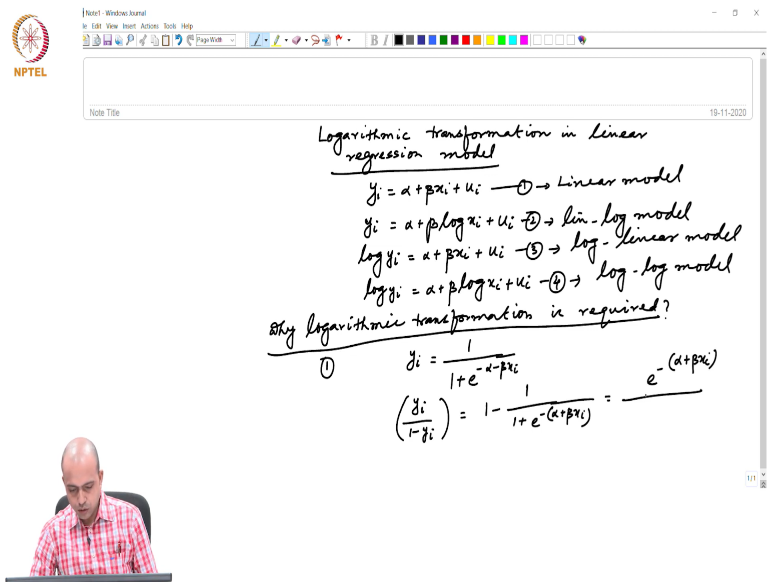
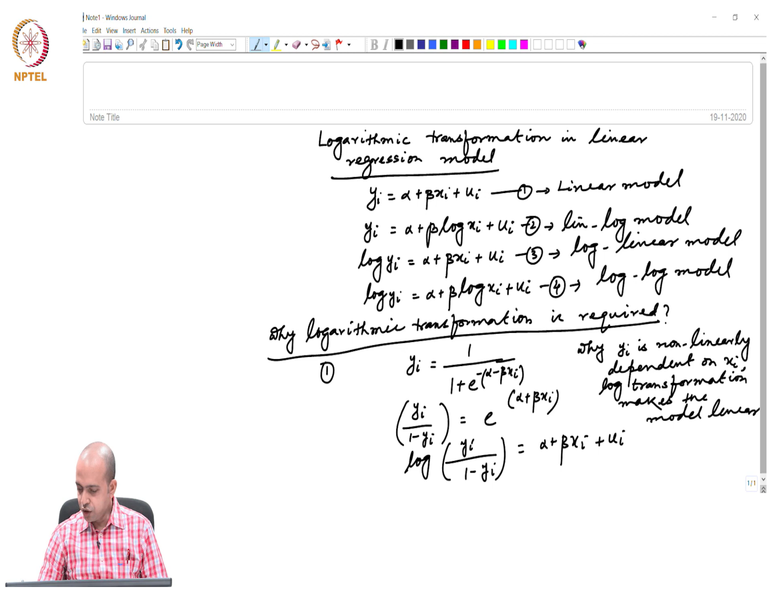
Introduction to Econometrics
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology Madras
Goodness of fit measure, ANOVA and Hypothesis Testing Part 2

(Refer Slide Time: 00:14)





Welcome. Today we will discuss about an interesting topic for empirical estimation of your econometric model. So, the topic of our discussion is logarithmic transformation in linear regression model.

So far, the econometric models, what we were discussing, they were linear in nature. That means our model was like this – $Y_i = \alpha + \beta x_i + u_i$. Let us say that this is model 1 which is called linear model. But sometimes when you work with data or when you read your textbook or some applied research paper you see that some of the variables, either independent or dependent variables is transformed into log.

And there are three alternative ways by which actually you can transform your model into a logarithmic transformation. You can either transform your independent variable x and keep Y as it is. Or, you can transform your dependent variable Y into log and keep x as it is. Or, you can transform both Y and x in log form. So, there are three alternative possibilities.

The second possibility, that means this is model number 1 where both Y and x are linear are unlogged, that means they are not log transformed, this is called linear model. The second model let us say that $Y_i = \alpha + \beta \log x_i + u_i$. This is, let us say, model 2 where we have transformed only x into log and the name of this model is called lin-log model. That means your left hand side is linear, right hand side you have transformed x into log. That is why it is called lin-log model.

