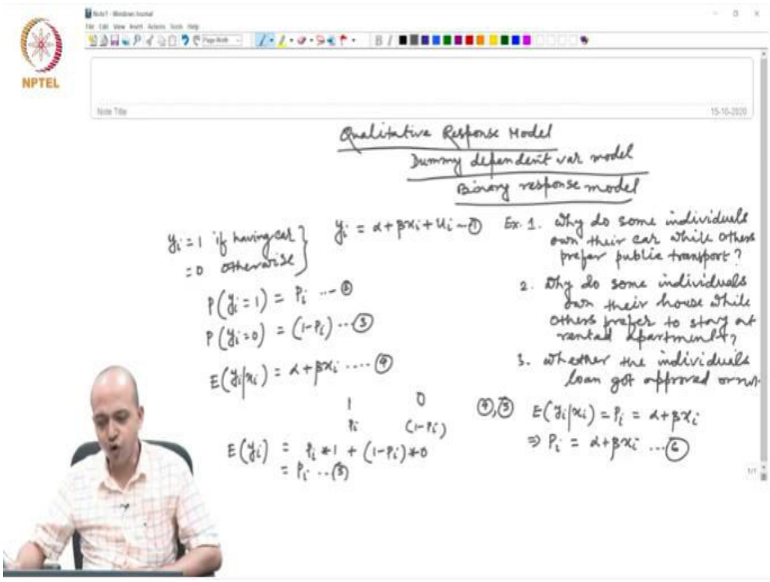
Environmental and Resource Economics
Professor Sabuj Kumar Mandal
Department of Humanities and Social sciences
Indian Institute of Technology Madras
Lecture 40
Qualitative Response Models- Linear Probability Model, Logit and Probit Models Part -1

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 times we get that type of situation where these models, what we are going to discuss from today onwards, these three four models are going to be extremely useful.

(Refer Slide Time: 0:57)



First of all, the name of this class of models, they are known as qualitative response model, 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, y i equals to alpha plus beta x i plus u i. Now in the context of dummy variable, what we discussed, sometimes our independent variables, this x i may become qualitative in nature and we were discussing about gender, caste, PhD, Non-PhD, so on and so forth about this x i 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 y i will take two values. I will give you some example right, let us give some example.

Example number 1; let us say that our research question is, why do some individuals own their car, while others prefer public transport? So you are going to explain the factors, 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 2; 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 3; 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 individuals 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 sign 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, 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? Education, sex, genders so on and so forth. With all these factors, you are going to explain whether the individual, what is individuals loan will get approved or not?

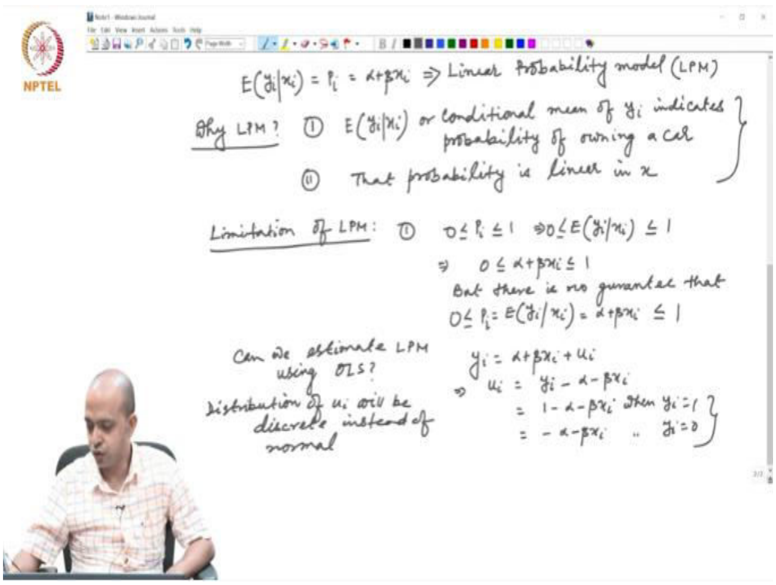
So basically, whether, here the research question is whether the individuals loan got approved or not. This is the question, so here that means what I am saying that your y i; y i they can take only two value y i equals to 1 if having a car, let us say that this is the car ownership problem. If having a car 1, otherwise.

Now, let us also assume that probability y i equals to 1 is denoted as p i and probability y i equals to 0, that is denoted as 1 minus p i. 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 expectation of y i given x i equals to alpha plus beta x i. Let us say that is equation 4.

Now I can find the expectation of y i from this formula also because y i can take two values: 1 and 0 and the probability that y will take value 1 is p i and 0 as 1 minus p i. So this is 0 and 1, these are the values that y can take and the probability is p i and 1 minus p i. So these are the two values y can take, so expectation of y i from here what I can write expectation of y i equals to sorry this is 1, this is 0.

So p i into 1 plus 1 minus p i into 0 equals to p i. Let us say this is equation 5. This is equation 5. Now combining 4 and 5, combining 4 and 5; 4 and 5 if I combine, then what I can write that p i actually expectation of y i given x i equals to p i and that again equals to alpha plus beta x i, so that means this implies p i equals to alpha plus beta x i this is, let us say equation 6.

