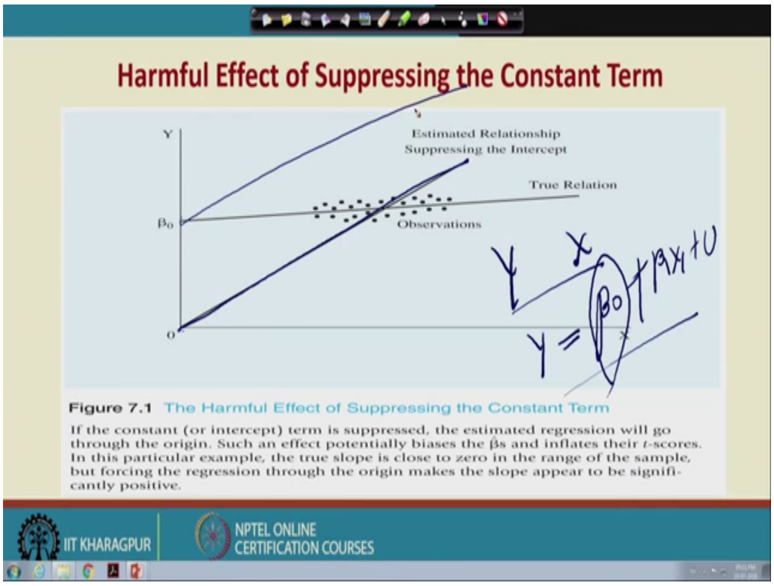
So, ultimately what is the or what is the process? So, what will you do; so number of you know functional forms you have to create and then again re you know re-estimate, and check the output with the previous output which we have already as per the miss specification test and then you declare that you know again the model is correctly specified with respect to its functional form. It is not that you know correct is model is correctly specified with respect to variables involvement, it is also required that you know the model should be declared that you know correct specifications with you know with respect to particular you know functional form.

(Refer Slide Time: 06:06)



Let us see what are the ways we can actually check and you know ultimate to go about it. So, ultimately when you start with you know estimation. So, we have actually a you know important things what we called as you know beta 0 and beta 1, two important you know what we called as you know coefficients. So, ultimately the coefficients you know significance will give you the indication that you know whether the variables is very important or whether the particular model is the free from all kind of you know bias before use for the you know requirement.

(Refer Slide Time: 06:47)



So, what we can do here? So, first we will check with you know the model like you know say starting with let us say we star with Y and X ultimately like this. So, let us say Y and X. So, usually what we start you know as per classical linear regression model. So, we start with the connecting beta 0 plus beta 1 X 1 plus error terms this switch weather there any requirement of beta 0 or whether there is no requirement of beta 0 that itself is a kind of a you know functional form. So that means, if it is actually no requirement of beta 0 then the function with linear linearity it can start with like this. And if you involve you know beta 0 then the model can start with predicting like this.

So, ultimately, so this is for the kind of you know a difference. So, whether the model will be start with the origin or start with you know above the origin that is what the intercept is all about. So, this itself is a kind of you know the kind of you know miss specifications. So, if the requirement is with you know intercept you are bound to have a intercept. If the requirement is with in origin then you can start within origin. For instance in engineering economics we have actually theoretical we have a cost functions. If you are you know means cost function if you say a cost function it is the relationship between cost and output.

So, now, cost can be variable cost can be total cost when it is total cost. So, it involves variable cost and fixed cost advantage variable cost, so fixed cost is not there. So, now, when you are in a spreadsheet showing you know output with a variable cost, so then you can simply start regressing from the origin because it is the meaningful kind of you know structuring. But when your excel spread sheet will give you the data of you know total cost and you know output against all output then you should actually involve you know intercept. So, because the intercept will take care the fixed cost part.

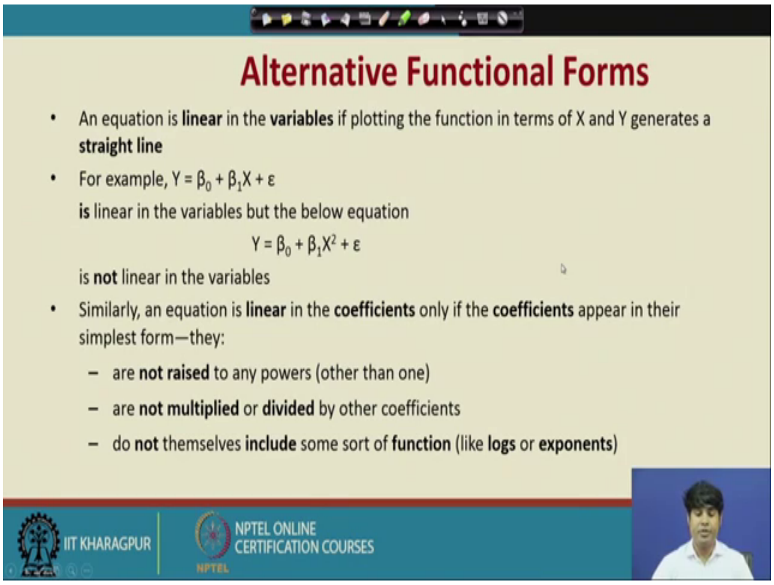
So, that is the theory will give you the indication whether to start with you know origin or whether you start with you actually above than origin. Ultimately a even if you have theoretically you have no clear idea. So, what is the requirement ultimate requirement that is what the miss specification is all about. You have to check and you know declare that you know this model is correctly specified, not only with respect to functional form, but also with respect to the involvement of intercept and you know with intercept without intercept. So, ultimately you can estimate the model with intercept we can estimate the model without intercept. Sometimes we can also estimate the model with you know trend vectors.

