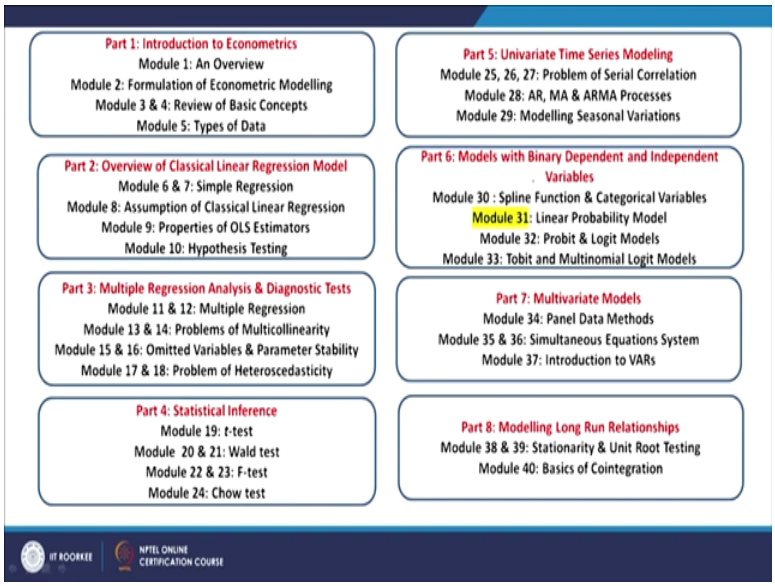
**Econometric Modelling**
**Professor Sujata Kar**
**Department of Management Studies**
**Indian Institute of Technology, Roorkee**
**Lecture – 31**
**Linear Probability Model**

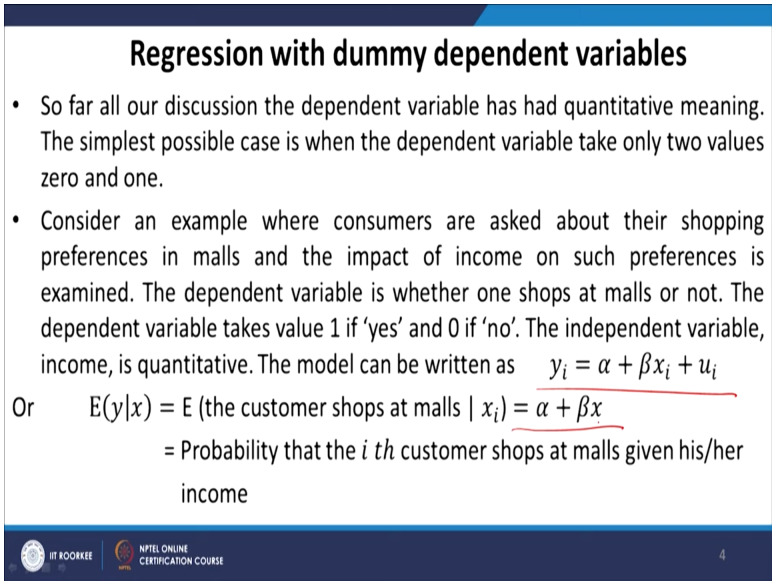
(Refer Slide Time: 0:41)




Hello everyone and this is module 31 of econometric modelling. So, far we have been discussing models with categorical or binary dependent and independent variables. So, under this part that is models with binary dependent and independent variables; so, far we have discussed spline function and categorical variables. Now, I am going to today discuss linear probability models. So, when we in the previous module that is in module 30, we actually dealt with categorical variables or binary variables in the independent variables.

So, one of the or one or more explanatory variables were either binary or having multiple categories. But, now I am going to deal with linear probability models which are basically one type of model, where we have the dependent variable as a binary variable.

(Refer Slide Time: 01:25)



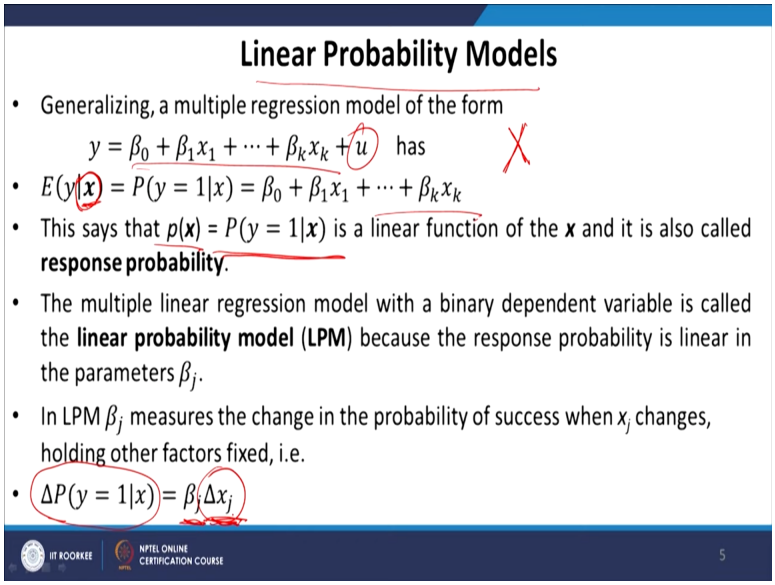

So far, in all our discussions, the dependent variable has had quantitative meaning; that is they were quantitative variables. The numbers themselves carried some implications. The simplest possible case is when the dependent variable takes only two values zero and one. So, that is one situation where we will be moving away from quantitative variables to qualitative variables. And there we have basically regression with dummy dependent variables.

