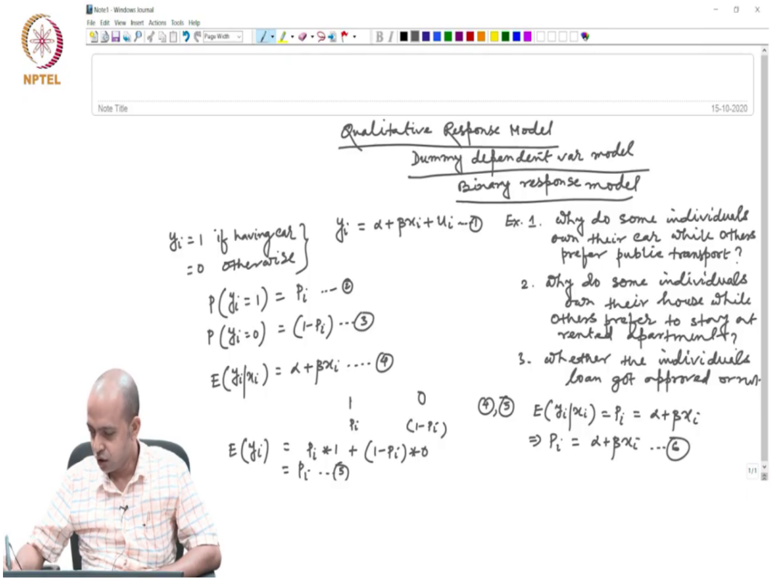
Introduction to Econometrics
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras
Lecture 56
Qualitative Response Models-Linear Probability Model, Logit and Probit Models Part 1

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.

(Refer Slide Time: 01:00)



Qualitative Response Model

Dummy dependent var model

Binary response model

$y_i = 1$ if having car
$= 0$ otherwise

$y_i = \alpha + \beta x_i + u_i$ --(1)

$P(y_i = 1) = P_i$ --(2)

$P(y_i = 0) = (1 - P_i)$ --(3)

$E(y_i | x_i) = \alpha + \beta x_i$ ....(4)

$$\begin{array}{ccc} & 1 & 0 \\ & P_i & (1-P_i) \end{array}$$

$E(y_i) = P_i * 1 + (1 - P_i) * 0$
$= P_i$ --(5)

Ex. 1. Why do some individuals own their car while others prefer public transport?

2. Why do some individuals own their house while others prefer to stay at rental apartment?

3. whether the individuals loan got approved or not

(4),(5) $E(y_i | x_i) = P_i = \alpha + \beta x_i$

$\Rightarrow P_i = \alpha + \beta x_i$ ...(6)

First of all, the name of this class of models are known as Qualitative Response Model or sometimes they are also known as Dummy dependent variable model or sometimes they are known as Binary response model. Now, this is called Qualitative Response or Binary response model because the other name of the dependent variable is response.

So, let us say that yi equals to alpha plus beta xi plus 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, cost, PhD, non PhD so forth about these 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 is called Dummy dependent variable model, Qualitative Response model or Binary response model. It is called Binary response model because your response variable yi will take 2 values. I will give you 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 that can explain the car or 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 i^{th} individual is having a car y equals to 0 indicates the household does not have a car.