So, without constant without two kind of you know trend we can estimate with constant you can estimate and again with constant and trend you can estimate. So, there are 3 levels you can actually estimate the process. And all these actually checks you are supposed to do before you declare or before use this model for the kind of you know engineering requirement. So that means, you are supposed to know what are the ways you can check and then finally, you can declare that means, the requirement here is miss specification.

If the require if the requirement is you know the involvement of intercept you should you know involve beta 0 if there is no requirement you can skip you know beta 0. So, ultimately you cannot take a decision until unless you check and you know verify the results. So, you know you first estimate the model with you know beta 0 and without beta 0. That is what called as you know for instance if we use a software like you know spss spss will give you two sets of output with coefficient and without coefficient that is what called as a standardize and standardize and standardize. The moment you will go for standardize then by default intercept will be removed in the process. So, it will be take you to the origin to have the output and again un standardize means it will take you the kind of you know estimations within you know intercept.

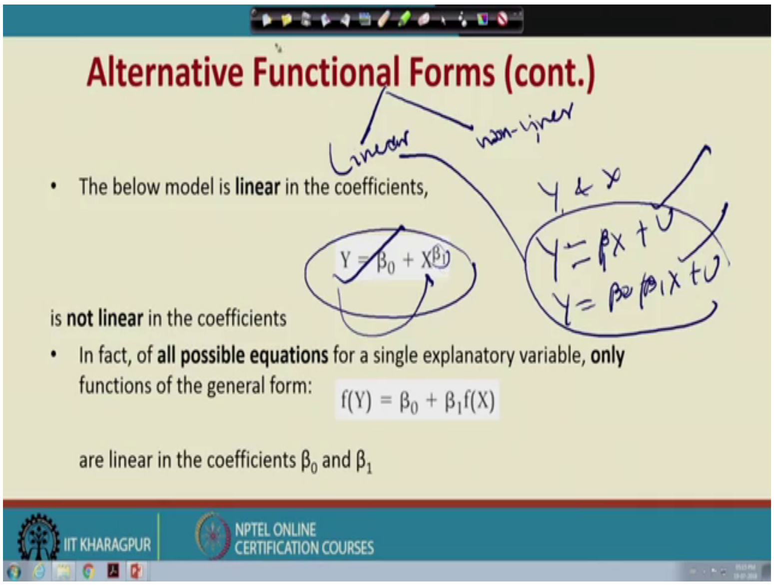
So, this is one way; so, for as a functionality is concerned because ultimately it is the function between Y and X. So, when I am writing Y as a function of X it may be starts within origins simply Y equal to beta X or simply we can say Y equal to alpha plus beta X. So, the both the functions are true but ultimately which one we have to you know use for the decision making process. So, until unless you check and verify you cannot you know use this particular you know model for the requirement. So, this itself is a kind of you know miss specification check you are supposed to do and go ahead with the a requirement.

(Refer Slide Time: 12:12)



So, let us see what are the other way we can actually observe this point and let us say various alternatives you know. Means one is the with intercept without intercept against there are various functionals you may have.

(Refer Slide Time: 12:16)



