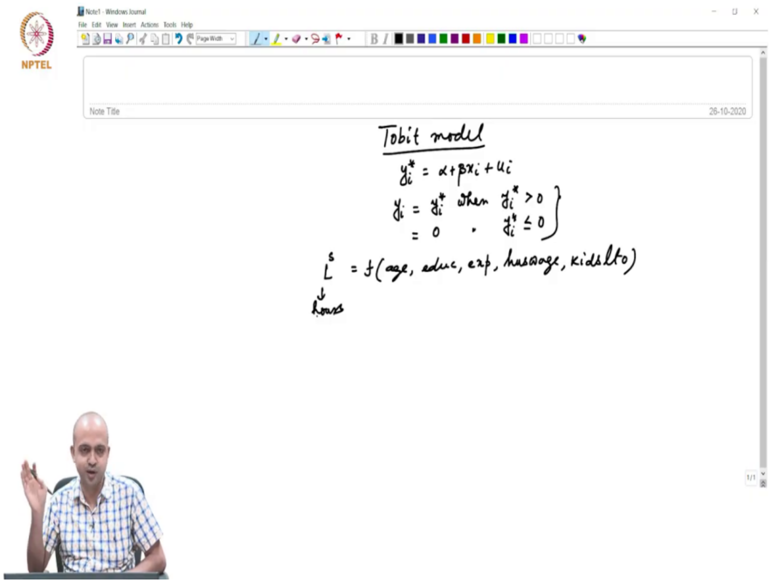
Introduction to Econometrics
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology Madras
Lecture 64
Qualitative Response Models - Probit and Tobit Models Part - 4

(Refer Slide Time: 0:15)



Welcome once again to our discussion of censored regression model or limited dependent variable model or Tobit model. And in our last class, we completed our discussion on the theoretical portion of Tobit model. And today, we are going to learn how to basically estimate Tobit model using a data set and how to interpret the coefficients.

So, before we start, we will just quickly recap the theoretical portion once again, so that our understanding would be better. So basically we were discussing Tobit model and this is the latent variable yi star equals to alpha plus beta xi plus ui. yi star is not observable and what we observe is actually yi and the relationship is yi equals to yi star when yi star greater than 0, and yi equals to 0 when yi star is less than or equal to 0.

So, that means the negative values of yi star, we do not observe, and we put 0 for all negative as well as 0 values of yi star, and we consider only yi greater than 0 for the uncensored model. So basically, this was the idea. We have two types of observations, one for which yi takes the value of 0 and other for which yi takes the value greater than 0. And we also discussed about the likelihood function combining these two and we said that we cannot simply throw away those observations for which we have 0 value.

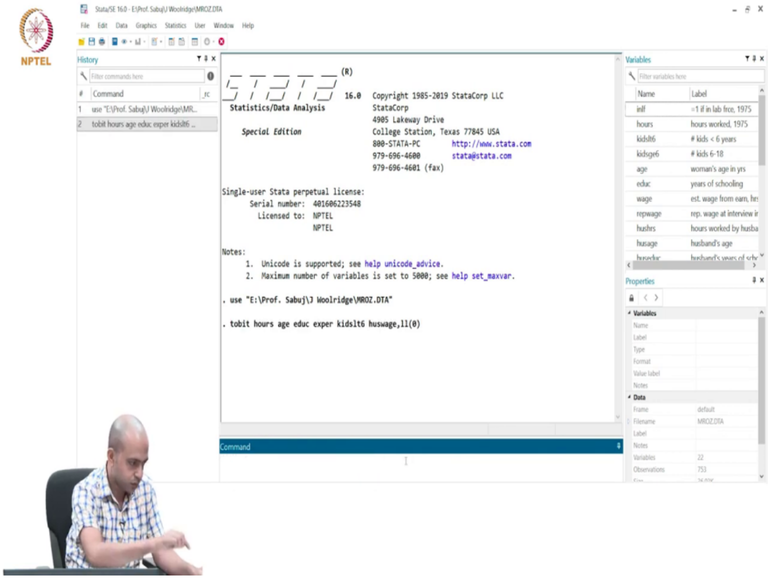
Because then your sample would be a truncated sample and in that truncated sample, your error term would be following a truncated normal distribution. So OLS cannot be applied there. So, there you basically require a truncated regression.