Possibility three is this – $\log Y_i = \alpha + \beta x_i + u_i$ and this is let us say model 3 which is known as log linear model. And possibility four is – $Y_i = \alpha + \beta \log x_i + u_i$ which is called log-log model or sometimes it is known as double log model.

Now, depending on which particular functional form you assume for your regression model, interpretation of beta will change. So, my today's lecture is to make you understand two things. First of all, why do you require this logarithmic transformation and secondly what would be the interpretation of your beta hat if you estimate model 1, model 2, model 3 and model 4.

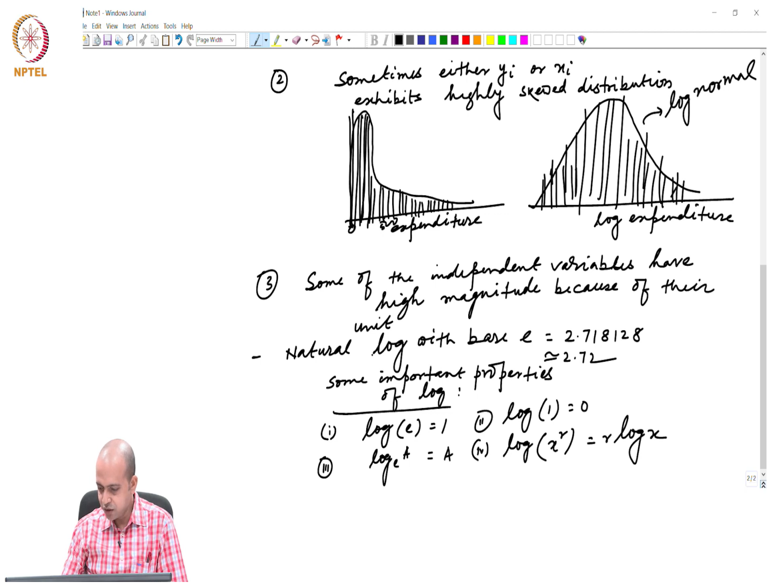
So, interpretation of $\hat{\beta}$ is different in these four alternative models. And that is what we are going to learn. So, let us first try to understand why logarithmic transformation is required in empirical research. There are several reasons. I will discuss one by one.

First reason is, what happens sometimes, your dependent variable Y and independent variable x, they have some kind of non-linear relationship. And we have not yet learnt non-linear estimation technique. So, we need to somehow linearize that apparently looking non-linearity between Y and x so that we can actually apply linear regression technique estimation to estimate your beta hat. What I am saying, sometimes, your dependent variable Y is non-linearly dependent with your x. How is this? Let us say that your $Y_i = \frac{1}{1+e^{-\alpha+\beta x_i}}$. This is your model. So, you can easily see that Y and x, they are non-linearly dependent. And we have not yet learnt non-linear estimation. So what we will do, we can linearize this model by logarithmic transformation.

How to do it? Let us say, (Yi/1-Yi) equals to (1-1/1+e to the power minus alpha plus beta xi) And that means, this would become $\overline{e^{-\alpha+\beta x_i}}$. So, you can easily solve this. And then if you take log of this, Yi/1-Yi, it would become $\alpha+\beta x_i$. So, when your Yi and xi, they are non-linearly dependent, then Yi/1-Yi you can easily understand that you can calculate 1-Yi first, then you divide this, divide Yi/1-Yi and you will get $e^{\alpha+\beta x_i}$

And then if you take log, it would become log(Yi)/1-Yi equals to $\alpha+\beta x_i$. You can add the stochastic error term here. Then you can easily estimate this model using linear estimation technique. So, that means one reason for logarithmic transformation is to overcome the non-linearity between your Yi and xi. So, I will write shortly, when Yi is non-linearly dependent on xi, log transformation makes the model linear. This is one reason.

(Refer Slide Time: 11:00)



Second reason is sometimes we see that your Y variable, dependable variable or some of your independent variables are highly skewed. So, sometimes either Yi or xi exhibits highly skewed distribution. For example, let us say, you are estimating a consumption function. And if you look at where consumption is a function of several variables, you are estimating a consumption function and if you plot the histogram of that consumption data what you see that this is let us say 0 and then this is 0 then 500 then 1000 like that it is having. So, there will be huge data here and then in this range and then you see this type of distributions.