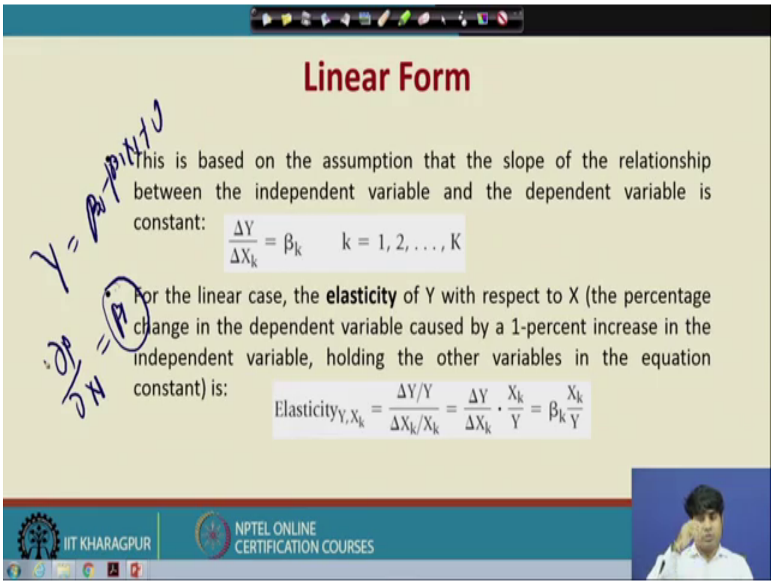
So, the first one is the linear one so that means, if we say functionality a technically if you say functionality this one. So, it may be actually two types linear one linear once and non-linear once. If it is between actually Y and X, Y and X, so the linear representation u will be like this Y equal to beta X plus u or beta 1 X let us say and Y equal to beta 0 plus beta 1 X plus u. So, this is the linear without intercept this is with intercept and again when there is a non-linear, so this will take you to the this basket. Then when it is a non-linear so then this may be many different forms again with intercept without intercept. So that means, different levels of you know structuring. So, every levels specify the functionality and then go for you know estimations.

So, the involvement of data is remains same now we are changing the functional form and estimating the model. That means, here variables variable is number of variables involvement is constant size of the data is a constant, what we are checking the kind of you know miss specification is with respect to functional form. One way we can call as you know is robustness check having different functional form with the you know specification of you know number of variables and the kind of you know size of the data then you are checking that you know whether the model is a correct place specified. Or that means, technically whether the model is good enough estimated model is a good enough to go ahead with the predictions or the kind of you know decision making process.

So, here is the game. So, we like to what are the number of ways you can check you know or you can estimate the models and declare that this these particular functions functional form is equal to go ahead with the forecasting. So, this is second levels where the model is actually Y equal to beta 0 towards X to the power beta 1. And what is happening here, so the variable means it is not linear in the coefficients the coefficients are means we will have a two different ways. The variables are linear in natures and coefficients are linear in natures, and variables are not linear in nature constants are not linear in natures for a.

For instance in this a in this functions variables are actually linear in format but its non-linearity in coefficients. So, means all kinds of you know specification will be there, but ultimately we will check actually and so for as a classical linear regression modelling is concerned. So, the parameters are you know linear in natures. We like to check the variables you know structural formality directly. So, let us see how is this particular you know flow.

(Refer Slide Time: 15:38)

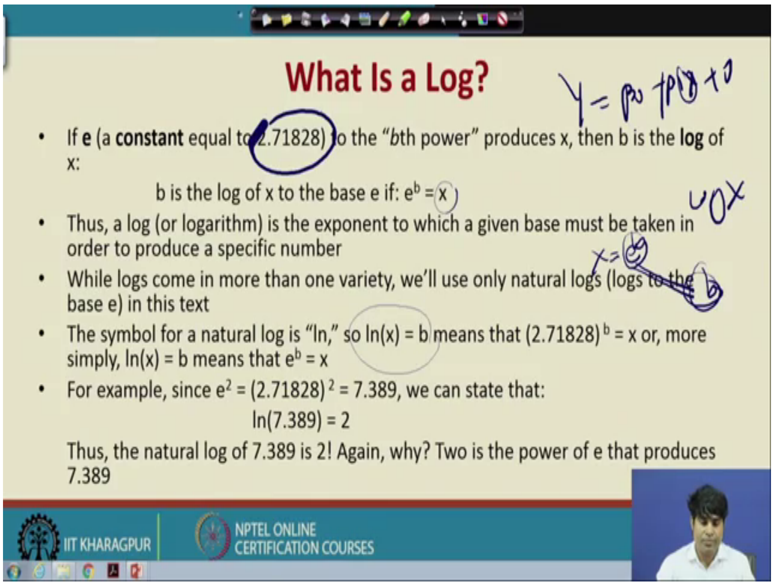


So, a linear form of you know in the case. Ultimately what will you do in the modeling after the estimations we like to study the impact of that particular variable with dependent variable and that is how we line to check the elasticity, ok, we like to check the elasticity coefficient. So, here is the elasticity coefficient. For instance if it is Y equal to you know beta 0 plus beta 1 X 1 plus u. So, d yy d X once will give you the beta 1 coefficient variable. So, that is that is the weight you can actually estimate.

So that means, once you estimate the models. So, we can get the beta coefficients that is elasticity which can represent that you know a one unit change of you know independent variable what is the impact on you know dependent variables. And the unit change of independent variable to the dependent variables varies means that particular you know weight will vary depending upon the various functional form. If the function is linear once then the particular elasticity may be different and whenever change the functional form then the weight will be slightly change.