Now, consider an example where consumers are asked about their shopping preferences in malls, and the impact of income on such preferences is examined. Basically, we are trying to find out that whether shopping preferences in malls are linked with the income levels of individuals. So, the dependent variable is whether one shops at a mall or not. The dependent variable takes value 1 if the individual or the person says 'yes, and 0 if he says 'no'.

The independent variable is income, which is a quantitative variable. So, we would be considering income as stated by the individuals. The model can be written as yi equals alpha plus beta xi plus ui. You can see that here we are so for considering only one independent variable that is income; and i refers here to the ith individual. Now, if we take the expected value, then the expected value y given x is the expected value; that the consumer or the customer shops at malls, given his or her income level *(refer to slide time 01:25)*.

So, this equals to alpha plus beta x; this measures the probability that the ith customer shops at malls given his or her income. So, the expected value measures the probability; because as you can see that y takes only two values 1 or 0. So, the expected value of y taking the value 1 is the probability that the ith customer shops at malls, given his or her income.

(Refer Slide Time: 03:45)




So, generalizing a multiple regression model of the form, we would be having y equals beta naught plus beta1 x1 and so on plus beta k xk plus u. So, you can see that we are now considering k plus 1 independent variable. Now, the expected value of y is given x, here I would like to mention that probably in this module and in the next module also. This small x in bold refers to the collection of all the independent variables. Now, if I write it x, capital X then it actually refers to the matrix consisting of all the observations and all the independent variables. But, here possibly this small x in bold refers to one individual's observations on all the independent variables *(refer to slide time 03:45)*..

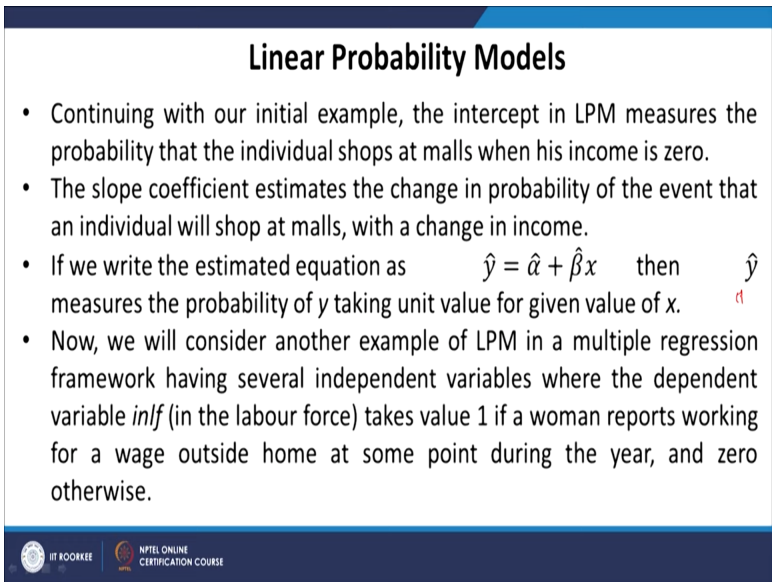
So, writing the expected value of y given x which is equal to the probability that y takes the value 1, for given values of x, is equals to beta naught plus beta1 x1 plus beta2 x2 and so on, plus beta k xk. Of course, we will not have the error term here. Now, this says that the probability of x which is equal to the probability of y, given y takes the value 1 for given values of x. And this is defined or denoted by p x small p x in bold is a linear function of the x; and it is also called the

response probability. So, ideally, this is called response probability. What is the probability that the ith individual is going to come up with a positive response, given the values of the x s.

The multiple linear regression model with a binary dependent variable is called the linear probability model; because the response probability is linear in the parameters beta j. Since we are considering linear functions that is why this is called linear probability models. Later on, we will also examine non-linear functional forms. In LPM beta j measures the change in the probability of success when Xj changes, holding other factors fixed. That is what we are trying to find out that if income changes of the jth individual by one unit, then how much does the probability or the response probability changes. How much the probability changes of the person shopping from the mall is given by delta P, y taking the value 1 for given values of x.

So, how much the probability is changing, is given by beta j multiplied by the change in the independent variable. So, when we have a large number of independent variables, we might consider all of them fixed except for one change that individual independent variable; say variable jth variable is changed. And when it is multiplied by its corresponding coefficient, then it shows that how much the probability of y taking value one changes, when the value of jth independent variable is changed by one unit.