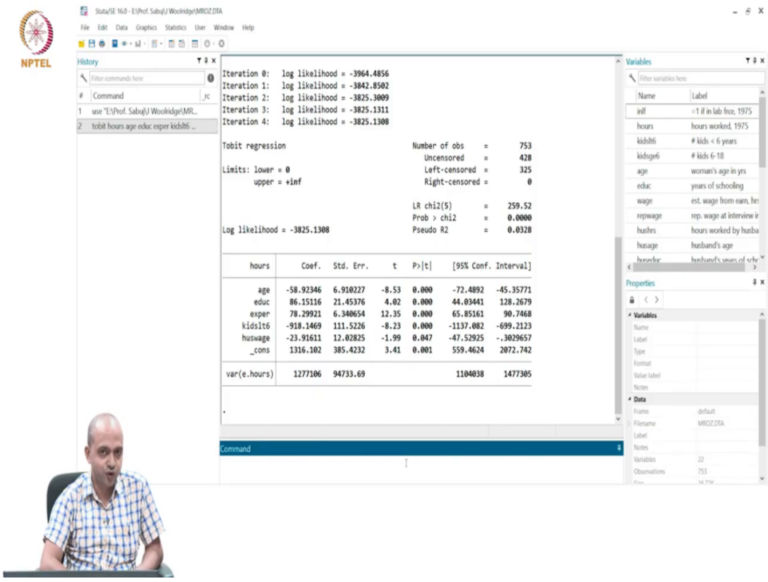
The example what we are going to discuss here is labor force participation. Our model is labor supply is a function of age, education, experience, husbands wage, number of kids. This is my labor supply function and this labor supply is measured by hours.

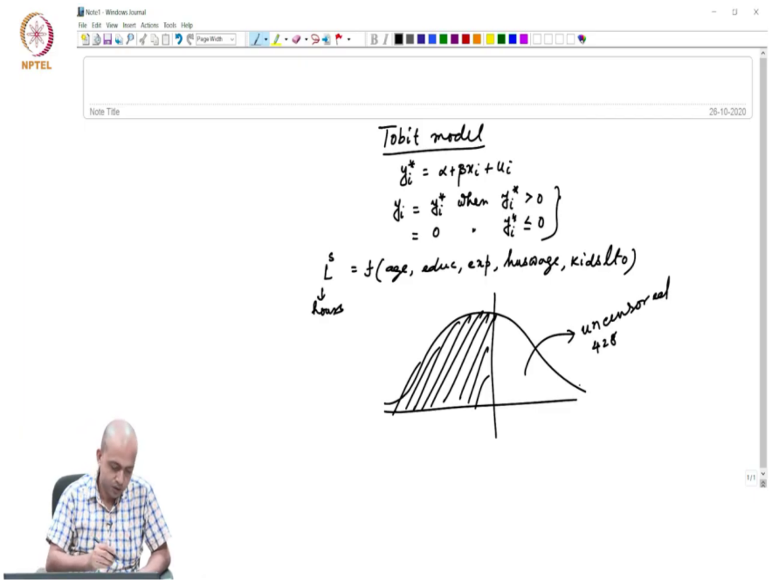
But the only problem is, we will observe the labor supply value for those who have actually participated in the labor market. For those who have not participated we will not observe any value and we will put 0. So, that is why the limited dependent or censored regression or Tobit model is used here. And as we discussed, there could be three types of censoring, one is censoring from below, censoring from top or censoring from below as well as top.

This case what we are going to discuss is only censoring from below. So, that means you have a minimum censoring at 0 level. So, that means we will observe someone's labor supply only when it is greater than 0. So, if that is the case, then we will pull left censoring or lower censoring in this particular case, but there are other cases. Depending on your situation you have to use either right censoring or censoring from both top and bottom.

(Refer Slide Time: 5:20)







With this, let us now go to our same data set that we were using earlier in the context of Logit and Probit. But, there is a difference in our objective in the case of linear probability model, Logit model and Probit model. We are basically interested in estimating the probability of labor force participation by a married woman given her socio-economic and demographic factors.

So, given a particular married woman's age, education, number of kids and husbands wage, we are basically going to predict what is the probability that a married woman will participate in the labor force. But here in the context of Tobit, we are no more interested in estimating probability of labor force participation, rather, here our interest is to estimate the responsiveness of labor supply, that means elasticity of labor supply with respect to these factors what we have just mentioned.