So, that means your distribution is like this. So, this data is left skewed. And for our estimates to be unbiased and efficient, what we want that they should exhibit some kind of normal distribution. And if you transform this expenditure into log and then if you plot the log expenditure here, then the histogram will look like this.

So, that means it will be more or less normal which is called a log normal distribution. So, that means the distribution becomes normal after logarithmic transformation. So, this is another reason when your either dependent variable or independent variable, they are highly skewed in nature, so that means there are too few observations with higher value or too few observations with lower value. Then that distribution is either left skewed of right skewed.

So, what you do, you transform the variable into logarithmic transformation and then what do you see? That your log, the result in distribution of the log transform variable shows more or less the normal distribution which is known as log normal distribution in the literature. So, these are the mainly two reasons for which we can transform our variable into either the dependent or independent variable into log transformation.

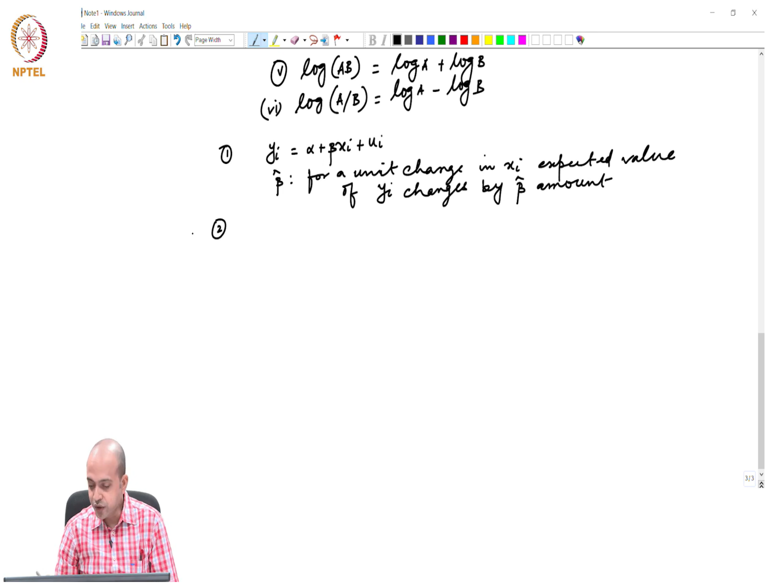
There are other reasons. Sometimes you see that some of your independent variable, because of their unit have high magnitude. For example, you are measuring let us say, your x1 is monthly income or yearly income measures by rupee and you have another variable, which is let us say, a family size.

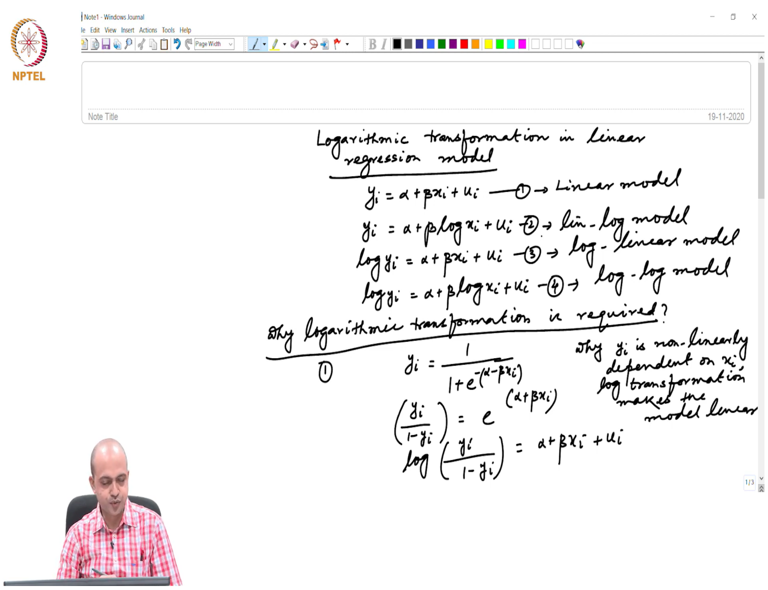
So, x1 will give you value like 1 lac where the family size is only 5. So, if you include this type of variables because of their unit, some variable shows high magnitude and some variables show very low magnitude because of their unit. So, you can transform the variable like income, expenditure, this so on and so forth into log transformation. So that you can get a parity in terms of their magnitude of all the variables in the model.