(Refer Slide Time: 07:08)

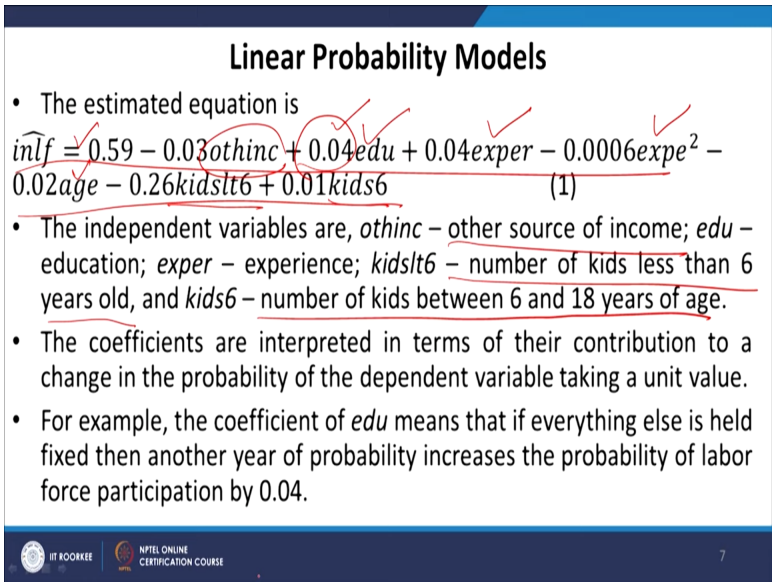


So, continuing with our initial example, the intercept in LPM measures the probability that the individual shops at malls when his income is zero. So, that is the usual interpretation of the intercept term that if we hold the values of independent variables zero; then what is the value of the dependent variable? Now, here the dependent variable takes only two values one and zero. And that is why here the intercept measures the probability that the individual shops at the mall; that is probability y takes the value one when his income is zero or the independent variable takes a value zero. The slope coefficient estimates the change in probability of the event that an individual will shop at malls, with a change in income.

If we write the estimated equation as y hat equals alpha hat plus beta hat x; then y hat measures the probability of y taking unit value for a given value of x. So, y hat essentially here is the estimated response probability. Now, we will consider another example of LPM. In a multiple regression framework having several independent variables, where the dependent variable (inlf) which is in the labor force takes value 1. If a woman reports working for a wage outside the home at some point during the year, and zero otherwise. So, this is a multiple regression equation where we will be having several independent variables; some of them are categorical, some of them are not.

But, the dependent variable most importantly is a binary variable having a value 1. If a woman so all the participants are women here if a woman has been in the labor force during last one year period time; then inlf will take a value one, otherwise, it will take a value zero.

(Refer Slide Time: 09:11)




So, suppose the estimated equation is like this *(refer to slide time 09:11)*. You can see that the other variables that are considered here are first of all we have the constant term; this other income othinc refers to other sources of income. So, other income sources the family has other than the income obtained by or income created by the woman herself. So, it is expected that other income is going to have a negative impact on women's labor force participation. That is if family income is very high, then given other things a woman may choose not to participate in the labor force. Then we have education as the second independent variable, experience as the third independent variable.

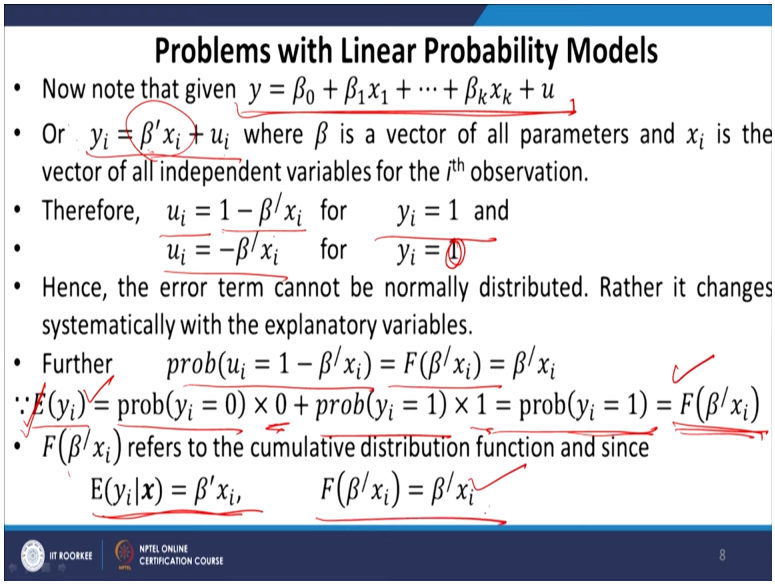
We also consider experience square; age is the sixth independent variable or fifth independent variable. This is kids lt6 implies the number of kids less than six years old, and kids6 implies the number of kids between 6 and 18 years of age. Now, this is an equation that we are also going to consider in a later module, or while comparing the linear probability model with other types of binary dependent variable models. So, this is an important equation *(refer to slide time 09:11)*. Now, the thing is that given all this we have already defined the equation and presented its estimates. The coefficients are interpreted in terms of their contribution to a change in the probability of the dependent variable taking a unit value.

For example, the coefficient of education means that if everything else is held fixed; then another year of probability increases the probability of labor force participation by 0.04. So, if education

increases by one unit that is by one year; because here education is measured in terms of the number of years, one has spent on getting an education. And if it increases by one more year, then 0.04 actually measures the probability of being in the labor force. Now, it has a positive sign which implies that an increase in education contributes positively to the probability of a woman being in the labor force.