So, that is why a picking up a you know perfect functional form will have a right choice for the decision making process, otherwise it may give some kind of you know bias results. So, you must be very careful about this particular you know structuring and that to first one is the linear form and if a if it is not linear form then we have a different you know kind of you know labels. So, one of the particular table is actually log form and we will call as you know log form that means, actually when we when we say that you know d Y by d X ok.

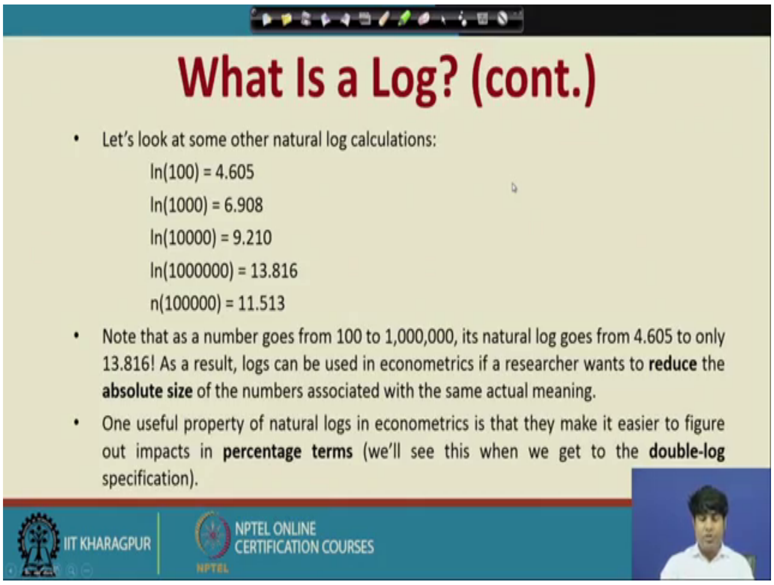
(Refer Slide Time: 17:37)



So, that is the coefficient through which actually where you know getting the particular you know weight factor. But now when the functional form will change then ultimately it will have a different value all together. For instance let us say a you know exponential type logarithmic means it is a counterpart is the exponential. So, let us say e to the power b equal to X so that means, so, here actually the value a value e e e value will be added here into the process to represent the impact of you know X; so, log X equal to b, so ultimately X equal to e to the power b.

So that means, see here if you compare this ones log X equal to b so that means, ultimately when I am writing Y equal to beta 0 plus beta 1 X plus u and ultimately X is nothing but actually log X. That means, we have transfer the data then in that case you know your X value will be your X value will be e to the power b only. Now, if it is only X then it is simply b. So, now, e to the power b and b there is a big difference and that is what the functional form will give you know kind of you know difference and that you know any minor difference may affect the decision making process. So, that is why this particular miss specification need to be checked and then it should be declared that you know the model is you know correctly specified before you use this model for any kind of you know decision making process as per the engineering requirement.

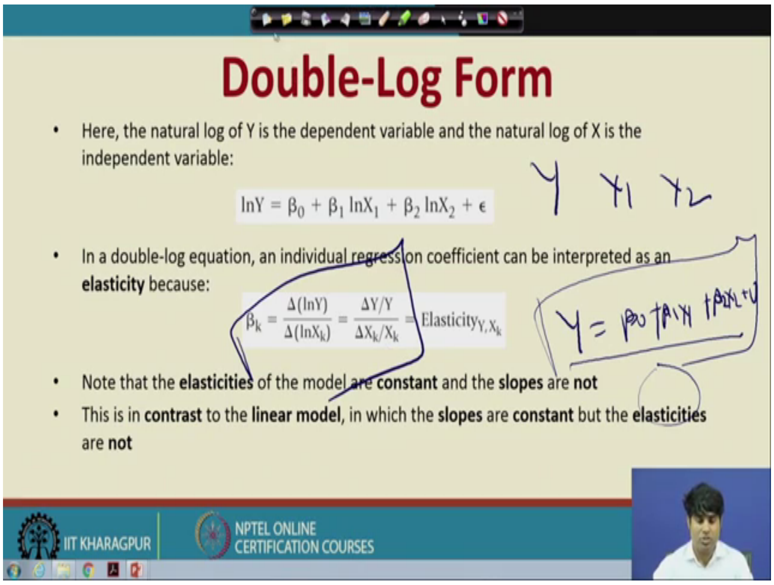
(Refer Slide Time: 19:34)



So, ultimately actually a this is all together understanding of the log and log 100, log 1000, log you know 1 lakhs like that you know how what is the weight we can actually put we have already put you know log e here and then log 10s, log 100. So, that that is the value actually that means, the weight of this particular variables will be you know added you know in a multiple ways. So, that is why a transferring the kind of you know variables and the kind of you know a functional form will have a different kind of you know weight structure and the different kind of you know kind of you know outcome altogethers.

So, so that means, technically the summation is that you know it is the or it is a kind of you know a mandatory requirement to check this particular misspecification that to with respect to various functional form.