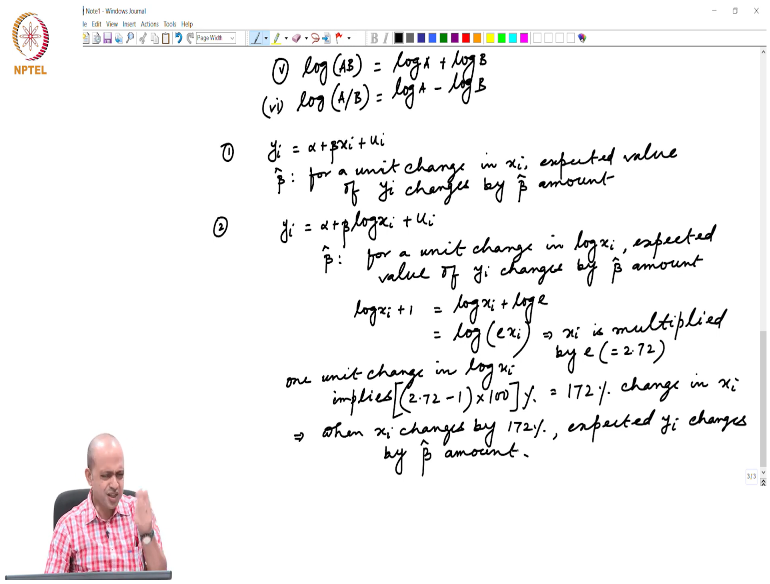
These are some of the reasons for which this logarithmic transformation. And when we transform the variables into log, please keep in mind that here, we consider natural log. Here, we are considering natural log with this e equals to 2.718128 something like that. Or you can say that this is 2.72 roughly. This is the natural log we are considering here with a base e.

And some of the important properties of logarithm would be useful to understand the interpretation of your beta hat. First of all, since we are considering natural log, which is base e, we can say that log(e) equals to 1. Secondly, log of 1 equals to 0. Thirdly, log of A base e equals to A. Then fourth, log of x to the power r equals to r log x. These are the important properties.

(Refer Slide Time: 19:58)







And then fifth, log of AB equals to log A plus log B and vice versa. Then, log of A by B equals to log A minus log B. These are the some important properties of logarithm that we are going to use here. So, we have understood why logarithmic transformation is required and some of the important properties. We are going to use natural logarithm for our purpose.

Now, we will take all those four models one by one and then we will try to interpret the $\hat{\beta}$. So, model number 1 which is a linear model, $Y_i=\alpha+\beta x_i+u_i$. This is the linear model. And what is the interpretation of $\hat{\beta}$ froms this model? As we all know, for a unit change in xi, expected value of Yi changes by beta hat amount. This is the interpretation we know so far.

For a unit change in xi, on an average or the expected value of Yi changes by beta hat amount. Now, I will take our second model which is basically, what was our second model? Yi equals to alpha plus beta log of xi plus ui. So, we need to understand the interpretation of beta hat from this model.

Now, if you apply the interpretation of beta from model 1, then what we can write? For a unit change, I am not doing anything, I am just replicating the interpretation of beta hat from model 1 into model 2. So, what will happen? For a unit change in, instead of xi, I am saying log xi because my independent variable is now xi. For a unit change in log xi, expected value or Yi changes by beta hat amount.

But, if you say for a unit change in log of xi, understanding this interpretation becomes little difficult, because what we want for a unit change in xi, we can easily understand the change in x, but understanding 1 unit change in log x becomes little difficult. So, we will convert this interpretation into a unit change in xi. How to do this? So, when I am saying that for a unit change in xi, let us say that log x is increased by 1 unit.

So, log x, if it is increased by 1 unit, it will become log xi plus 1, this is 1 unit change in xi. This we can write as log of xi plus log of e from the properties we have discussed earlier, that log e equals to 1. And this we can write easily as log of e into xi. This is also another property- log A plus log B equals to log of AB.