So, there is some difference in our objective and that is why the modeling is also different. Tobit is quite different from the other probabilistic models because we are no longer interested in estimating probability. So, now, we will first estimate the model and the command is very simple instead of Logit and Probit, we have to just use Tobit that is a command.

So, this is Tobit and then your dependent variable is hours, then your age, then your education, then experience and number of kids less than 6 years of age and husbands wage. Then you have to put the censoring. Since, we are putting censoring from below it is called l l. l l is for lower limit.

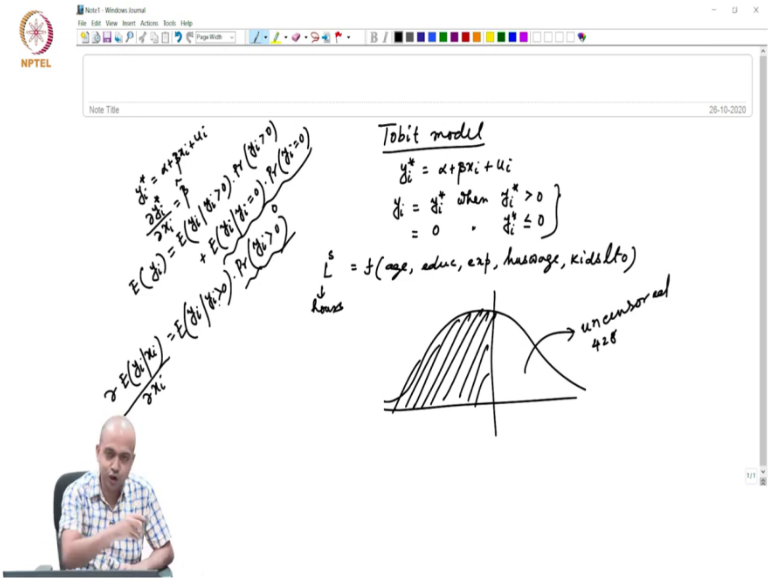
Look at this. So, this is the output. This is the output where again the estimation procedure as we have discussed in our previous class that estimation requires again formulating a likelihood function and what is the likelihood function? Likelihood function if you recall that is basically a product of two types of observations. One is for yi star greater than 0 and another is for yi star less than or equal to 0.

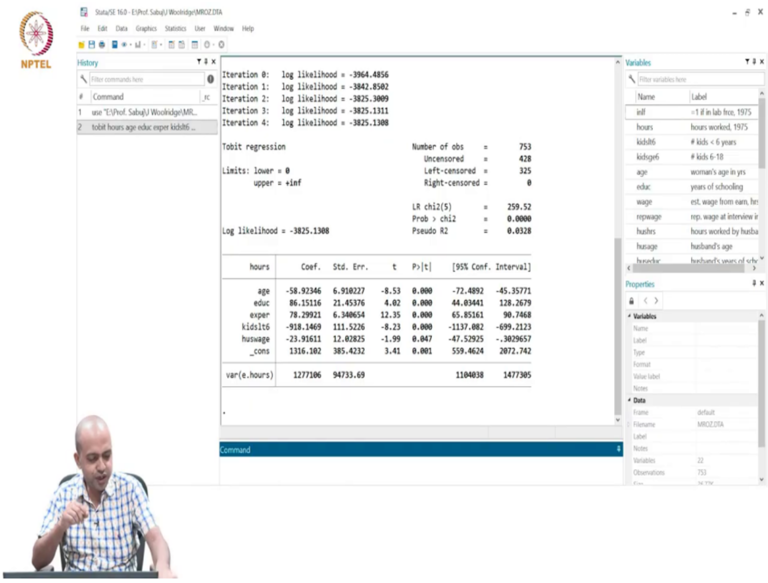
So, combining these two we formulated the likelihood function and that likelihood function, we need to maximize with respect to your parameters which were alpha, beta and sigma square. And if you do so, then the same likelihood estimation method will give you this type of output. And here in the context of likelihood estimation you have total 753 number of observations, but out of which you get 428 uncensored observations. That means what?

