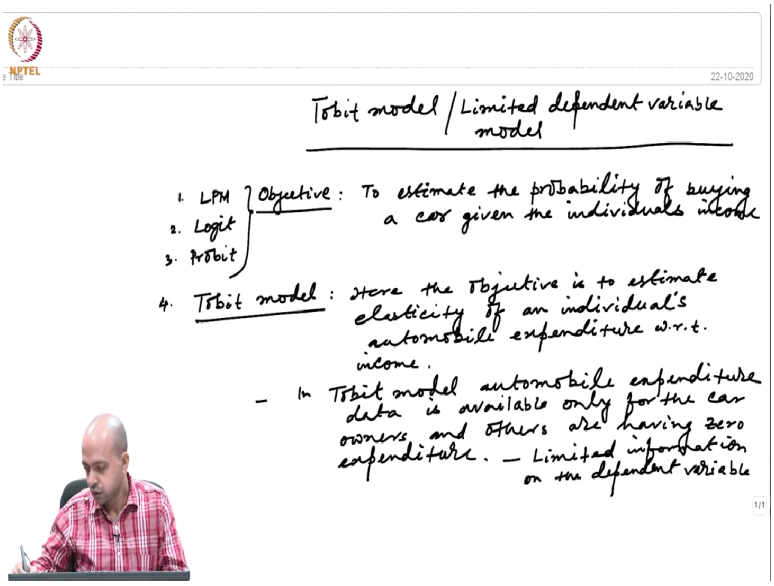
Introduction to Econometrics
Professor. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras
Qualitative Response Models- Probit and Tobit Models Part - 2

Welcome, today we will start discussing about a new model which is called limited dependent variable model or Tobit model.

(Refer Slide Time: 00:35)



So, models where we have limited information on the dependent variable. Before we discuss about Tobit model, let us go back and see, in linear probability model, Probit model and Logit model, these are the three models we have discussed so far. And if you recall, in all these three models, our objective is basically to estimate the probability of let us say owning a car, owning an house or anything else given the information of the individuals let us say income. we have only one explanatory variable income and we would like to predict what is the probability that the ith individual will buy a car depending on the ith individual's level of income. That was the objective.

So, that means in LPM or Logit or Probit our objective is to estimate the probability of owning or let us say buying a car given the individual's income. We are considering only one explanatory variable that is income, it is the probability.

But now, in Tobit model instead of predicting or estimating the probability, let us say that our objective here is little different. We want to estimate the elasticity of a particular individual's automobile expenditure with respect to income. So, it is not probability, I want to say that if the income changes from 10,000 to 15,000 or 15,000 to 20,000 or 20,000 to 25,000, how does individual's automobile expenditure respond to that change in income, that is our objective.

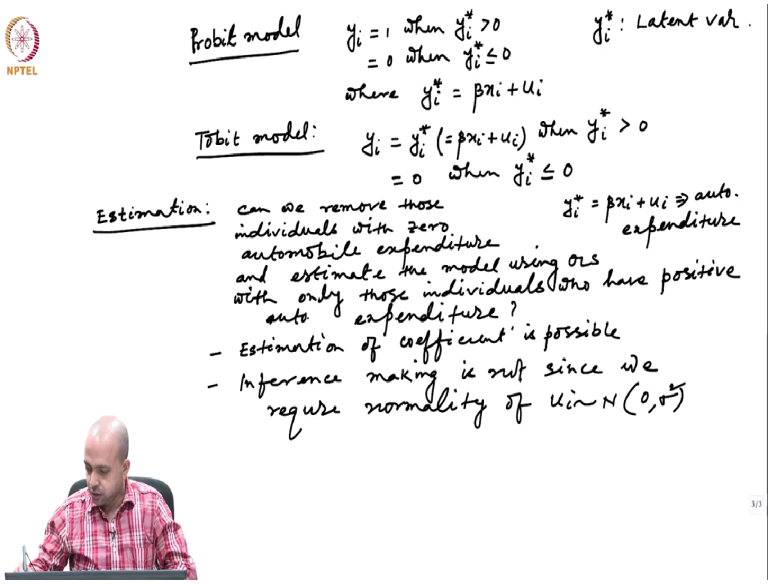
So, that means basically, in Tobit model, what is our objective? Here, the objective is to estimate elasticity of an individual's automobile expenditure with respect to income. So, I am not at all interested in estimating the probability whether you will have a car or not, rather I want to see how does your automobile expenditure respond to your income.

