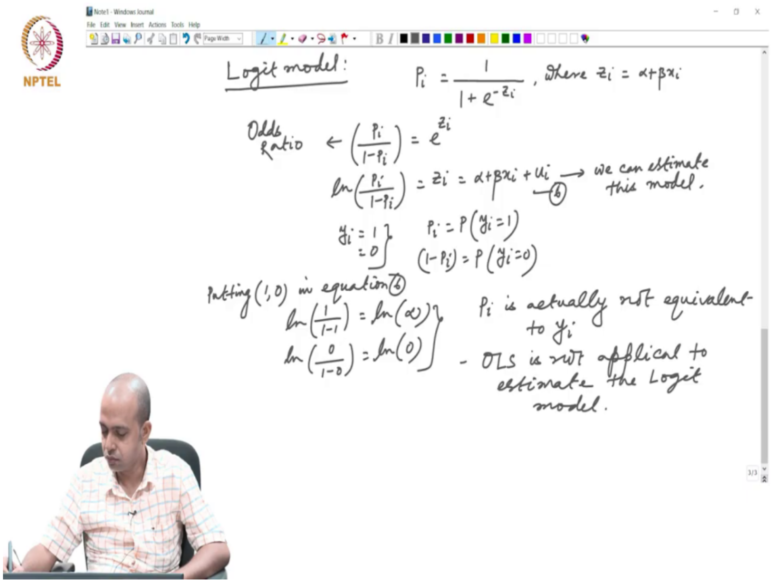
Introduction to Econometrics
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras
Lecture 57
Qualitative Response Models-Linear Probability Model, Logit and Probit Models Part 2
(Refer Slide Time: 00:14)



pi is probability of owning a car or probability of owning a house, pi by 1 minus pi is called odds in favour of happening the event or odds in favour of owning a car since the numerator is pi will say that this is odds in favour of owning a car. If we calculate 1 minus pi by pi, that is also odds ratio but then that will indicate odds against happening that event so here pi by 1 minus pi is your dependent variable and we can take log of this.

Now, once if you think of estimating this model you have information on y and y can take two values yi equals to either 1 or 0. yi equals to 1 and which basically indicates the individual is owning a house and this is indicating not owning a house that means pi is probability yi equals to 1 and 1 minus pi is basically probability yi equals to 0.

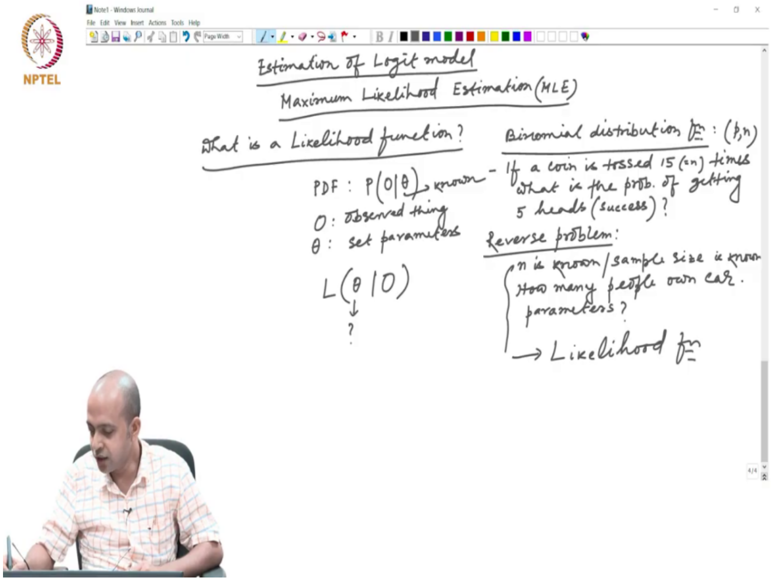
Now, to estimate this model, first of all you need to have information on the dependent variable, so we do not have information on pi rather we have information on yi. Now, apparently you may think of you can put yi value in this equation and you will construct the dependent variable.

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, so if you put 1 then that would become 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 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 y because see here, what 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 observed is the realisation yi is a realisation some people they have owned the car, that is why you have put 1 some people they do not have got that is 0, so that is the realised thing, but probability of owning a car is not observed rather what you observe is actually the ultimate realisation whether the person has taken or not that is the decision.

After taking the decision, what we observe is a realised fact but what we are thinking here in terms of this model 6 is log of pi by 1 minus pi, which is not actually observable, so what do we can say that pi is actually not equivalent to yi so that is why 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.

(Refer Slide Time: 06:07)



So, we are going to discuss estimation of Logit model using maximum likelihood function or maximum likelihood estimate. So let us say I am writing maximum likelihood estimation which in short known as a MLE. Before we talk about maximum likelihood estimation we need to know what is a likelihood function. This is very important likelihood function.

The concept of likelihood function is actually closely related with probability distribution or probability density function. So if you know probability distribution function or probability density function then you can easily understand the concept of likelihood function.