That means 400 in terms of a simple diagram I can say that, so this is your censored part, this is your uncensored part and total number of observations is 753 and here it is 428. So, this portion you have 428 observations, that is what it means. And then you have got this model. Now, the next question is how are you going to interpret this coefficient? Are we going to interpret these coefficients as marginal effects that means can we say that as age increases by 1 unit or 1 year then the labor supply goes down by 58.92 hours?

If we can do so, then we can directly say that these coefficients are actually the marginal effects. Can we do that? The answer is no. Why this is so, because these coefficients they look exactly like your OLS estimates. But still we cannot use these coefficients as our marginal effects. Why this is so? We cannot interpret the coefficients directly as marginal effect in Tobit model as well.

(Refer Slide Time: 11:22)





See here we say that yi star equals to alpha plus beta xi plus ui. So, the beta that we got from that output is actually from here that means, we can say that del yi star and del xi equals to beta. So, if you estimate the coefficient that is basically showing for the unit change in xi what is the change in that latent variable, but latent variable is not observable rather what you want is yi. That means this is our prime variable of interest.

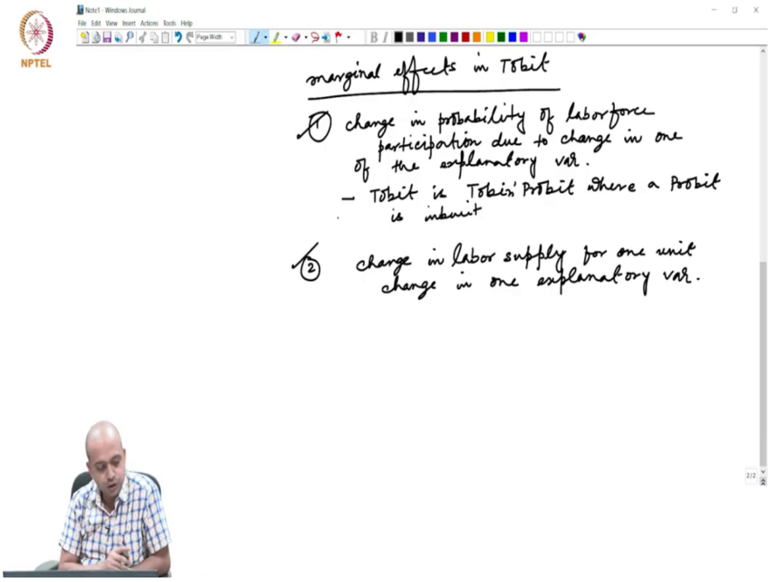
So, that means, what do you have to do? First let us see what is expectation of yi. Now, this expectation of yi is basically like this, there are two expectations- expectation of yi when yi is actually greater than 0 and that you have to multiply with the probability that yi greater than 0 plus expectation of yi given yi equals to 0 multiplied by probability that yi equals to 0.

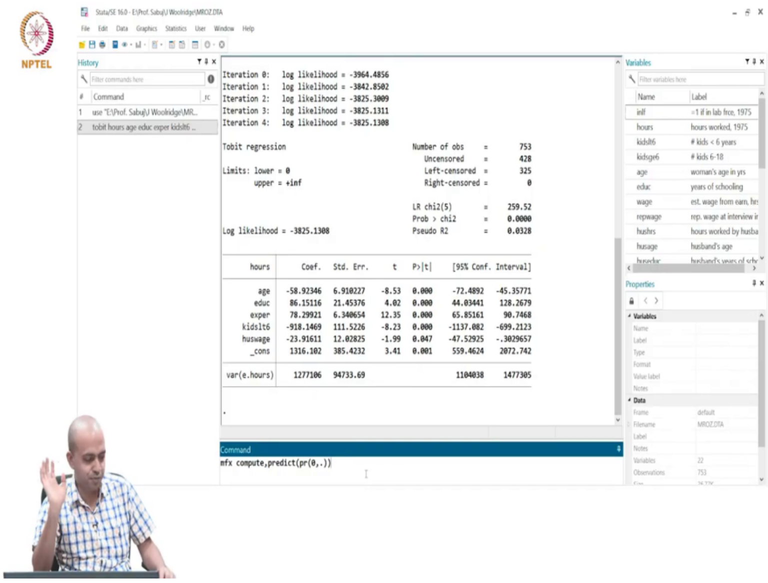
So, this portion is actually 0, so if you multiply this, this becomes 0. So, that means expectation of yi given xi you can write equal to actually expectation of yi for yi greater than 0 multiplied by the probability that yi greater than 0. That means, we cannot directly take those coefficients as marginal effect because this probability yi greater than 0 should also be there. First of all the married woman should participate in the labor market, then only we can observe their labor supply and then only we can get those coefficients as marginal effect.