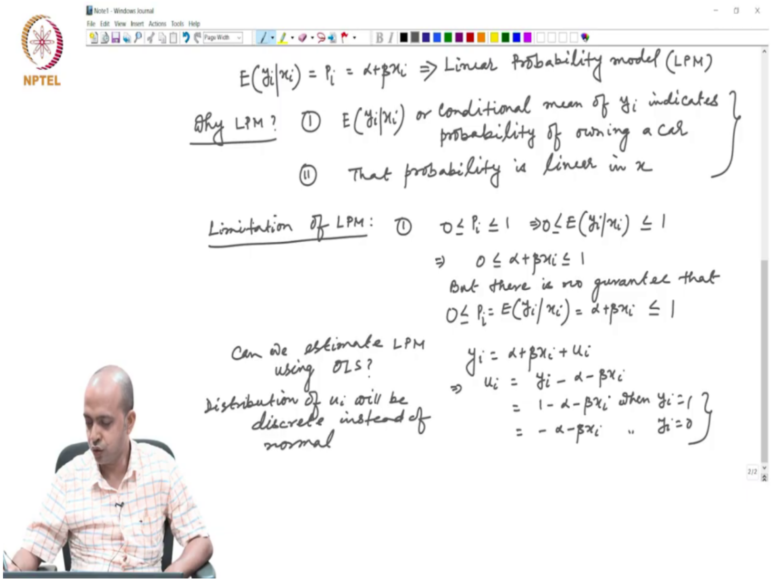
Example number 2, why do some individuals own their house, while others prefer to stay at rented apartment, this is another question that you might be interested. Example number 3, suppose several individuals have applied for loan some of the individuals loan got approved and some individuals loan got rejected and we want to know 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 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 or education, sex, gender, so on and so forth, with all these factors you are going to explain whether the individual's loan will get approved or not. So, basically here the research question is whether individual's loan got approved or not.

So, here that means, what I am saying that yi can take only 2 values. yi equals to 1 if having a car and 0 otherwise. Now, let us also assume that probability yi equals to 1 is denoted as pi and probability yi equals to 0 is denoted as (1-pi). Let us say this is equation 1, this is 2, this is 3.

Now, if I take expectation of equation 1, now I can write expectation of yi given xi equals to alpha plus beta xi. That is equation 4. Now I can find the expectation of yi from this formula also because yi can take 2 values 1 and 0 and the probability that y will take well 1 is pi and 0 is (1-pi). So, expectation of yi equals 2, this is 1 this is 0, so pi into 1 plus 1 minus pi into 0 equals 2 pi let us say this is equation 5. Now, combining 4 and 5 if I combine then what I can write that pi actually, expectation of yi given xi equals to pi and that again equals to alpha plus beta xi, so that means this implies pi equals to alpha plus beta xi this is let us say equation 6.

(Refer Slide Time: 11:34)



$E(y_i|x_i) = P_i = \alpha+\beta x_i \Rightarrow$ Linear Probability model (LPM)

Why LPM ? ① $E(y_i|x_i)$ or conditional mean of y_i indicates probability of owning a car

⑪ That probability is linear in x

Limitation of LPM : ① $0 \leq P_i \leq 1 \Rightarrow 0 \leq E(y_i/x_i) \leq 1$

$\Rightarrow 0 \leq \alpha+\beta x_i \leq 1$

But there is no gurantee that

$0 \leq P_i = E(y_i/x_i) = \alpha+\beta x_i \leq 1$

Can we estimate LPM using OLS?

Distribution of u_i will be discrete instead of normal

$y_i = \alpha+\beta x_i + u_i$

$$\Rightarrow u_i = y_i - \alpha - \beta x_i \\ \left.\begin{aligned} &= 1-\alpha-\beta x_i \text{ when } y_i = 1 \\ &= -\alpha-\beta x_i \quad \text{"} \quad y_i = 0 \end{aligned}\right\}$$

Now, this is a probabilistic model that we have derived. So that means when I am saying expectation of yi given xi equals 2 pi equals to alpha plus beta xi this model is known as 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 equals to alpha plus beta xi.

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, like other cases here expectation of yi given xi denotes actually the means or conditional mean of yi basically indicates probability of owning a car.

So, here the conditional mean of yi, expectation of yi given xi, they actually indicates a probability when we are talking about y equals to alpha plus beta xi plus ui in the context of consumption function, their expectation of yi given Xi or alpha plus beta xi was denoting the 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 xi, because of these two reasons, this model is called linear probability model.