Now, if this is my objective, then when you go and survey and ask the people the expenditure for automobile, some individual who is having let us say i10, they will say 5 lakhs, 7 lakhs, 10 lakhs, 15 lakhs, 20 lakhs, 1 crore like that and then correspondingly you will note down your income also. But there would be some households who will say, I have no car actually. So, that means my expenditure for automobile is 0. So, that means you cannot actually absorb that particular individual's automobile expenditure unless and until that individual has his or her own car. That is what we have to keep in mind.

So, that means, we do not have information on the dependent variable for all the individuals rather we have limited information on the dependent variable about automobile expenditure only for those households who are having a car. That is why this model is known as limited dependent variable model. So, that means, what do we can say in Tobit model, automobile expenditure data is available only for the car owners and others are having zero expenditure. I cannot absorb actually your expenditure.

Now, the question is, how will you model this situation? That is why the name limited dependent variable. So, limited information on the dependent variable that is why limited dependent variable.

(Refer Slide Time: 09:53)



Now, how will you model this situation? The econometric model of the Tobit is actually closely related with the Probit model. So, in Probit, what we are doing? In Probit model, if you recall, our model was yi equals to 1, when yi star greater than 0 and this is equals to 0, when yi star less than equals to 0, where yi star equals to beta xi plus ui. I am just ignoring the intercept for simplicity. So, this is how we have modeled the Probit.

But in case of Tobit what we are writing? We are saying that yi equals to yi star, which is actually equals to beta xi ui when yi star greater than 0 and equals to 0 when yi star is actually less than equals to 0. So, my yi dependent variable is actually yi star and in Tobit model this yi star equals to beta xi plus ui is actually for example is automobile let us say, automobile expenditure, where yi star is actually a latent variable.

So, this is how we model a Tobit equation, which is closely related with Probit. So, if you know Probit, since we are interested in probability yi takes only 1 and 0 value, but in Tobit since we are estimating our interest is to estimate elasticity of automobile expenditure, I need an expenditure function yi star for yi. So I am saying, yi equals to yi star when yi star greater than 0 and for 0 or negative expenditure kind of thing, we are putting 0.

So, that means in my sample while estimating, I will include only those who are having positive expenditure for this. But the question is, how will you estimate the model? Can you throw away or can you remove those households from your sample who are having zero expenditure for the

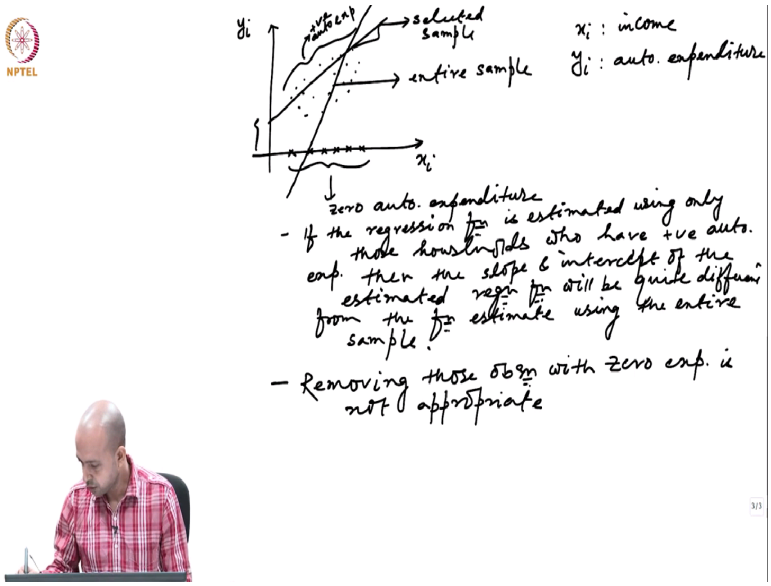
automobile and estimate the model using OLS? That is the first question that comes to our mind. So, that means how do we estimate these estimation?

And we are posing this question, can we remove those individuals with zero automobile expenditure and estimate the model using OLS with only those individuals who have positive auto expenditure? That is what we are asking. Since you have 0 observation on yi for some individuals, you cannot observe all your dependent variable. So, an easiest solution is just to remove all those individuals with zero automobile expenditure and then estimate using the data for all those households or individuals with positive automobile expenditure using the ordinary least square method, can we do so?

Now, as you know that is for estimation purpose there is absolutely no problem, but the problem happens in the next stage where we try to make inferences that means, hypothesis testing, how does it happen? See for inference making first of all estimation of coefficient is possible. Inference making is not since we require normality of ui, that mean that should follow that should follow a normal distribution with 0 mean and 1 variance or whatever sigma square let us say this is 0 mean and sigma square variance.