(Refer Slide Time: 20:38)



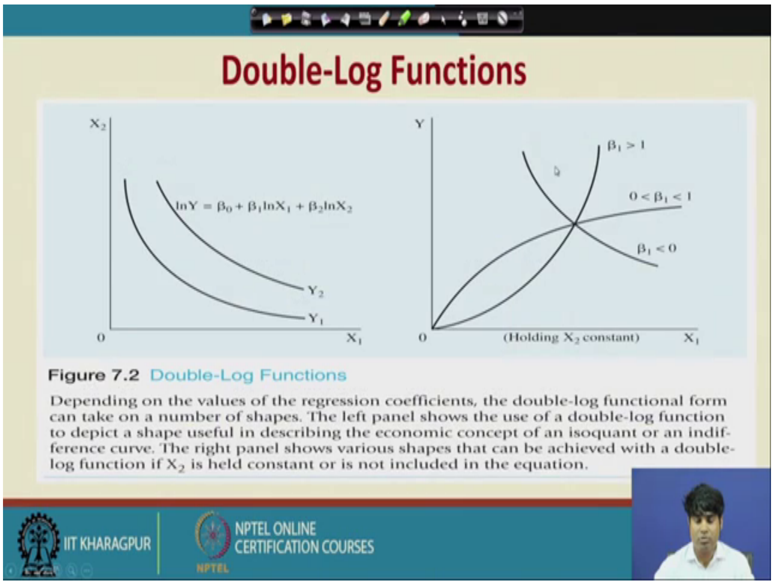
So, here if it is log then there are 3 forms, either you know kind of you know double log, or it can be a semi log or it can be you know either way with you know original variables with log variables. So, that is what kind of you know semi log and the kind of you know reverse way also.

So, if it is a double log then that means, both the sides you have to take the a kind of you know logarithmic for instance dependent variable and independent variables. So, this is the right choice. Here the ultimately we have variable Y X 1 and X 2 we can simply start with Y Y equal to beta 0 plus beta 1 X 1 plus beta 2 X 2 plus U ok, but ultimately a you can go ahead with estimations and had the output that is what called as you know output with respect to linear functional form.

Now, what will we do? We change the orientations now instead of Y X 1 and X 2 we transfer the data log Y log X 1 and log X 2. So, now, this same models we can you know estimates, but here the data will be not the original data or the actual data it is the transfers data and the transformation is the logarithmic transformations. So, ultimately we have to instead of Y we use logo Y instead of X 1 we will use log X 1, instead of X 2 will be use log X 2 and the difference will take care actually at the elasticity level only.

So, here it is not d Y by d X ones it is the d log Y by d log X 1. So, this will give you a different kind of you know (Refer Time: 22:31), so log. So delta of log Y means if you simplify it is actually the kind of you know elasticity coefficient. So, Y by X divided by d X so that means, Y by X that is the coefficient has having more you know another weight will be added into this particular process if you change that particular function functional form. So, means a simple to some you know develop functional form. So, that is the difference you will find here.

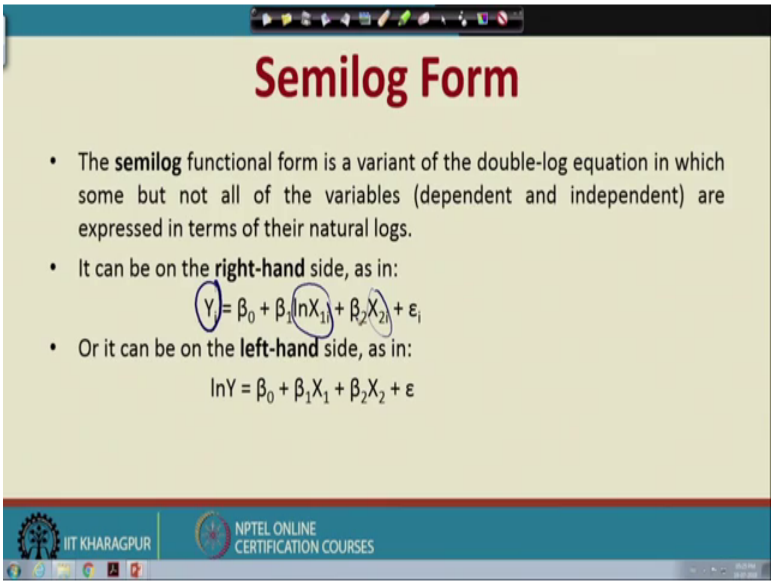
(Refer Slide Time: 23:03)



And that that is how there is high requirement of you know checking the miss specification of the models by changing the functional form. And you see here and this is how the you ultimate you know graphical look about the you know estimations and it is a logarithmic function. So, plotting can tell you somehow you know like you know we need theory to integrate. So, what kind of you know functional form we can use whether it is the log X or simple X or log Y or with the simple Y.

