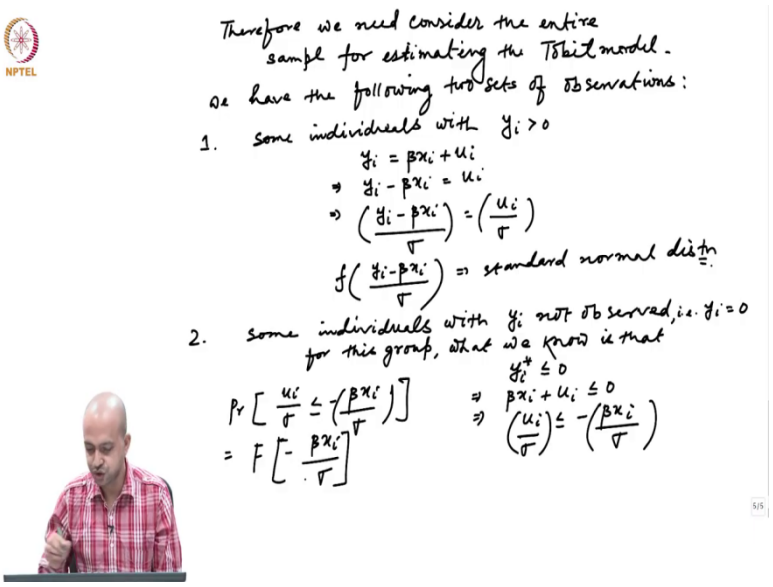
Introduction to Econometrics
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras
Lecture 63
Qualitative Response Models - Probit and Tobit Models Part - 3

(Refer Slide Time: 0:15)



So, that means we need to consider the entire sample for estimating the Tobit model. Now, how will you proceed? So, if you look we have two set of observations. First of all, some individuals with yi greater than 0. And when yi is greater than 0 positive expenditure, so that means we can easily observe the function which is yi equals to beta xi plus ui and then that implies yi minus beta xi equals to ui or yi minus beta xi by sigma equals to ui divided by sigma, this happens.

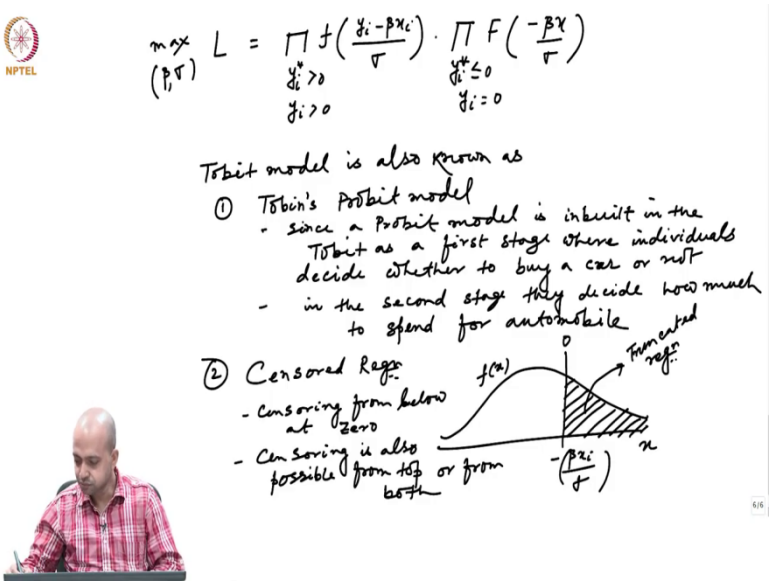
And probability of observing those households with yi greater than 0 can be represented by the standard normal distribution. So, that means, probability of yi greater than this can be represented by a standard normal distribution, which is yi minus beta xi by this. And the second set of observation, some individuals with yi is, yi not observed. That means, yi not observed meaning yi equals to 0.

For these set of people, for this group, what we know is that yi start less than equals to 0, that means or you can say that beta, this implies beta xi plus ui less than equals to 0, that means ui less than equals to minus beta xi or we can say that by, ui by sigma less than equals to beta xi by sigma. And what is the probability of this? Probability that means, in terms of probability,

probability ui by sigma less than equals to minus beta xi by sigma is actually F of minus beta xi by sigma that is what we know.