Now, what is a probability distribution function in any probability distribution function if you think of let us say binomial, pi show or any other type of probability distribution function, suppose you are thinking of the simple probability distribution function the binomial function and the probability distribution function how it is characterised, let us am writing PDF, Probability Distribution or Probability Density Function, PDF is characterised by a observed thing given the parameters, I will explain this, when I am specifying a particular Probability Density Distribution Function lets a binomial I can find out from there.

If I toss a coin, let us say 15 times, what is the probability of having, let us say, 5 heads? so let us say I am thinking of binomial distribution function, which is given by 2 parameters in n p, so I can answer this type of question if a coin is tossed n times that means, let us say 15 times, what is the probability of getting 5 heads. let us this getting a head we are calling it this is success, that you can find out if the probability distribution function is given now, if I reverse this question, so if I reverse this question here the parameters are known and you are trying to find out the probability of getting 5 heads or out of 15 trials.

So, you are tossing the coin 15 times that means, number of trials is 15 and you can find out what is the probability of getting 5 successes or 5 heads in 15 trials. What is the reverse of this problem? Can you think of the reverse problem? I know in my sample how many trials I have made, let us say 15 that means, let us say I am asking 15 individuals. So, my sample size n is 15 so there are the reverse problem n is known that means, sample size is known, I also know that, in that out of these 15 individuals, how many people they have car so how many people owned a car, here I was trying to find out that how many head I will get that means, what is the number of successes here if I define that owning a car is actually success of that trial that means I already know out of 15 individuals how many of them are actually having a car they know. What I do not know is actually the parameters. Parameters are unknown.

So, that means in a sample size of 15 individuals, 5 individuals are having a car. That is my sample and I have already observed that sample here in terms of O is actually the observed thing. In probability distribution function you are trying to find out probability of getting 5 success that is denoted by O and what is known to you is theta which is the set of parameters, so here O indicates observe thing and theta basically indicates set up parameters in binomial distribution function, this P the parameter is actually known for a single toss it is half that is known to you.

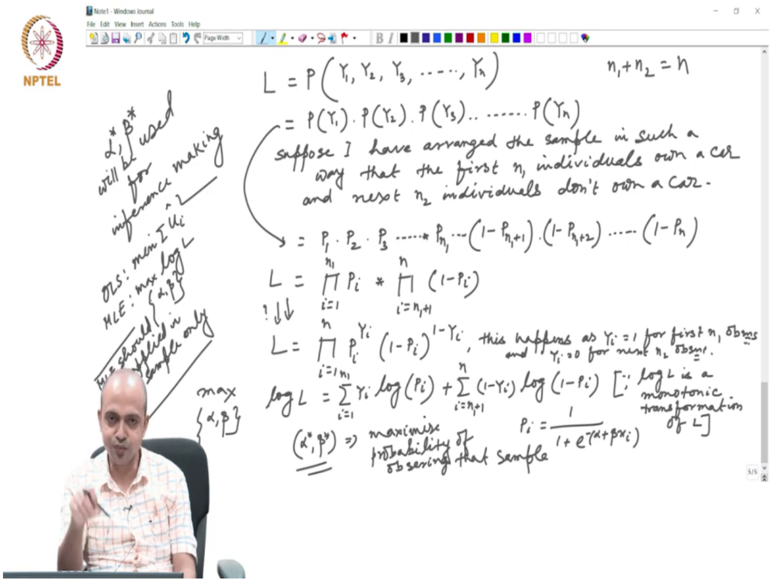
So, given theta set of parameters in probability distribution function, you are trying to find out probability of O that means probability of 5 heads, but in a likelihood function I already know that, that is known to me, what I try to find out is theta, so this is unknown that means parameter are unknown here this is known that is why I say that likelihood function is just the reverse of probability density or distribution function in likelihood function, I would like to get the parameters because a set of parameters will only define the PDF so that I can observe that sample with maximum probability that is why it is called maximum likelihood function.

That means, in short what I want, I already observed the sample where out of 15 individuals 5 individuals are owning a car, you give me the set of parameters that will maximise the likelihood of observing that particular sample, because from a given population, I could have observed some of the samples, some other samples of 15 individuals and in that sample, it would have happened that out of 15, 2 individuals that are having a car or 14 individuals having a car or 10 individuals having a car.

So, if you change the parameter, you will get different type of probability function distribution function, so I would like to get that set of parameter that will maximise the likelihood of observing that particular sample and that is the technique, what we apply here in maximum likelihood method instead of minimising the error term that we use to do in ordinary least square method.