So, after plotting a you can get to know little you know structure. Even if you know you cannot able to plot. So, no harm actually you just arbitrarily change the format and then re-estimate and again a compare these outputs and finally, declare that this is the functional form through which you know model should estimate and then the estimated outcome can be used for the kind of you know engineering requirement.

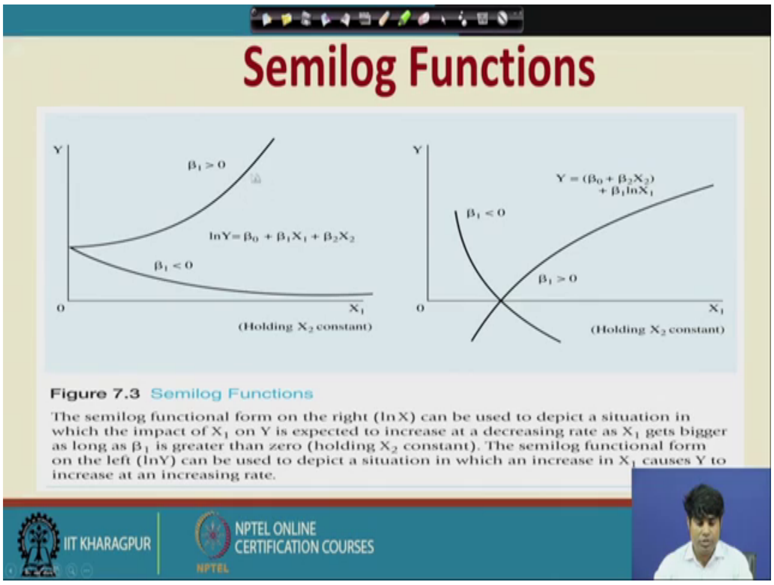
(Refer Slide Time: 24:10)



So, that is the choice and you will have different kind of you know setup. So, a if not double log and then we have a semi log. So, here Y is the original actual and X 1 is actually transport X instead of original X 1 so we will use log X 1, and then again X 2. So, that that is a allowed. So that means, log Y and X 1 and X 2 against Y with log X 1 and log X 2 against Y with log X 1 and X 2 and another way Y X 1 with a X 1 with log X 2.

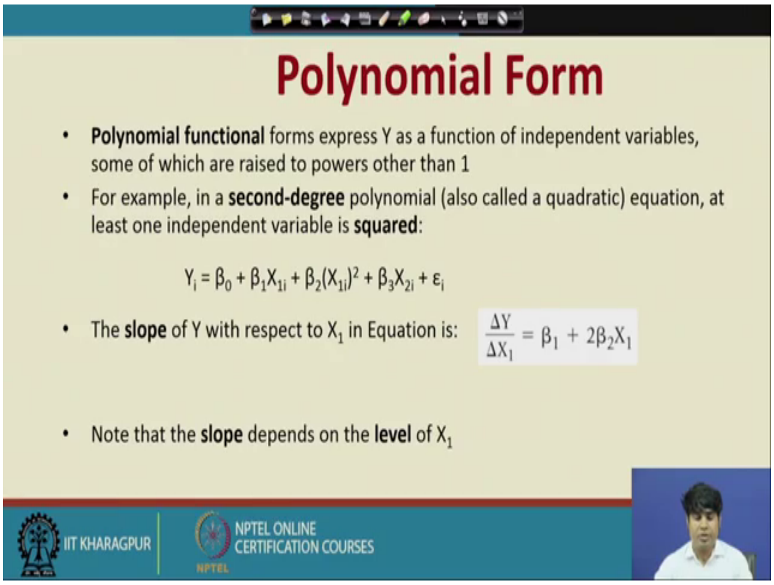
So that means, these are all various options single log, double log against you can actually allow another log both the sides and then you can go ahead with the estimation. Ultimately doing all this things you are normalizing the data, and ultimately there is a high chance the accuracy of this particular estimation will be very high. Not only always true, but sometimes it is also it is you know it is the fact, so no harm again. To do this processing re-estimate and then re-check the kind of you know out come and then compare and declare which one is the final choice as per the particular requirement.

(Refer Slide Time: 25:41)



So, this is the graphical look of you know semi log functions, ok.