Now, this linear probability model is the starting point of this Qualitative Response model, it has some limitations. The first one is as you know, from the properties of probability that pi should always lie between 0 and 1, but that implies expectation of yi given xi should also lie between 0 and 1 and that implies that alpha plus beta xi 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 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 will 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 expectation of yi given xi will always lie between 0 and 1, and if that is the case that means you are actually violating the properties of probability. So it may so happen that your estimated probability is 1.56 which does not make any sense.

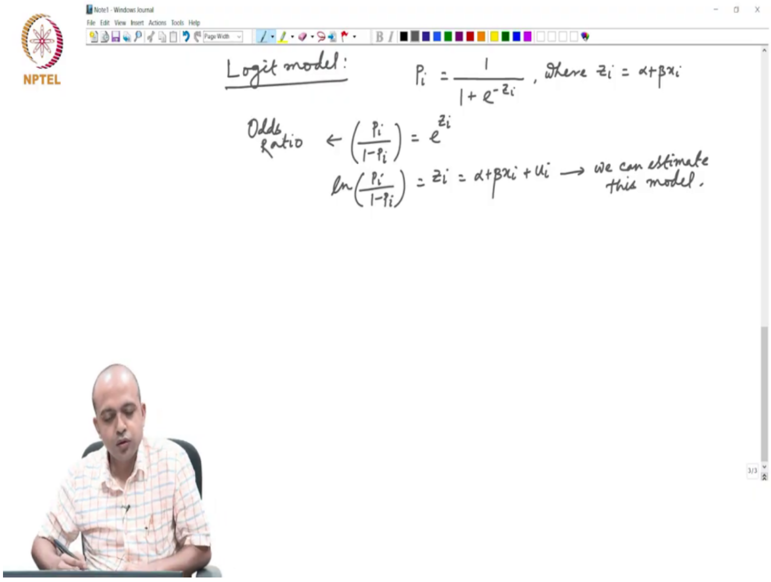
So, that is the limitation of linear probability model and then secondly, can we estimate the model pi equals to alpha plus beta xi using OLS can we estimate the mode., When I am saying that expectation of yi given xi equals to alpha plus beta xi, can we estimate the model using OLS, what will happen if we estimate the model using OLS?

yi equals to alpha plus beta xi plus ui and y will take only 1 and 0, so that means depending so from here we can say that ui equals to yi minus alpha minus beta xi, so equals to either 1 minus alpha minus beta xi or equals to minus alpha minus beta xi when yi equals to 1, when yi equals to 0.

So, that means here you see ui can take only two values or do not be the distribution of then ui, so ui that means, this will say that 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 probability model that first of all, there is no guarantee that the probability will lie between 0 and 1 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 minus alpha minus beta xi and equals to minus alpha minus beta xi when yi equals to 0. So, Ui follows a discrete distribution.

(Refer Slide Time: 22:42)



Logit model:

$P_i = \frac{1}{1+e^{-z_i}}$, where $z_i = \alpha+\beta x_i$

Odds Ratio $\leftarrow \left(\frac{p_i}{1-p_i}\right) = e^{z_i}$

$\ln\left(\frac{p_i}{1-p_i}\right) = z_i = \alpha+\beta x_i + u_i \longrightarrow$ we can estimate this model.

So, this is the problem and to overcome this, econometrician developed another model which is called Logit model. So, here instead of assuming probability is a linear function of x, this model assumes that probability pi which is actually probability yi takes the value 1, pi equals to 1 by 1 plus e to the power minus zi, where zi equals to alpha plus beta xi. This model apparently looks like a non-linear model but you can always linearize this model.

How you can do that? If you take pi by 1 minus pi that would become e to the power zi and then if you take a log of this, then log of pi by 1 minus pi equals to zi equals to alpha plus beta xi 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. But, even in this model also, what is your dependent variable? Dependent variable is log of pi by 1 minus pi and this pi by 1 minus pi has a specific name called odds ratio.