Applied Econometrics

Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology-Madras

Lecture - 29
Course outline for Applied Econometrics

Qualitative Response Model-Part I

Welcome once again to our discussion of Econometrics and today we are going to discuss about a specific type of econometric model which is also very interesting and which has very interesting applications also. Many a times we get that type of situation where these models what we are going to discuss from today onwards these 3-4 models are going to be extremely useful. First of all the name of this class of models they are known as Qualitative Response Model. Or sometimes they are also known as Dummy Dependent sorry Dummy Dependent Variable Model or sometimes they are known as Binary Response Model. All are same Binary Response Model. Now why this is called Qualitative Response or Binary Response Model because the dependent variable the other name of the dependent variable is response.

So, let us say that $yi = \alpha + \beta xi + ui$. Now in the context of dummy variable what we discussed sometimes our independent variables this xi may become qualitative in nature and we were discussing about gender, caste, PhD, non-PhD, so on and so forth about this xi independent variable. So, when independent variable is qualitative in nature we said that we have to convert this qualitative information into a quantitative one using the dummy variable approach. Now the same dummy variable we can apply in this context also when your dependent variable is qualitative in nature.

That is why the name called Dummy Dependent Variable Model, Qualitative Response Model or Binary Response Model. Why it is called Binary Response Model? Because your response variable yi will take two values, y take two values. I will give you some example. Let us give some example. Example number one, let us say that our research question is why do some individual, own their car while others prefer public transport? So, you are going to explain the factors that can explain the car ownership.

So, when you go to individuals you will ask do you own a car they will say either yes or no. So, this is a qualitative information. So, that yes or no information you have to translate into a quantitative format by assigning let us say 1 for yes and 0 for no. So, that means y equals to 1 indicates the ith individual is having a car, y equals to 0 indicates the household does not have a car. Example number two, why do some individuals own their house while others prefer to stay at rented apartment? This is another question, this is another question that you might be interested.

Example number three, suppose several individuals have applied for loan, some of the individuals loan got approved and some individuals loan got rejected and we want to know what are the factors that can determine whether an individual s loan will get approved or not. So, then basically you will ask the individuals whether your loan got approved they will say either yes or no and you have to assign 1 for yes, 0 for no that means again you are converting that quantitative information, qualitative information into quantitative one using the dummy variable. But in all these cases, in all these cases the qualitative information is only for the dependent variable and that you have to regress with what is the collateral that household is having, what is the monthly income that the household is having, what is the dependency ratio, what is education, sex, gender, so on and so forth. With all these factors you are going to explain whether the individual, what is individual s loan will get approved or not. So, basically whether here the research question is whether the individuals loan got approved or not, this is the question.

So, here that means what I am saying that your yi, yi they can take only two value, yi equals to 1 if having a car, let us say that this is the car ownership problem, if having a car 0 otherwise. Now, let us also assume that probability, probability yi equals to 1 is denoted as Pi and probability yi equals to 0 that is denoted as 1 - Pi. Let us say this is, this is equation 1, this is 2, this is 3. Now, if I take expectation of equation 1, then what I can write $E(\frac{yi}{xi})$ = α + βxi – (4). Now, I can find the expectation of yi from this formula also, because yi can take two values 1 and 0 and the probability that y will take value 1 is Pi and 0 as 1 - Pi.

So, this is 0 and 1, these are the values that y can take and the probability is Pi and 1 - Pi, 1 - Pi. So, these are the two values y can take. So, E(yi) from here what I can write

expectation of yi equals to, sorry, this is 1, this is 0. E(yi)= Pi*1+(1-Pi)*0, = Pi. (5).

Now, combining 4 and 5, then what I can write that Pi actually E(yi/xi) = Pi and that again equals to α + βxi. So, that means this implies Pi = α + βxi –(6). Now, why this model? This model what I have, this is a probabilistic model that we have derived. So, that means when I am saying $E(\frac{yi}{xi})$ = α + βxi = Pi, this model is known as linear probability model, linear probability model. So, that is the first model in this class of models, that means linear probability model is the first kind of model of the binary response models.

We have many other models, but this is the starting point, Pi = to α+βxi. Now, this is called linear probability model. Why this is called linear probability model? That means this is in short I will say LPM and why this is called LPM? There are two reasons. First of all, unlike other cases here E(yi/xi), xi denotes actually that means or conditional probability, conditional probability of sorry conditional mean, conditional mean of yi basically indicates probability of owning a car. So, here the conditional mean of yi E(yi/xi), they actually indicates a probability.

When we are talking about yi = α + βxi + ui in the context of consumption function, their $E(\frac{yi}{xi})$ = α + βxi was denoting only the mean income, but here it is a conditional, it is a probability. Conditional mean of yi indicates probability of owning a car and secondly, that probability is a linear function of x, that probability is linear in x. Because of these two reasons, this model is called linear probability model, that is all, linear probability model. Now, this linear probability model it has, though it is the starting point of this quantitative response model, it has some limitations. What are those? What are the limitations of LPM can you think of those.

The first one is, as you know from the properties of probability that pi should always lie between 0 and 1. Pi should always lie between 0 and 1, but that implies E(yi/xi) should also lie between 0 and 1 and that implies that α + βxi should lie between 0 and 1. But as you can see, suppose this xi denotes income, that means we are trying to understand the probability that a particular household will own a car from that household's income. Since this is a linear function, as income increases, probability of owning a car will also increase. But as you can think of, let us say income is increasing from 40,000 to 50,000, there would be some increase in the probability.

