Introduction to Econometrics
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras
Lecture 65
Qualitative Response Models - Probit and Tobit Models Part - 5

(Refer Slide Time: 0:15)

After this we should also remember what are the limitations of Tobit model? There are basically two limitations of Tobit model. Firstly, there is some kind of theoretical limitation because tobit model is mostly applicable for a censored sample.

But the example what we were discussing determinants of labor supply or elasticity of labor supply or elasticity of automobile expenditure, econometricians say that, they are more of a coordinate solution rather than a proper censoring. What does it mean? It means, our theoretical structure says that yi equals to yi star, when yi star is actually yi star greater than 0 and this is 0 when yi star is actually less than equals to 0.

So, that means theoretically there is a possibility that yi star can take negative values for which we are putting a censoring, but in this case, when I am writing the labor supply function or automobile expenditure, we are saying yi star, yi equals to alpha plus beta xi also plus beta xi plus ui when this entire amount, this is actually for those when labor supply is positive and 0 otherwise.

Now, these 0 in this particular case, it does not make any sense to think that labor supply can actually take negative values as well. Same is true for automobile expenditure. So, even though we equate this situation with the censored model, they are not exactly the same. There

are cases where this yi star can actually take negative values, but that is not for this particular case what we are discussing.

So, that means, econometrician say that whenever you observe some kind of 0 value, it is not always the case of censoring, rather it might be a case of coordinate solution. What does coordinate solution means? That means if you think about our micro economics theory of labor supply, let us say in x axis I am measuring labor supply and here it is leisure.

You have two goods, leisure and labor supply, then it may so happen that individuals indifference curve is this and optimality is achieved here for which you have 0 labor supply. Now, why I am not supplying any labor in the labor market, because it may so happen that as individual labor, I might be having some kind of reservation wage or minimum wage.

And if the market wage what is offered to me is less than my reservation wage, then I will prefer to enjoy my leisure rather than participating in the labor market and supplying labor. So it is basically the 0 is happening out of purely individuals' decision making rather than thinking this is actually negative, because negative labor supply does not make any sense.

So, we need to apply a different type of model, theoretical model to represent this situation, which is not like this. So, this is the theoretical limitation, but what should be the actual model for this case and all, that is actually beyond the scope of our discussion in basic econometrics, but at least you should remember that there is something called a coordinate solution for which we observe 0, which is different from censoring, because censoring means less than or equal to 0.

So, theoretically there should be a possibility of having negative values of yi star, which is not the case in the context of labor supply in the context of automobile expenditure. Of course, we still apply Tobit model to this situation, but there are some theoretical limitations for which these type of other models have also been developed by econometrician. But we are not going to discuss in this basic econometrics course. Probably, we can discuss those advanced model in our applied econometrics course, if you are interested in.

Secondly, the second limitation is that as I said that Tobit involves two stages. In stage one, individuals decide whether to participate in labor market or not. And in the second stage, they actually decide how much labor to supply.

Now, the simple Tobit model, they assume same type of explanatory variables are involved in both the stages, that means, what are the explanatory variables that we consider? It was age,

education, experience, number of kids and husbands wage. So, what we assume here these variables decide whether they will participate in the labor market or not and the same variables also decide how much labor he will supply.

But it needs to happen in reality. In reality, there are some variables which will decide whether you will participate in the labor market or not but they do not actually decide your second stage decision making. So, that means the first stage is called participation and this stage is known as outcome.

So, same explanatory variables are involved both in participation as well as outcome which is unrealistic. It may so happen that there are some variables which will appear only in the first stage and not in the second stage. For example, number of kids may decide whether he will actually participate in the labor market or not but once you decide to participate, then that number of kids may not play a significant role in your second stage decision where actually you are thinking how much labor to supply.

This example will become more clear if your dependent variable becomes wage. For example, the wage is visible, I can observe your wage only when you have a positive amount of labor supply. So, what are the factors that will decide whether you will participate in the labor market or not? It is basically how many kids you have less than 6 years of age and what is your husband's wage.

But do you think once you participate in the labor force, the employer will decide your wage based on how many kids you have and what is your husband's wage? Quite unlikely, that is why involving same set of explanatory variables both for participation as well as outcome makes Tobit application problematic. For which to overcome this, so they say that same set of explanatory variables are involved in both participation and outcome. This is the limitation.

To overcome this problem, famous econometrician Heckman, he came up and gave one solution to this, he designed a model where some variable will appear only in the participation, but those variables would be excluded in the second stage where the individuals are deciding about the labor force or by the end the individuals are getting their wages and all and that is called Heckman's selection model.

So, Heckman selects and model is basically an improvement over simple Tobit. So, this is an improvement over simple Tobit model and once again this Heckman selection model is again beyond the scope of my discussion of basic econometrics. So, that also comes in a little

advanced course, which is applied econometrics, but you should always keep this thing in mind that simple Tobit model has this type of limitation.

First of all, Tobit model assumes all cases are censored, they do not differentiate between coordinate solution we have discussed, coordinate solution with censoring. A proper censoring requires theoretical possibility of yi star to become less than 0. That is actually not the case here, when you are estimating wage or labor supply or automobile expenditure.