(Refer Slide Time: 25:46)



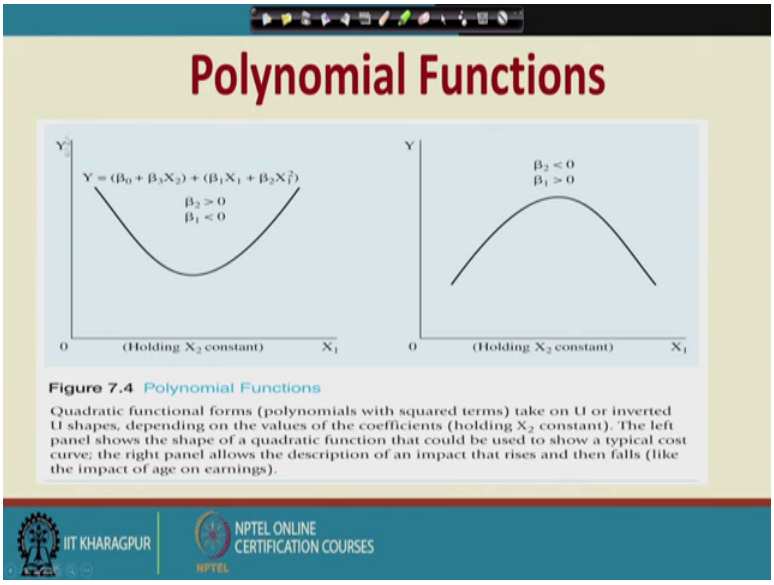
So, if not actually single log, or double log or semi log or not linear then there can be another form also that is called as a you know polynomial. Polynomial is actually multiple kind of you know structure with you know single degree means polynomial of order 1, order 2, order 3. Like you know that means, the moment you restart you know using polynomial forms so you can have actually multiple option all together.

And again it will give you lots of you know alternatives that to with respect to functional forms. So, you just specify the functional form and then you go ahead with the estimation. So, polynomial of degree 1, polynomial of degree 2, polynomial of degree 3, polynomial of degree 1 is nothing but called as you know linear format if not then it will start with you know quadratic cubic and so on and every times you like to check whether the particular functional form is very effective a separate you know theory and data and the kind of you know engineering requirement is concerned.

So, for instance if you say polynomial at least one of this particular variable should have you know polynomial structures that is the degree should be starting with you know second degree. So, it is in a quadratic format and where we are integrating Y with X 1 and then X 1 here and with you know X 2 sometimes you know you know it doing this the estimated process will be very effective and it will add value to the systems and as a result the decision making process will be very efficient all together.

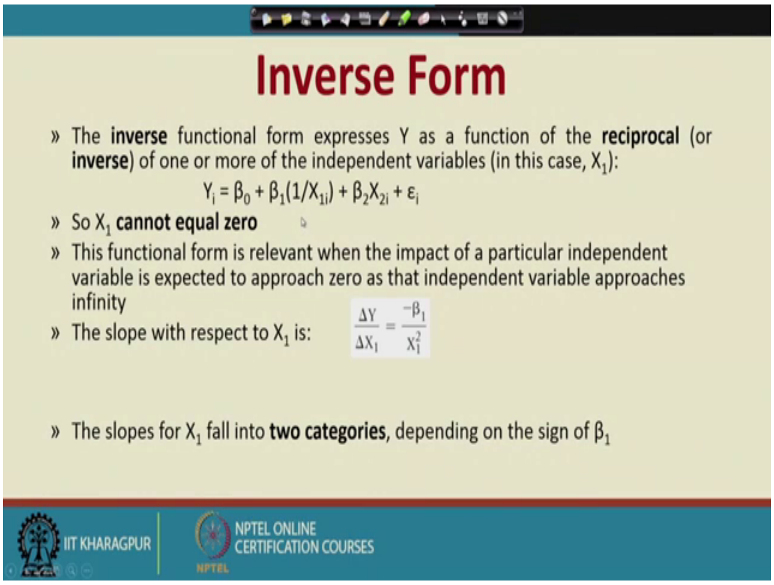
So, again there is no harm to you know use this form and re-estimate and check and verify the output. Ultimately it is the it is the search process and in the search process there is a opportunity cost. So, every time you have to check which one is the good you have to jump you have to kick the you know the other one. So, like that this is the procedure you have to follow and then go ahead with the kind of you know prediction and decision making process.

(Refer Slide Time: 28:12)



So, this is the polynomial form of the you know graphics and that to polynomial of degree two and when the increase the degree. So, ultimately graphical you cannot check what you can mathematically you can connect and estimate, there is a there is no harm at all.

(Refer Slide Time: 28:32)