So, that means, now, what do we need to construct like the Logit and Probit model a likelihood of observing the entire sample that means, within that sample some individuals with 0 observation and how do you them in terms of probability that can be represented as a standard normal distribution because when yi greater than 0, this is nothing but yi equals to beta xi plus ui. So, after some manipulation and dividing both sides by sigma, you can say that F of yi minus beta xi by this, this is basically a standard normal distribution and this is basically the cumulative standard normal distribution.

(Refer Slide Time: 7:19)



So, if you combine these two, then you can actually write your log likelihood as some individuals for which yi star greater than 0, what you can write is yi minus beta xi by sigma and then multiplied by then for some household for which yi star less than equals to 0 that is nothing but minus beta xi by sigma.

So, this is your likelihood function which you are trying to maximize and then with respect to beta and sigma. And if you maximize this likelihood, you will estimate your beta hat and sigma hat by maximum likelihood method. Now, detailed procedure of again deriving the log likelihood and other things I am not going in detail about that. What I am trying to make you understand is how intuitively you can understand the Tobit model, wherein my objective is to estimate the elasticity.

That means, I am not removing those individuals with 0 value for the automobile expenditure rather I am trying to construct a likelihood with both sets of people, for the first set of people yi star greater than 0 or yi greater than 0, you have positive automobile expenditure and remaining is yi equals to 0 because yi star less than or equals to 0. So, here yi also greater than 0, here yi equals to 0, this is one.

This is the second set of people, this is the first set of people and probability of observing the entire sample is then the likelihood. And how will you maximize this likelihood? If you maximize this likelihood with respect to beta and sigma, then again you will get the beta hat and sigma hat which maximizes the likelihood of observing a particular sample.

Now, this Tobit model is basically known as Tobin's Probit model, because in Tobit model there is a Probit model also which is inbuilt, why this is so? Because if you think closely to observe the positive expenditure for automobile, that particular household or individual, firstly have to decide that yes, I will buy a car and then the next question is how much will you spend.

If you do not decide to buy the car, I am not able to observe your automobile expenditure that means, you can think of the Tobit model in terms of two different steps. In the first step the individual is deciding whether to buy a car or not, which is exactly like the Probit model probability of buying a car. If I pass the first stage that means, if probability of yi equals to 1 in the first stage, then in the next stage what would be my expenditure?

That is why since the Probit model is inbuilt in the Tobit model as a first stage, the Tobit model is known as Tobin's Probit model. Why this is so? Since, a Probit model is inbuilt in the Tobit as a first stage, where individuals decide whether to buy a car or not. And in the second stage they decide how much to spend for automobile. And secondly, this is also known as a censored model, censored regression. Why this is censored regression?

Because if you look at the diagram once again this is x, this is f(x) and this is minus beta xi by sigma. So, that means, I can observe a positive expenditure only after this. So, that means I am putting a censoring at 0. So, 0 or less than that I am not able to observe, if your expenditure is more than 0, then only I will be able to observe your automobile expenditure. That is why this is called a censored regression.

Now you can do censoring from the below or upper limit also. For example, let us say I am saying you are observing the efficiency of a firm, where efficiency is defined as your actual