Then again 50,000 to 75,000, another increase in the probability of owning a car. Then 75,000 to 1 lakh, 1 lakh to 1.5 lakhs, 1.5 lakhs to 2 lakhs, 2 lakhs to 2.

5 lakhs. So, the probability of owning a car will keep on increasing as x increases since it is a linear probability. So, it may so happen that at some point of time your probability will go beyond 1, since you are calculating probability with a linear function. So, that is why there is no guarantee that this pi or E(yi/xi) will always lie between 0 and 1. But there is no guarantee that pi or $E(\frac{yi}{xi})$ = α + βxi, they will lie between 0 and 1. And if that is the case, that means you are actually violating the properties of probability.

So, you may, it may so happen that your estimated probability is 1.56, which does not make any sense, which does not make any sense. So, that is the limitation of linear probability model. And then secondly, can we estimate the model Pi = α + βxi using OLS, can we estimate the model? When I am saying that the $E(\frac{yi}{xi})$ = α + βxi, can we estimate the model using OLS? What will happen if we estimate the model using OLS? So, can we estimate LPM using OLS? That is also we need to think about. Now, E(yi), that means the model what you are estimating Yi = α+βxi+ui and y will take only 1 and 0.

So, that means depending, so from here we can say that ui = yi – α-βxi ui. So, equals to either 1 – α-βxi or equals to - α-βxi, when yi equals to 1, when yi equals to 0. So, that means here you see ui can take only two value, what would be the distribution of then ui? So, ui that means this will say the distribution of ui will be discrete instead of normal. Now,if ui follows a discrete distribution, we cannot go for hypothesis testing as you know, because for that we need the normality assumption of ui. Otherwise, we cannot construct the test statistic for conducting hypothesis testing.

So, this is another problem of linear program, of linear probability model that first of all there is no guarantee that this will lie between 0 and 1 the probability then secondly we cannot estimate this model using OLS method, because ui takes only two values depending on what y takes. When y equals to 1, ui equals to 1 – α- βxi and equals to – α-βxi when yi equals to 0. So, ui follows a discrete distribution. So, this is the problem and to overcome, to overcome this econometrician they developed another model which is called logit model. So, here instead of assuming probability is a linear function of x what this model assumes that probability pi which is actually probability yi takes the value 1, Pi = $(1/1 + \mathrm{e})^{-Zi}$ where zi = α+βxi ,Zxi.

Now, from here what you can do that this model apparently looks like a non-linear model, this looks like a non-linear model but you can always linearize this model. How you can do that? If you take 1 minus, if you take Pi,(Pi/1-Pi) that would become e^{zi} . And then if you take log of this then log(Pi/1-Pi) = zi = α+βxi , and then you can estimate this model because now this model becomes a linear model. So, you can add the error term here and then that is basically the estimable function. So, this mathematical model you can convert into statistical one by adding the error term and this particular specification you can now estimate.

So, you can estimate this model. We can estimate this model. But even in this model also what is your dependent variable? Dependent variable is log(Pi/1-Pi) and this (Pi/1-Pi) it has a specific name. What is the name? The name is called odds ratio, odds ratio. pi since this is pi is probability of owning a car or probability of owning a house, (Pi/1-Pi), is called odds in favor of happening the event or odds in favor of owning a car.

Since the numerator is pi we will say that this is odds in favor of owning a car. If we calculate (1-Pi/Pi) that is also odds ratio, but then that will indicate odds against happening that event. So, here (Pi/1-Pi)   is your dependent variable and we can, you have take log of this. Now, once if you think of estimating this model, see how will you estimate this model? You have information on y, and y can take two values, yi equals to either 1 or 0, this is y, this is yi. yi equals to 1 and which is basically the indicator indicates the individuals is owning a house and this is indicating not owning a house that means what is pi and pi is probability yi equals to 1 and 1 - pi is basically probability yi equals to 0.

Now, to estimate this model, this is let us say model, 6. To estimate this model, first of all you need to have the information on the dependent variable. So, we do not have information on pi rather we have information on yi. Now, apparently you may think of you can put yi value in this equation and you will construct the dependent variable. Now, if you put 1 and 0 here, what will happen? If you put 1 and 0 putting 1 0 in equation 1, in equation 6, what you will get? You will get, so if you put 1, then that would become log of 1 by, sorry, log of 1 by 1 minus 1, so that would become 1 by 0, so that means log of, this would become your dependent variable.

And if you put 0 here, then what will happen? This would become log of 0 by 1 minus 0 equals to log of 0. So, this would become your dependent variable. Can we estimate this model? No. That means we cannot actually put 1 and 0 in this equation, why? Because see here we are thinking that pi is actually equivalent to yi, but that is actually not the case. Pi indicates the probability which will lie between 0 and 1, but here what you are, what you observe is the realization, yi is the realization, some people they have owned the car that is why you have put 1, some people they do not have car that is 0, so that is the realized thing.

But what you are thinking of the probability which is unobserved, probability of owning a car is not observed, rather what you observe is actually the ultimate realization, whether the person has taken or not, that is the decision. After taking the decision what we observe is a realized fact, but what we are thinking here in terms of this model 6 is log (Pi/1-Pi)   which is not actually observable. So, what we can say that pi is actually not equivalent to yi. So, that is why in this model we cannot estimate using the OLS method, because the dependent variable itself we do not have information. We cannot put 1 and 0 here and we cannot estimate the model.

So, that means OLS is not applicable to estimate the logit model. We need to go for a different route.