(Refer Slide Time: 11:33)



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 expectation of y i given x i equals to p i equals to alpha plus beta x i. This model is known as linear probability model. So that is the first model in these 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 p i equals to alpha plus beta x i. 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? Why this is called LPM? There are two reason, first of all, unlike other cases here expectation of y i given x i, denotes actually that means or conditional probability of, sorry conditional mean of y i basically indicates probability of owning a car.

So here the conditional mean of y i expectation of y i given x i they actually indicates a probability, when we are talking about y i equals to alpha plus beta x i plus e y in the context of consumption function, there expectation of y i given x i or alpha plus beta x i was denoting the only the mean income, but here it is a conditional, it is a probability.

Conditional mean of y i 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? What are the limitations of LPM? The first one is, as you know, from the properties of probability that p i should always lie between 0 and 1, p i should always lie between 0 and 1, but that implies expectation of y i given x i should also lie between 0 and 1 and that implies that sorry, and that implies that alpha plus beta x i should lie between 0 and 1.

But as you can see suppose this x i 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 will, probability of owning a car will also increase. But as you can think of let us say, income is increasing from 40000 to 50000, there would be some increase in the probability.

Then again, 50000 to 75000, another increase in the probability of owning a car, then 75000 to 100000, 100000 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 p i or expectation of y i given x i will always lie between 0 and 1. But there is no guarantee that p i or expectation of y i given alpha plus beta x i, 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. So that is the limitation of linear probability.

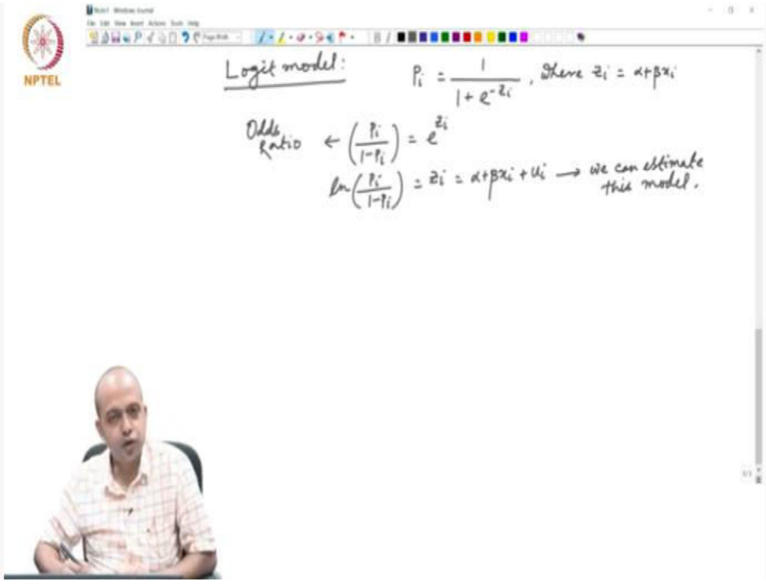
And then, secondly, can we estimate the model p i equals to alpha plus beta x i using OLS? Can we estimate the model? When I am saying that the expectation of y i given x i equals to alpha plus beta x i, can we estimate the model using OLS? What will happen if we estimate the model using OLS? So can we estimate LPM using OLS? That is also we need to think about.

Now expectation of y i that means the model what you are estimating y i equals to alpha plus beta x i plus u i and y will take only 1 and 0, so that means, depending, so from here we can say that, u i equals to y i minus alpha minus beta x i. So equals to either 1 minus alpha minus beta x i or equals to minus alpha minus beta x i, when y i equals to 1, when y i equals to 0 right. So that means here you see, u i can take only two value.

What would be the distribution of then u i? So u i that means this will say the distribution of u i will be discrete, instead of normal. Now, if u i, follows a discrete distribution, we cannot go for hypothesis testing as you know, because for that we need the normality assumption of u i. Otherwise we cannot construct the test statistic for conducting hypothesis testing. So this is another problem 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, u i takes only two values depending on what y takes, when y equals to 1, u i equals to 1 minus alpha, minus beta x i and equals to minus alpha minus beta x i when y i equals to 0. So u i follows a discrete distribution. So this is the problem. This is the problem and to overcome this, econometrician they developed another model which is called logit model.

(Refer Slide Time: 22:42)



So here, instead of assuming probability is a linear function of x, what this model assumes that probability p i, which is actually probability y i takes the value 1 p i equals to 1 by 1 plus e to the power minus z i, where z i equals to alpha plus beta x i. 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 p i, p i by 1 minus p i that would become e to the power z i. And then, if you take log of this, then log of p i by 1 minus p i equals to z i equals to alpha plus beta x and then you can estimate this model, because now this model becomes a non-linear model, sorry 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 of p i by 1 minus p i and this p i by 1 minus p i it has a specific name. What is the name? The name is called odds ratio, odds ratio.