Now, I have not mentioned the t values or the standard errors of the statistics. Originally, the equation tends to have all significant variables.

(Refer Slide Time: 11:57)




Now, note that given this expression *(refer to slide time 11:57)*. So, we have already understood that how LPM are constructed, how do they operate, and how the parameter estimates are interpreted. Now, we talk about the problems with the LPMs or linear probability models. Note that given y equals beta naught plus beta1 x1 plus beta k xk plus u that is we are considering a multiple regression equation. Or, which alternatively can be written as yi equals beta i beta prime xi plus ui, where beta is a vector of all parameters; and xi is the vector of all independent variables for the ith observation.

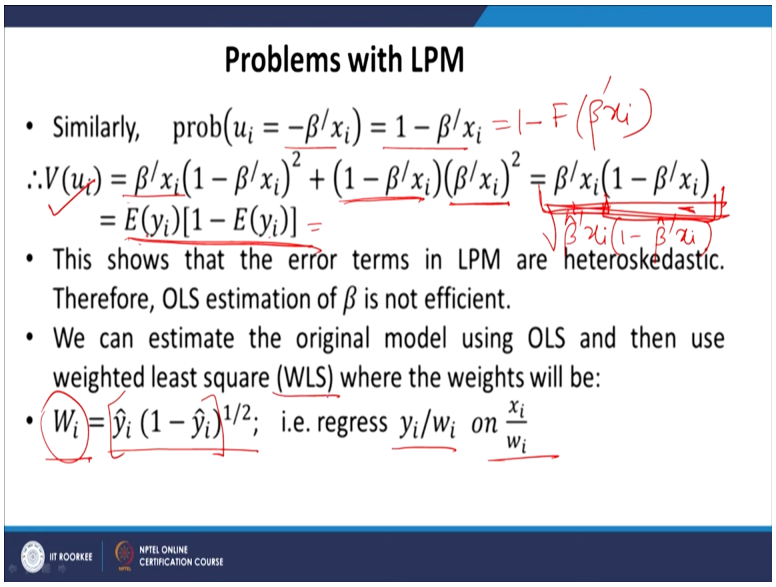
Then, we can write ui as 1 minus beta prime xi; right simply for yi equals 1. So, when yi takes the value 1, then taking this expression to the other side will be having ui equals 1 minus beta prime xi. And similarly, when yi takes the value 0, then ui is simple minus beta prime xi. Hence the error term cannot be normally distributed; rather it changes systematically with the explanatory variable. So, this is the first problem that we encounter with the LPM that the error terms are might not be normally distributed; because it is linked to the independent variables. Further, the probability that ui takes the value 1 minus beta prime xi, is equal to F beta prime xi, which is equal to beta prime xi.

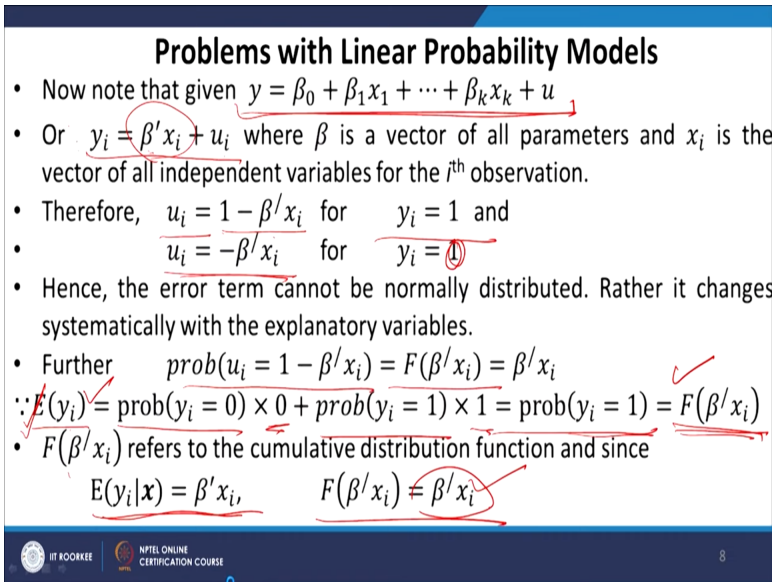
Why this is so? Because you know the expected value of u yi, how do we calculate it? We calculate it as probability yi equals to 0. The probability that yi takes the value 0 multiplied by

the value it takes which is actually 0, plus the probability of yi taking the value 1, and multiplying the value it takes which is 1. So, this is how we calculate the expected values; and this simply turns out to be the probability of yi taking the value 1, which is the cumulative distribution function. We denote it by F beta prime xi. And you can see that probability of yi when it is the probability of yi takes the value 1.

If beta prime xi refers to the cumulative distribution function, and since the expected value of yi given x equals to beta prime xi. So, ideally, we first show that the expected value yi is equaled to the cumulative distribution function of the probability of y, when y takes the value 1*(refer to slide time 11:57)*. And then since expected value yi is equaled to beta prime xi. Therefore, we have the cumulative distribution function also taking the value beta prime xi. Now, how we are going to use it.