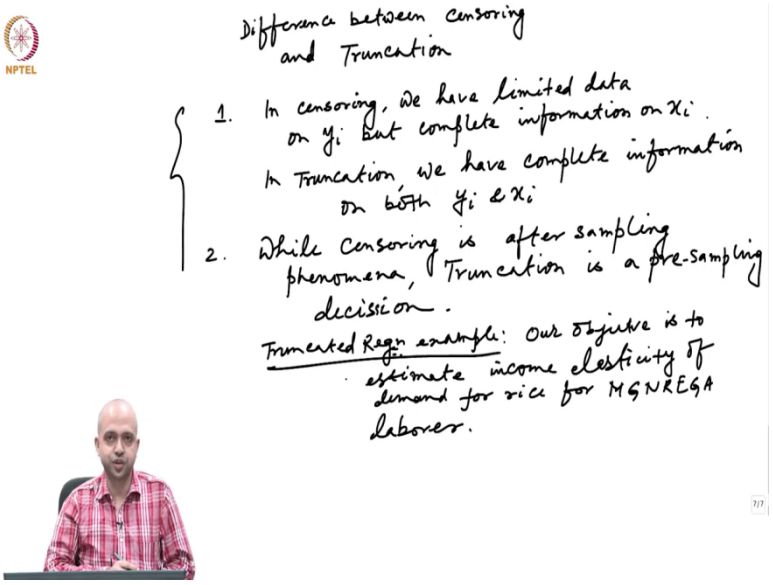
output to potential output. And in that case, the maximum of the technical efficiency of a firm value would be just 1 where actual and potential is same.

In that case, any value of technical efficiency which is more than 1 is not observable, it is meaningless. So, that means technical efficiency should be censored from top. What would be the censoring point? So, this is a censoring at 0. So, that is why this is censoring from below at 0. Censoring is also possible from top or from both.

So, in this particular automobile expenditure case, we can say that we will observe the individual also automobile expenditure if it is greater than 0 and there is no upper limit. So, in that case, it is known as censoring from the below or lower censoring or left censoring.

Now, a closely related concept is called truncated regression. what we are discussing earlier that for, if you remove this set of people with 0 or less than 0 expenditure and try to construct the model using only those individuals who have positive expenditure then here you have to apply truncated regression. ui follows a truncated normal distribution. So, that means there is a difference between censoring and truncation and what is the difference?

(Refer Slide Time: 18:13)



The first difference is in censoring we have limited data on yi which is a dependent variable, but complete information on xi, because even if you do not buy a car, I can always observe your income. So, that is why irrespective of whether you buy a car or do not buy a car, I can always observe your income, but unless you buy a car, I cannot observe your automobile expenditure. That is why in censoring we have limited data on yi but complete information on xi. But in truncation, we have complete information on both yi and xi because I have already decided not to take those into my sample. This is the first difference.

The second one is while censoring is after sampling phenomena, truncation is a pre-sampling. So, that means, when our objective is to let us say find the automobile expenditure of the car owner, in the process of collecting the sample itself, I will not ask you the question unless and until I see you have a car. So, my interest is now finding the automobile expenditure of the car or not, but what was happening in case of censoring? I am going and asking each and every individual what is your expenditure for automobile and you will respond either with actual expenditure or 0 value.

After collecting the sample, I will now put a censoring on 0. I will say that no these are the samples, not households with 0 observation, 0 expenditure. So, I will put a 0 expenditure here and then I will estimate the Tobit model. That is why while censoring is actually after sampling phenomena, truncation is a pre-sampling decision.

So, these are the 2 basic difference between censoring and truncation we need to keep in mind and if at all you want to remove those observations and consider only the selected

sample of car or not to estimate this, that means, you need to apply a specific technique of estimation which is called truncated regression, which is quite different from the Tobit model estimation using the likelihood method.

So, this is basically, these are the two differences between Tobit regression which is also known as limited dependent variable or censored regression model and this is a truncated regression model. We are not discussing the truncated regression here, we will discuss only the Tobit model or limited dependent variable or censored regressive model, but you should always keep this thing in mind.

Because sometimes your objective might be such that you have to apply only the truncated regression. For example, if your objective is let us say is to estimate income elasticity of demand for rice for MGNREGA laborer. That means, I am not going to ask each and every individual how much do you spend for rice? Rather I will focus only those individual who are MGNREGA laborer. So, that means, while collecting the sample itself it is truncated. I will take only those who are MGNREGA people. Like the previous case, I will ask only those people who have, for example, here if your income is not coming from MGNREGA, then you are not, I am not considering you. That is why income elasticity of demand for rice only for MGNREGA people.

So, there is no question of considering the entire sample, which is basically a truncated regression here. So, with this, we are just closing our discussion today. And tomorrow, what we will do, we will discuss about the estimation part of the Tobit model and we will also see how to interpret the coefficients. Thank you.