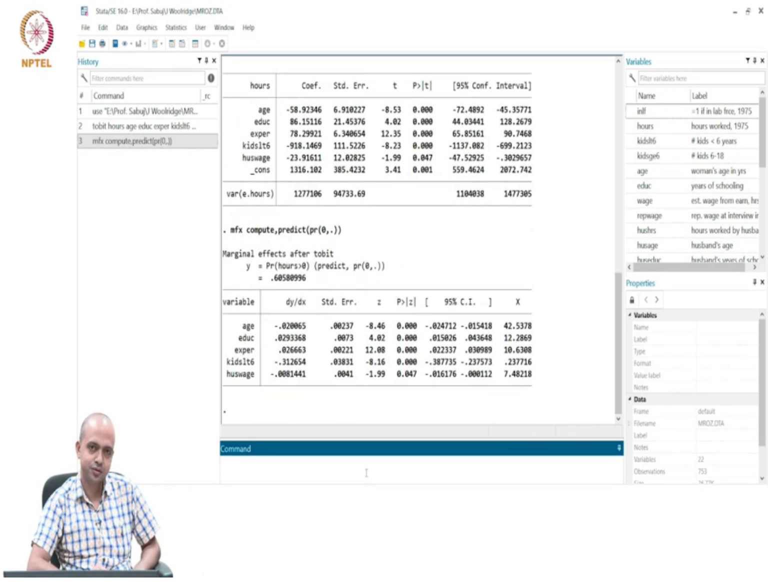
Because of this second component which is also multiplied with this expectation, we cannot take the coefficients after Tobit estimation as marginal effect. So, please keep in mind, we can take the coefficients as marginal effect only when we are estimating the model using OLS. We are using OLS because here the beta hat is coming from that yi star which may or may not have direct interpretation all the times and which is not observable.

So, that is why delta yi delta xi is basically this, but what do we want is delta expectation of yi given xi, which is actually this. So, that is why since the probability yi greater than 0 is also multiplied with this, we cannot take those as direct interpretation of marginal effects. So, again like the Logit and Probit model we need to put specific command for getting the marginal effects.

(Refer Slide Time: 15:39)







Now, here in the context of Tobit, marginal effects are of two types. One, if you recall we said that Tobit model is basically a Tobin's Probit that means, a Probit model is inbuilt in the Tobit model. So, there are two stages; in the first stage, the individual decides whether to participate in the labor force or not.

That means, from the first stage we can calculate what is the probability a married woman will participate in the labor force that is why from Tobit’s model, we can actually estimate the change in probability in labor force participation for the unit change in any of these explanatory variable. That is type 1 marginal effect. So, what I am saying the type 1 marginal effect is change in probability of labor force participation due to change in one of the explanatory variables. And why we are able to do this? Because Tobit is basically Tobin's Probit where a Probit is inbuilt.

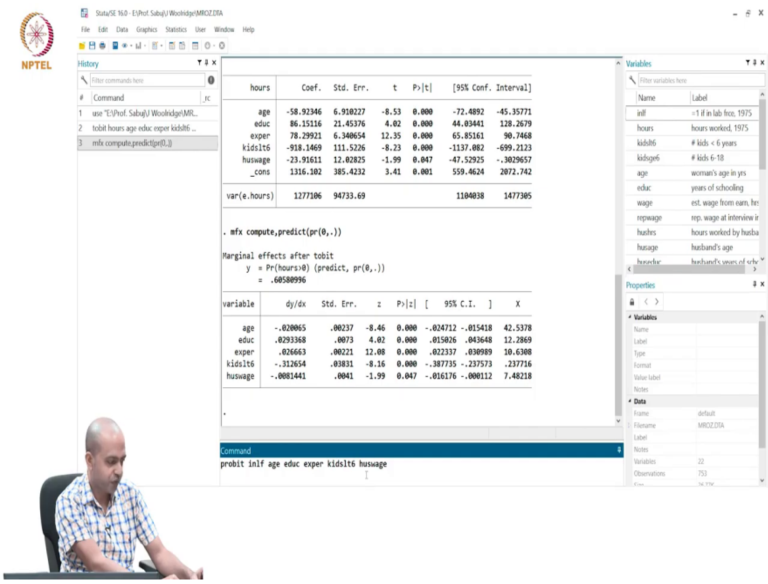
And the second type of marginal effect, once you decide to participate in the labor force and we can actually observe, so, this is basically for the censored observations. And then in the second stage, when you actually participate in the labor force and we can observe your labor supply from there what you can get is change in labor supply for one unit change in one explanatory variable.