(Refer Slide Time: 15:14)






So, similarly, the probability of ui taking the value minus beta prime xi will be equal to 1 minus beta prime xi. That is similarly 1 minus the cumulative distribution function; so that is why we have the probability of ui taking the value minus beta prime xi is 1 minus F beta prime xi. Now, when it comes to the calculation of the variance of ui how do we do it? We do it as probability multiplied by the square of the values. So, the probability that ui will take a value beta prime xi is 1 minus beta prime xi; so, 1 minus beta prime xi square multiplied by the probability.

And similarly, the probability that ui will take the value beta prime xi is actually 1 minus beta prime xi. So, 1 minus beta prime xi multiplied by the beta prime xi square. By rearranging terms we would find that this is equal to beta prime xi multiplied by 1 minus beta prime xi. And since the and this would be exactly equal to expected value yi multiplied by 1 minus expected value of yi.

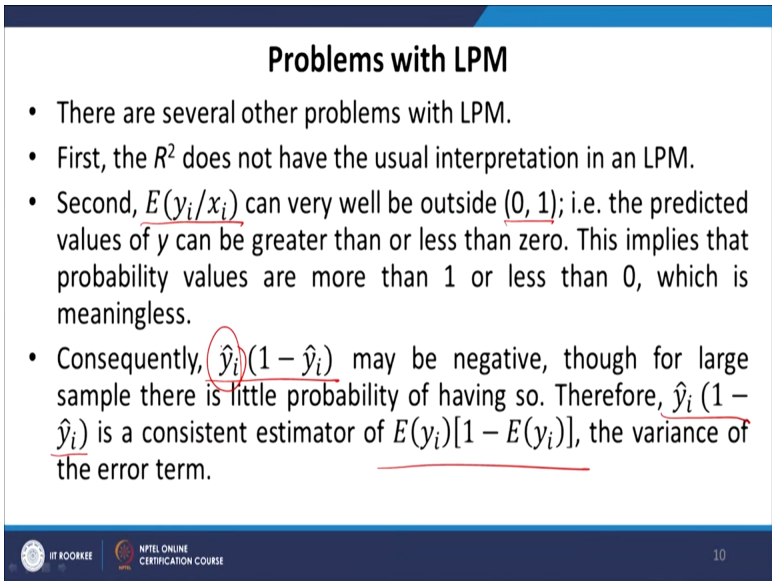
Because we have seen that the expected value of yi given x is beta prime xi. So, that is why this is equal to the expected value of yi given xi; and this is the expected value of yi given xi when subtracted from 1. So, 1 minus the expected value of yi given xi. So, this shows that the error terms in LPM are heteroskedastic; therefore, the OLS estimation of beta is not efficient. The error terms are linked or the error population error variance is linked systematically with the independent variable. So, we cannot have homoskedastic errors, when the errors are not

homoskedastic or they are heteroskedastic. Then we do not have the best estimation or best estimators of beta. Beta is not the most efficient one.

We can estimate the original model using OLS and then use weighted least squares or WLS, where the weights will be denoted by the wi. The weights are yi hat multiplied by 1 minus yi hat raised to the power half; this is because as you can see that it is the error variance *(refer to slide time 15:14)*. Now, the estimated counterpart of this error variance will be or this is the error variance. And the estimated counterpart of this error variance would be beta hat xi, multiplied by 1 minus beta hat prime xi. And root over this is our standard deviation of the estimated value of the population error variance, or standard deviation estimated standard deviation of the population error term.

So, that is why this is the weight that would be used. And by using this weight, we regress yi divided by wi on xi divided by wi. So, both yi and xi both the dependent variable and the independent variables are now converted or transformed rather; and then we go for this another set of regression, which is the WLS with these transformed variables in order to have efficient estimators.

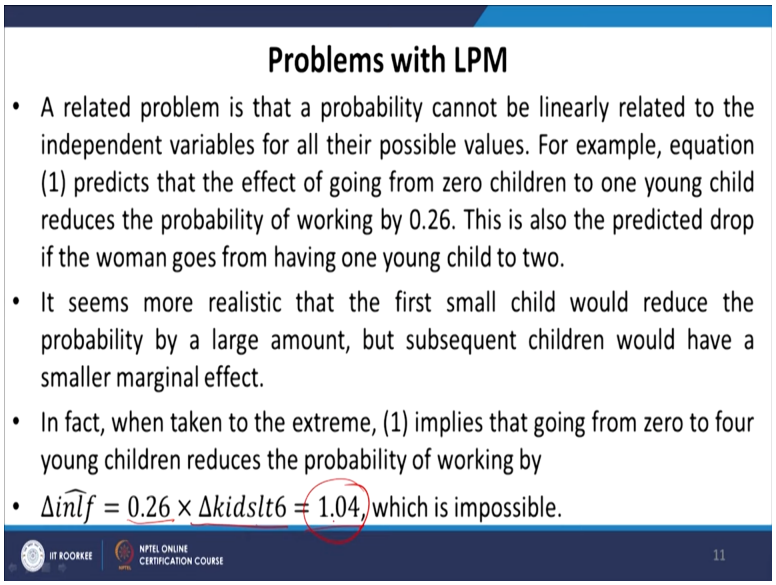
(Refer Slide Time: 18:46)