But still it is applied we see many people are actually using Tobit model even these contexts as well. And secondly, Tobit model assumes that both the stages participation and outcome are determined by same set of explanatory variable while in reality that may not be true always. We have some variable which will appear only in the participation deciding whether the married woman will participate in the labor force or not, but not in the second stage where the actual labor supply decision making is made.

Now, this is basically the limitations of Tobit model and with this we are closing our discussion on Tobit. And before we wind up our discussion of this portion of probabilistic model and limited dependent variable model, let us just revise what we have discussed.

(Refer Slide Time: 15:24)

So, in LPM first we discussed about linear probability model. There, our model was just like the OLS yi equals to which is basically yi is nothing but the probability of labor force participation alpha plus beta xi plus ui and this yi is nothing but your actually Pi, what is the probability that individual will own a house or not. And then, we say that this linear

characterization of probability between income and within income is problematic because probability does not behave linearly with the explanatory variable.

And we come came up with the Logit model where the probability is 1 by 1 plus e to the power minus z and this z actually belongs to minus infinity to plus infinity. And when z belongs to minus infinity to plus infinity, this P actually belongs to 0 and 1 which was not the case in the context of linear probability model.

Then, we have also discussed about Probit. Probit is also like the Logit one, the only change is that the density function, the CDF or do we assume cumulative density function which is like logistic here, we assume that in the context of Probit, it is like minus infinity to alpha plus beta xi this is fz.dz, where z is basically, where zi is basically 1 by root 2 pi into sigma squared e to the power minus zi square by 2 and zi basically follows a standard normal.

So, this is basically minus zi by 2 and zi equals to z minus mu by sigma whole square and this zi 0 mean and 1 variance. So, this is the Probit model that we have discussed and then we say that while these models have their objective of estimating probability, we have the fourth model that we have discussed that is Tobit where our objective is not to estimate the probability but to estimate the direct elasticity.

We said that yi equals to yi star, when yi star is actually positive and 0 when yi star is either negative or 0. This is also called censored model, which is different from the truncated one. We have also introduced the concept of truncated model, but truncated regression model is not within the scope of our discussion here.

So, if you throw away the observations with 0 labor supply or 0 expenditure for automobiles, then probably you are dealing with a truncated sample, where the error term follows a truncated normal distribution. And to estimate that type of model, you need to have a truncated regression model, that also we have introduced, but we have not discussed in detail.

There are several extensions of this probabilistic model that is possible. For example, here in the context of simple Logit and Probit what we assume that the dependent variable is binary that means, either you buy a car or do not buy a car, but in reality there are situations where you might be interested in whether to buy a sedan, whether to buy a hatchback, whether to buy an SUV or do not buy anything.

Suppose my yi actually takes this type of scenario; no car then this is let us say sedan, then this is let us say hatchback and let us say this is SUV. So, that means, this is a situation where

dependent variable instead of binary, instead of a dichotomous choice, now this becomes a polychotomous choice. And to represent this we need to extend the simple Logit and Probit model into three types of model again; multinomial logit, conditional logit or Probit and mixed.

To represent this type of situation, we need to explain the simple Logit and Probit model into multinomial, conditional and mixed logit we need to understand. So, these three models, again is beyond the scope of my basic econometrics, that comes in the advanced or next level course of applied econometrics. Then there are situations, here when I am talking about no car, sedan, hatchback and SUV, we do not distinguish between these options that means there is no order of preference, so individual can take, any select any of this.

But suppose, we are ordering these options, where we say that no car situation with lowest order then a little better situation is having a hatchback then sedan is even better than hatchback an SUV is the best that you can get. So, that means I can see that SUV is greater than sedan than hatchback than no car situation. This is my order of preference.

When your dependent variable can be ordered in this way, many a times we face questionnaire where the options are like agree, somewhat agree, strongly agree, not at all agree. So that means these options are actually ordered, where we say that strongly agree is better than somewhat agree which is better than not agree. So when a dependent variable can be ordered in this way, then we need to have ordered logit ordered Probit type of model. So indeed require either ordered logit or ordered Probit, this is also some kind of extension of your simple Logit and Probit model.

So, with this we are closing our discussion on this probabilistic model and limited dependent variable model. And as I said there are different ways by which we can extend our discussion to have advanced level of econometric models and that comes under the scope of applied econometrics. So, the objective of my, this course was to only introduce the course of econometrics to you, who are just a beginner of this.

So, that is why we made our analysis very simple, because our objective is only to introduce and once you get the basic foundation of the subject, then the only there is a possibility of extending this course and having a discussion of more advanced model. For example, in this discussion of basic econometrics throughout the course, we dealt only with cross sectional data, but it is possible to have two other types of data set also; panel data and time series data.

When I am saying yi equals to alpha plus beta xi plus ui, we assume that xi's are non-stochastic, but there are situations when xi's can become stochastic, that means, endogenous in nature due to simultaneity.

So, simultaneous equation model, if we introduce then there are other related discussion on instrumental variable estimation, all those things come under the scope of advanced econometric models. But, at best what I assume that my objective if you can understand the basics of this course, if your foundation is clear, then there is absolutely no problem in understanding the advanced courses.

If you understand these basics, then my objective is fulfilled. And with this, I just I would like to wind up our discussion on basic econometrics, wishing you all the best in your learning. And if you find this course interesting, then probably you can go for our next level of applied econometrics course. Thank you.