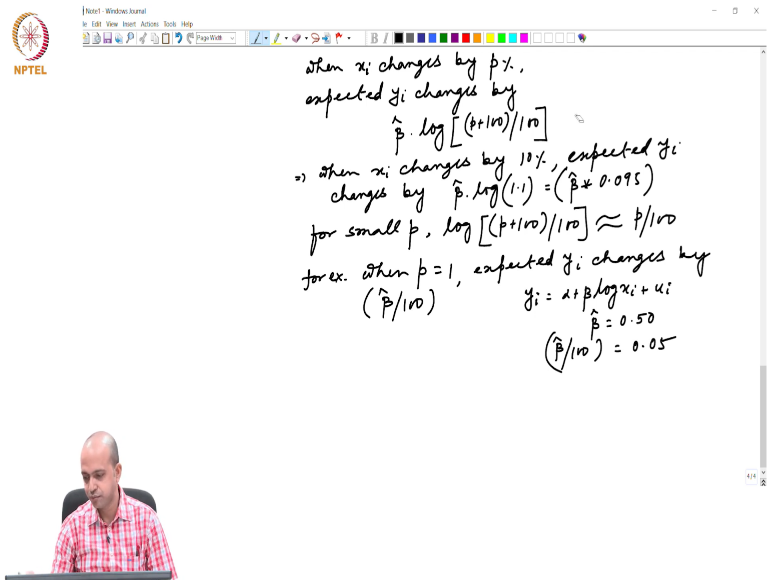
So, that means we can write that when log xi is actually increased by 1 unit, that basically means that your x is multiplied by e. So, that means it implies that x is multiplied by e which is equals to 2.72. So, 1 unit change in log xi implies 1 unit that means ultimately what I can write? That 1 unit change in log xi implies 2.72 minus 1 into 100 percentage change in x.

So, that means we can say that 172 percent change in xi. Please try to understand. First I said that what is the interpretation of beta hat? For a unit change in log xi, Yi changes by beta hat amount that is what we are putting here. So, for a unit change in xi, expected value for a unit change in log of xi, expected value of Yi changes by beta hat amount.

Now, 1 unit increase in log xi is nothing but log of xi plus 1 which is log xi plus log e. Log xi plus log e and that is nothing but, that means I am saying that 1 unit change in, increment in log xi is nothing but multiplying x by e which is equals to 2.72. Now, if you convert this into proportional percentage change, that is nothing but then 2.72 minus 1 into 100 which is nothing but 172 change.

So, that means what does beta hat indicate here in this model? When x changes by 172 percent that means ultimately what we got when xi changes by 172 percent, expected value, expected Yi changes by beta hat amount. But this is also something difficult to understand. That means when I am saying xi changes by 172 percent, this is something max, this is not very comfortable for us to understand. Rather what we want some smaller percentage change in xi.

(Refer Slide Time: 29:49)



Let us say when x changes by p percentage, we can use the result of this and say that when x changes by p percentage, expected value of Yi changes by beta hat into log of p plus 100 by p percentage. So, that means if you say that let us say that p is 10, there is 10 percent change in x, so this implies when x changes by let us say 10 percent, expected Yi changes by beta hat into, if you put a 10 here, 110 by 100 that means log of 1.1.

So, equals to beta hat into let us say, log of 1.1 equals to 0.095 percentage. For small p, for very small p, this log of p plus 100 divided by 100 is actually equivalent to p by 100. So, that means let us say, for example, when p equals to 1, that means we are saying there is 1 percent change in x, when p equals to 1, expected value Yi changes by beta hat divided by 100 percentage.

That means, let us say, after estimating your model, this is your model Yi equals to alpha plus beta log xi plus ui and you estimated beta hat equals to let us say 0.50, this is your estimated beta hat. So, what is the interpretation? For 1 unit change in x, expected value will change by beta hat divided by 100. So, that means this would become 0.05 percent.

So, this is how we can understand the interpretation of beta hat when your independent variable is transformed into logarithmic transformation. Sorry, this is not percentage, this is unit, beta hat. So, for a 1 percent change in xi, expected Yi will change by your beta hat divided by 100. That is what you have to understand.