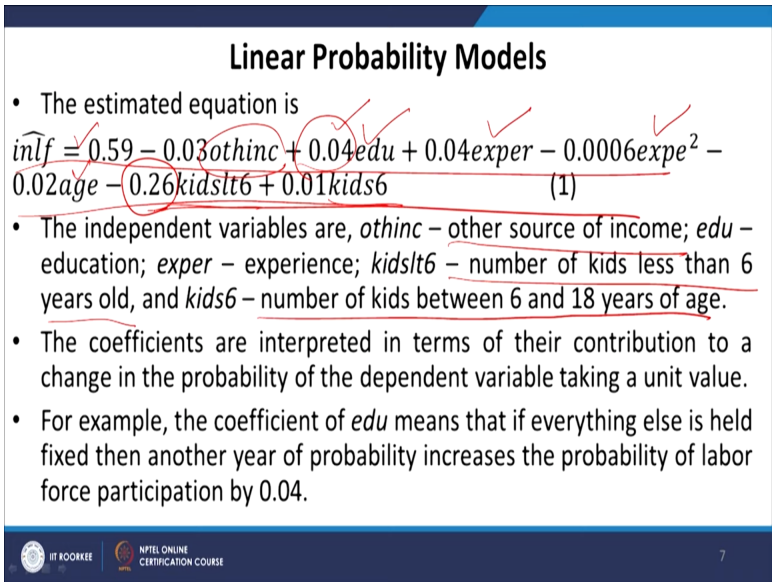


But, there are some more problems with LPM. So, the other problems are first of all the R square does not have the usual interpretation in an LPM, the way we have it for the simple OLS regression or multiple OLS regression. The second expected value of yi given xi can very well be outside 0, 1. That is the predicted values of y can be greater than or less than zero. This implies that probability values are more than 1 or less than 0, which is meaningless; because you know that expected values are response probabilities. So, if expected values are the estimated or predicted values of yi, if they are greater than 1 or less than 0; then this implies probabilities are greater than 1 or less than 0, which is meaningless.

Consequently, yi hat multiplied by 1 minus yi hat may be negative; though for a large sample there is little probability of having so. Therefore, yi hat multiplied by 1 minus yi hat is a consistent estimator of the expected value of yi multiplied by 1 minus the expected value of yi, the variance of the error term. So, what we expect LPM to be is to provide us asymptotically efficient estimators; or we can also say that the estimator of the error variance is consistent. So, as sample size increases, we expect that yi hat multiplied by 1 minus yi hat to be non-negative. But, otherwise, there is always the possibility of the value yi hat taking positive greater than 1 or less than 0 values.

(Refer Slide Time: 20:36)






A related problem is that a probability cannot be linearly related to the independent variables for all their possible values. For example, equation-1 predicts that the effect of going from zero children to one young child reduces the probability of working by 0.26. This is also the predicted drop if the woman goes from having one young child to two. So, when we go back to equation-1; the equation having multiple independent variables. The value of this 0.26 measures the positive probability or decline in the probability, when a woman has one more child; who young child that is the age of the child is below six years.

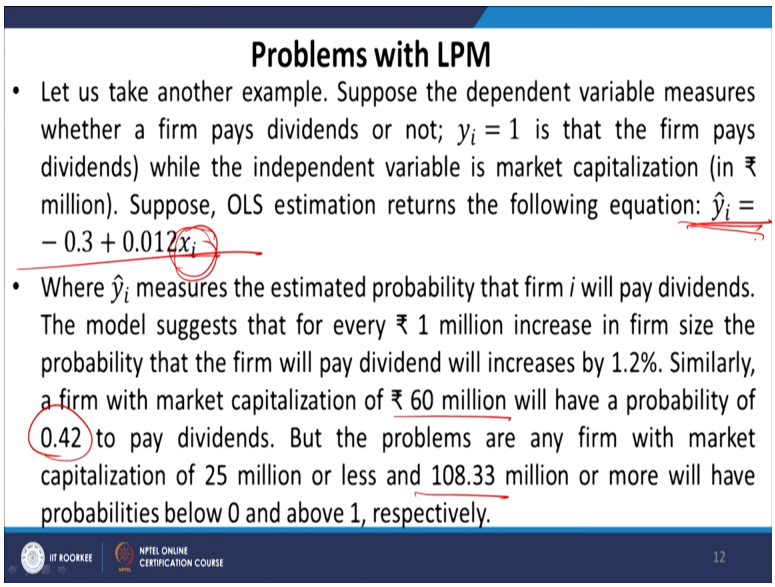
Now, since it is 0.26, it implies that a woman having no child to when she has the first child; then the probability of being in the labor force declined by 0.26. Similarly, a woman who has a single child when she goes for the second child; then also the probability decreases by 0.26 and so on. So, when a woman has three children and she goes for the four children, then also the probability of being in the labor force decreases by 0.26. But, this is not probably much realistic.