So, this approach is little different maximum likelihood estimates is different from OLS in that regard in OLS to get the parameters, we minimise some of the error terms that mean summation Ui hat square, here we are maximising the likelihood, what is the likelihood? Likelihood of observing that particular sample were out of 15 individuals I already observed that 5 individuals they own a car, so that is just reverse or probability distribution function. In probability distribution function given theta I would like to estimate the probability of observe thing and what is observe thing, that probability of heads that I would like to observe here in likelihood function I have already observed the O, but I would like to get theta set of parameters that will maximise the probability of observing O that means, probability of observing a particular sample in which 5 out of 15 individuals own a car that is actually the likelihood function, so this reverse problem will give you a likelihood function. Now, how will you apply this likelihood function in the context of Logit model estimation?

(Refer Slide Time: 18:40)



So, let us say that the likelihood function is L denoted as probability of observing a sample and as you know, probability of observing a sample means probability of observing y1 probability of observing y2 probability of observing y3, dot dot dot probability of observing yn.

So, your sample consists of n number of responses basically y takes the value 1 and 0 this is the response function, so if you observe all your yi is that is basically the likelihood function and this so that means, likelihood function is the join probability of observing y1 y2 y3, dot dot dot yn so that means, this is nothing but probability of y1 then probability of y2 then probability multiplied by probability of y3 then probability of observing yn this is the likelihood function, which is nothing but the joint probability distribution function.

Now, suppose I have arranged the sample in such a way, that the first n1 observation n1 individuals, own a car and next n2 individuals do not own a car that is how I have actually arranged first that means, where n1 plus n2 equals to n which is the total sample size. So, that means, what I am saying probability y1 y2 dot dot dot n1 up to n1 it is actually pi, because pi is the probability of observing the sample.

So, from here what I can write equals to p1 multiplied by p2 multiplied by p3 how long this will continue up to p of n1, because first n1 observations, first in 1 individuals will own a car that is why p1 p2 p3 dot dot dot pn1 and what would be the next term, next term would be 1 minus p n1 plus 1, because next person does not have a car that is why 1 minus p for first n1 p1 p2 p3 dot dot dot pn1, next observation, next individual does not own a cat that is why 1 minus pn1 plus 1 then 1 minus pn1 plus 2 dot dot dot 1 minus pn, that is how I have arranged the data.

Now, if you write this expression in a concise format, then what you can write this is nothing but product of i running from 1 to n1 pi and then this is multiplied by i running from n1 plus 1 to n1 minus pi and this again you can write as 1 grand product pi to the power yi into 1 minus pi to the power 1 minus y, now this step is little interesting, how I could write this step from this step to this step you will need to understand how.

See this happens, because yi equals to 1 for the first n1 observations and if yi equals to 1 then this would become pi and yi equals to 0 for the next n2 observations that is why, so it will become 1 minus pi because this will become 0 so ultimately this will result in pi n2 1 minus pi this happens because as yi equals to 1 for first n1 observations and yi equals to 0 for next n2 observation now, if you take log then what will happen?

If you take log of L equals what you can write this you can say that yi log pi plus 1 minus yi log of 1 minus pi i running from 1 to n1 and here i running from n1 plus 1 to n why I have done that, because log L, why I did this because log L is a monotonic transformation of L, so this is the thing log L equals to yi log pi and plus 1 minus yi into log 1 minus pi now, you can substitute the value of pi, what is the value of pi where pi equals to 1 by 1 plus e to the power minus alpha plus beta, this is your pi, so you can substitute pi here and then you are trying to maximise these with respect to alpha and beta, so that means, I am trying to maximise this with respect to alpha and beta.

So, I want that particular set of alpha and beta that will maximise this my likelihood function and this likelihood function is basically probability of observing that particular sample in which out of 15 individuals, 5 individuals own a car this is the mechanism of maximum likelihood estimates for estimating the Logit model instead of minimising the errors some summation Ui hat square, we are actually maximising log L with respect to alpha and beta and you will get the optimum alpha star and beta star that will maximise probability of observing that sample.

So, I will take that particular alpha hat and beta hat, so that means alpha hat and beta hat then will be used for inference making, so in OLS I was minimising some of Ui hat square so OLS in case of OLS what you are doing, you are minimising summation U hat square, but in MLE, what you are doing is actually maximising log of L with respect to alpha and beta that is the difference between this and this and there are certain advantages of this MLE the advantage is that, this alpha star and beta star what you get after maximising log L that is basically the asymptotically they are asymptotically efficient and they approach normality that is also asymptotically, so they are efficient asymptotically that means what?

That means they are efficient in large sample and they approach normal they follow the normal distribution that is also asymptotically that means, the MLE method should be applied in large sample, so whenever you are estimating Logit model you have to keep in mind since the underlying estimation strategy is maximum likelihood and alpha hat and beta hat they are efficient asymptotically that means, they saw the efficiency property in large sample you must have a large sample to estimate your Logit model.

So, we should not apply the Logit model in a small sample like 330 or 40 we should have minimum 200 to 250 observations for Logit model to estimate, so with this we are closing our discussion today, tomorrow we will discuss about another class or another particular model of Binary response model or Qualitative response model and then we will also estimate this we will see how to estimate the model using a particular data set.