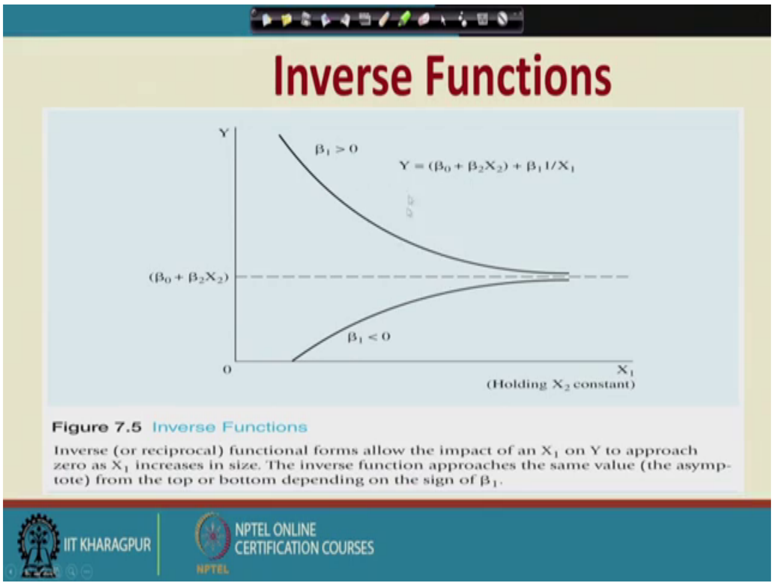
And you can also have inverse functions. So, now that means, technically same way Y X 1 and X 2 instead of using log X 1. So, we can use actually 1 by X 1, ok. So, against we are actually standardizing the particular you know variables. So, doing this sometimes you know relatively reduced and in that case the model accuracy maybe high but while doing this. So, be ensure that you know data will be very supportive for instance one of the conditions by using inverse function is that you know X 1 cannot equal 0. So, in the dataset if X 1 is coming in any sample is coming 0 so by default you cannot use this particular you know functional form, so that is the you know restrictions or we can say that you know conditions to use you know inverse form.

So, when you use inverse form first check is that you know check whether any data for a particular variable is having 0 or non zero. If it is having at least one 0 then it should not use this ones because 1 by 0 by ultimately it will give infinite and when we use the this you know value and software then the model will not run and ultimately will not get any kind of you know outcome.

So, that is why before you start the process you must sure that you know it is satisfying the criteria of inverse functions and then build the functional form and re-estimate the model. And again check the reliability compare with the other one and then see whether there is improvement or not if there is improvement go aheads and then you block the airs that itself give you the kind of you know what we can called as you know specifications.

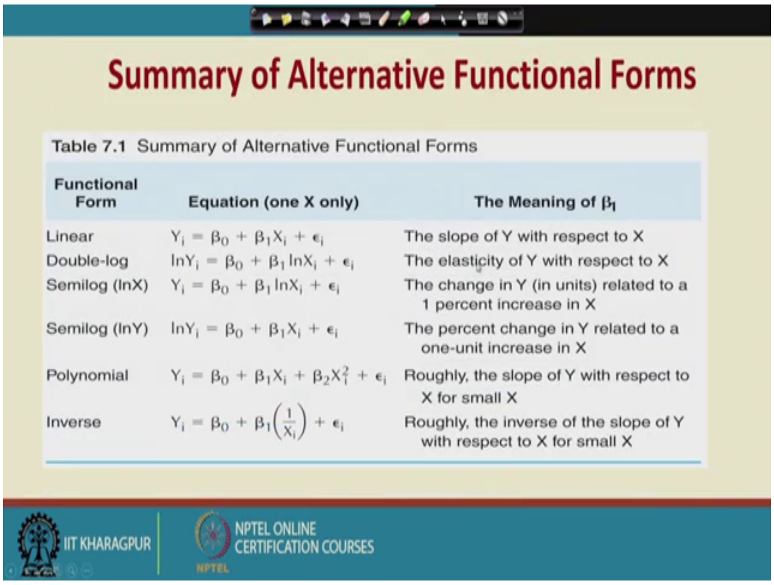
Other, the moment you will be say that you know for instance you estimate the model in a linear form and then you re-estimate the model in inverse form and you will found there is better result here. Then by default the previous one will declared as you know miss specified model. So, you use the inverse form of estimated output and that will that output by default will be declared as you know correct specification that is what the flow you have to maintain every time.

(Refer Slide Time: 31:03)



So, this is the graphical look of you know inverse functions.

(Refer Slide Time: 31:07)



Ultimately we have no business about the graphical loop, it is how it the requirement ultimate requirement is the estimated model output. And we like to check whether the model output is very good by changing the various functional forms; so, if start with linear check the diagnostics and again check the miss specifications by fixing the optimum variable structures. And then again look for the various functional form and then finally, you declare that you know the model is actually correctly specified and free from all the all the errors and I have gone through all the diagnostics and then the model is ready to apply for the decision making process and to solve some of the engineering problems.

So that means, it is not so easy actually it is very complicated again, and its very time consuming process, but it is the need and it is the requirement. Until unless you do all these processing you cannot get a good models and you cannot declare that you know this is the model through which you can do the decision making as per the particular you know engineering requirement. By the way these are all mandatory requirements and you are bound to check and then you will be declare that you know model is you know correctly specified, and the best, and then can be used for the kind of you know particular engineering problems requirement.

With this we will stop here and this is all about the diagnostics of you know engineering econometrics, and in the next lectures we will be starting with you know non-linear econometric modelling. So, where we have a various forms of the models and some of the engineering problems requires such kind of you know modeling to forecast or predict the engineering output or you know engineering requirements. With this we will we will close this lecture.

Thank you very much.