Because the thing is that generally, we would expect that a woman when she has the first child, probably the probability of not being in the labor force would be a sharp drop. But, a woman who already has a few children or maybe one or two children, and she goes for the second or the third child; then the probability drop or drop in the probability of being in the labor force would be somewhat lesser. So, this is the this that is why it seems more realistic that the first small child

would reduce the probability by a large amount; but subsequent children would have a smaller marginal effect.

In fact, when taken to the extreme, equation-1 implies that going from zero to four young children reduces the probability of working by 0.26 multiplied by 4. So, this actually turns out to be 1.04, which is impossible. So, if we go straight away from one to four children, the number of kids less than six years of age, when increases from zero to four; we would be multiplying 0.26 by a number 4. And would be getting a probability or change in probability by 1.04.

(Refer Slide Time: 23:28)

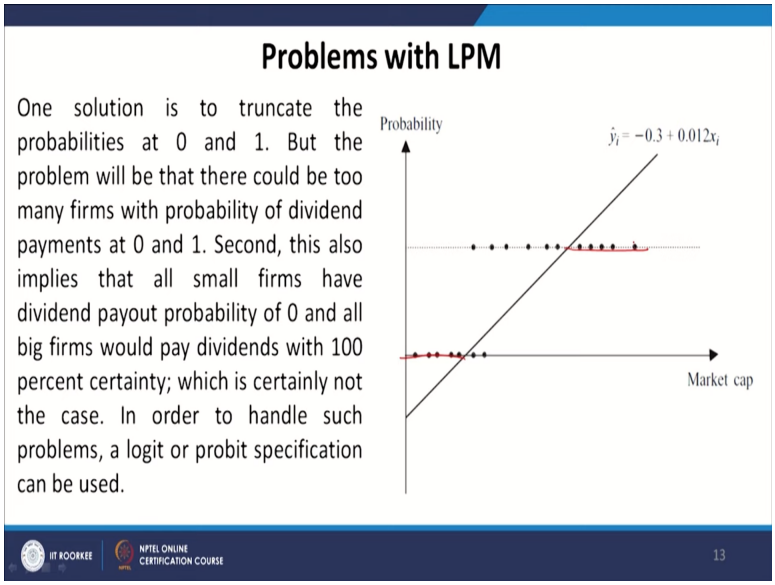



So, let us take another example. Suppose the dependent variable measures whether a firm pays dividends or not, yi equals to 1 when the firm pays dividends; while the dependent variable is market capitalization that is measured in terms of millions of rupees *(refer to slide time 23:28)*. And yi takes value 0, when the firm does not pay dividends. Suppose OLS estimation returns the following equation. We estimate this model and have yi equals minus 0.3 plus 0.012 xi. Now, where yi hat measures the estimated probability that firm i will pay a dividend. The model suggests that for every 1 million rupees increase in firm size, the probability that the firm will pay dividends will increase by 1.2 percent.

Similarly, a firm with a market capitalization of 60 million will have a probability of 0.42 to pay dividends. This how we obtain? Because you can see that here the value of xi is 60 million. So, we are having 60 million here, 60 million multiplied by 0.012 minus 0.3 gives us this value 0.42. So, the probability of a firm having a having market capitalization value worth 60 million rupees is 0.42; that the firm will pay dividends. But, the problems are any firm with a market capitalization of 25 million or less, and 108.33 million or more will have probabilities below 0 and above 1 respectively.

So, if we put 25 million here or any number other than 25 million, any number lower than 25 million; then we will be having yi value that is less than 0, and 108.33 million is the upper threshold limit. So, for any value of x which is greater than 108.33 million, our yi hat value will be greater than 1. So, we encountered the problem of having predicted probabilities, which are either less than 0 or greater than 1 which are meaningless.

(Refer Slide Time: 25:45)

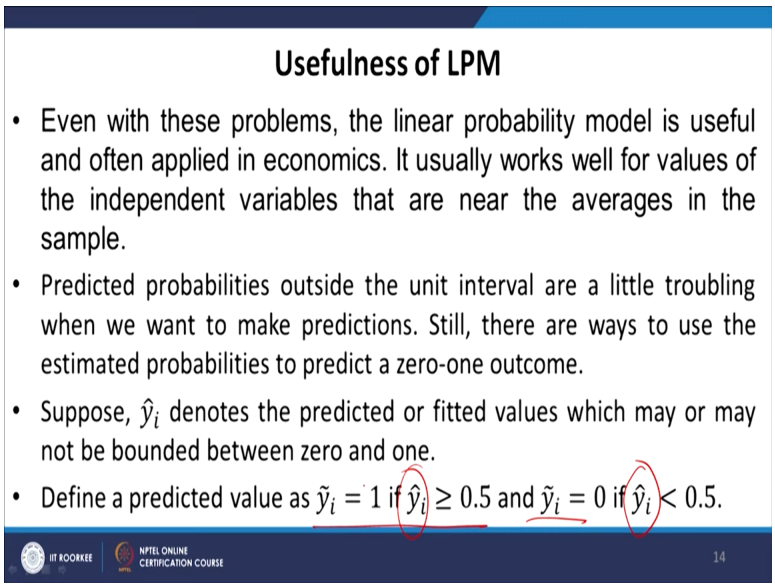



One solution is to truncate the probabilities at 0 and 1. But, the problem will be that there could be too many firms with the probability of dividend payments at 0 and 1. So, instead of considering 0, below 0, and greater than 1 probability, we all truncate them at 0 and 1. So, all firms having estimated probability greater than 1 are assigned a value 1, and all firms having estimated values less than 0 are assigned value 0. But, then the problem is that there could be too

many firms having probability 0, and all big firms would pay dividends with 100 percent certainty; which is certainly not the case. In order to handle such problems, our logic or profit specification can be used.