Whether this normality assumption is valid or not, we have to check in the sample where we have removed those observations with zero automobile expenditure. Now, first we will try to answer this question using a simple diagram. And then we will try to understand the situation more intuitively let us say.

(Refer Slide Time: 18:20)



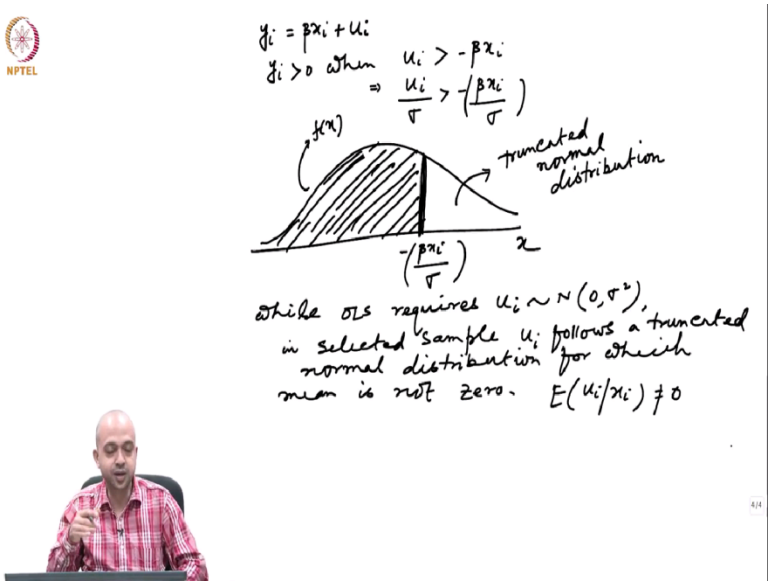
So, suppose, this is a simple diagram where here it is xi and this is yi. Xi is basically income and yi is basically automobile expenditure. So, if you plot the raw data then what will happen? For some households you have zero automobile expenditure that means they will lie in the x axis. Let us say, these are the household who have zero automobile expenditure and some household they have positive expenditure. So, these households are actually zero expenditure and these households are actually positive expenditure.

Now, if you ignore these households with zero automobile expenditure and run the regression obviously your regression would be something like this, with this intercept and this slope. But if you want to take the entire set of observation with positive as well as 0, then your regression function should be something like this, because now, you have to represent the entire data set covering the zero observation as well as the positive.

So, that means, if you remove those observations with zero automobile expenditure, then both the slope and the intercept of the estimated equation would be quite different from the equation estimated using the entire set of observations, which is quite clear intuitively from the simple diagram. So, that means, what I am saying that if the regression function is estimated using only those households who have positive auto expenditure, then the slope and intercept of the estimated regression function will be quite different from the function is estimated using the entire sample.

So, this is what you can understand intuitively, that if you remove this, this would be the observation. So, this is only for limited observation. This is the equation for selected sample and this is for entire sample. So, this is how so that means removing those observations and applying OLS. So, that means removing those observations with zero expenditure is not appropriate.

(Refer Slide Time: 24:06)



Secondly, what you can think of see what I am seeing that yi equals to beta xi plus ui, so that means, yi greater than zero, when ui greater than minus of beta xi, which says that ui by sigma is actually greater than minus of beta xi by sigma. Now in a simple diagram again, this is x and let us say this is fx. So, what I am seeing that I can observe yi only when this distribution, this is let us say minus or beta xi by sigma, this is the distribution. So, I am considering only this. So, that means this is a truncated normal distribution.

So, that means, you are trying to apply OLS in a situation where ui follows a truncated normal distribution while OLS requires ui to follow a standard normal distribution. That is why removing these set of observations and then applying OLS for the remaining is not appropriate. So, that means while OLS requires ui to follow normal distribution will let us say 0 mean and sigma square variance in selected sample ui follows a truncated normal distribution for which mean is not 0.

So, that means expectation of ui given xi, what we said is 0 is here not equals to 0. Mean is not equals to 0, because I am not considering the entire sample. I am considering only this part. So,

this is only a truncated normal distribution and where the truncation is happening? Truncation is happening here minus beta xi by sigma. Why I am dividing with sigma, just to ensure that this follows a standard normal distribution nothing else.

So, OLS is not applicable in a selected sample because of these two reasons we have just discussed. So, initially we said that when you apply OLS in the selected sample, the slope and intercept of the estimated equation are quite different from the slope and intercept of the equation you estimate using the entire sample. And secondly, ui does not follow a normal distribution which is required for OLS to be applicable rather ui follows a truncated normal distribution.