So, these are the problems with LPM, and in order to handle this kind of problem that probabilities getting into negative values. We would be using alternative models and some of the simplest possible and the most common type of dependent variables. Binary dependent variables are a common types of models with binary dependent variables is probate and logic models.

(Refer Slide Time: 27:03)



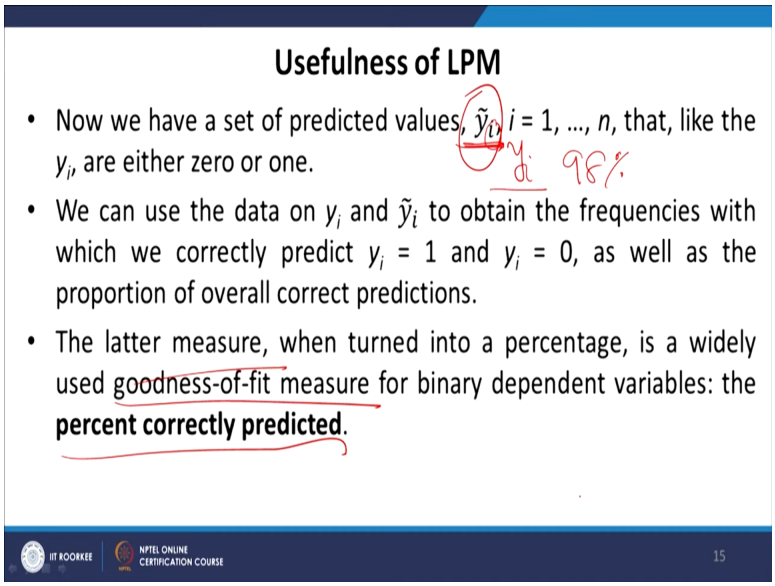

But, there is certain usefulness of LPM. Even with these problems, the linear probability model is useful and often applied in economics. It usually works well for values of the independent variables that are near the averages in the sample. So, if the sample is not very diverse, the majority of the sample observations are around the average values. Then probably we would not encounter the problem of having estimated probabilities below 0 or greater than 1.

So, if only a few observations, for example, the sample from which the estimated equation was reported for women being in the labor force or not, had 96 percent of the women having most of the values around the average. So, which implies that only 4 percent of the values were outliers, and probably their predicted probabilities would be less than 0 and greater than 1. So, those 4

percent can essentially or can be easily removed from the sample. And one can work with 96 percent or for 96 percent of the results; we would be getting desirable results. So, in that case, LPM could well be applicable. Predicted probabilities outside the unit interval are a little troubling when we want to make predictions.

Still, there are ways to use the estimated probabilities to predict a zero-one outcome. Suppose, yi hat denotes the predicted or fitted values, which may or may not be bounded between zero and one. So, define a predicted value as yi tilde equals to 1, if yi hat takes a value greater than equal to 0.5. And yi tilde takes a value 0, if yi had values are less than 0.5; so, we are these are our estimated probabilities. So, estimated probabilities are assigned values 1, when the value is greater than equal to 0.5; and all the estimated values are assigned a value 0, whenever, the predicted probabilities are less than 0.5.

(Refer Slide Time: 29:21)




Now, we have a set of predicted values yi tilde that like the yi are either 0 or 1. So, my original values are original series the dependent variable had only values 1 and 0. And now my predicted values are also 1 and 0, because I am not using the estimated predicted values; but, I am have coded them. And now I am using yi tilde. We can use the data on yi and yi tilde to obtain the frequencies with which we correctly predict yi equals 1, and yi equals 0; as well as the proportion of overall correct predictions. So, next what we do is matching that is if my yi tilde is

1, then we cross check that whether corresponding to this i, my yi was also equal to 1 or not. And on that basis we actually arrive at a goodness-of-fit measure; so this is called the percent correctly predicted.

The closer or the higher the value, it implies the better is the estimate. So, if roughly I observe that 98 percent of the time, when yi tilde takes a value 1, yi is also 1. Then it implies that the model is able to somewhat correctly predict the possible outcome in terms of 1 and 0 value. So, the latter measure, when turned into a percentage is a widely used goodness-of-fit measure for binary dependent variables; and it is called the percent correctly predicted.

So, that is all about the linear probability models. In the next module, I will be discussing profit and logic models, which are supposed to take care of some of the problems that we encounter, while dealing with LPM. Thank you.