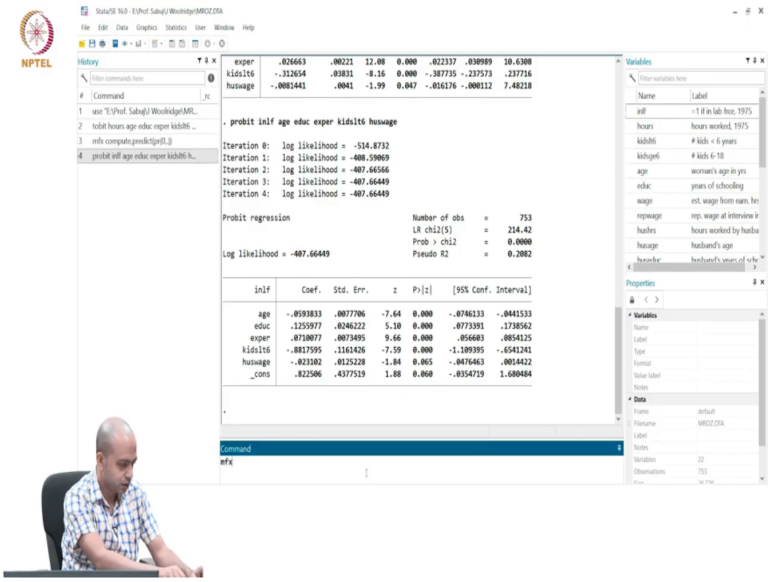
For example, as education increases by one unit, what is the probability of labor force participation is one type of marginal effect that we can get. And secondly, for the unit change in education, what is the change in labor supply amount that means the elasticity. These are the two types of marginal effects that we can get from the Tobit model.

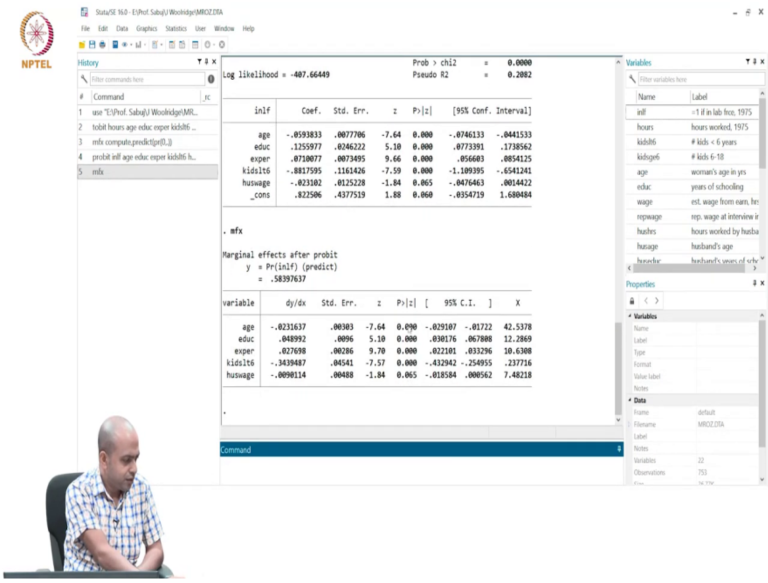
Now, so this is once you estimate the model, Tobit model; two types of marginal effect as we said we can get and for that two types of specific command is required and the command is this mfx compute predict pr. So, the command is little complicated but if you understand the logic mfx compute means anyway I am interested in computing marginal effect. After that we are asking Stata to first predict what is the probability that woman participate in the labor force.

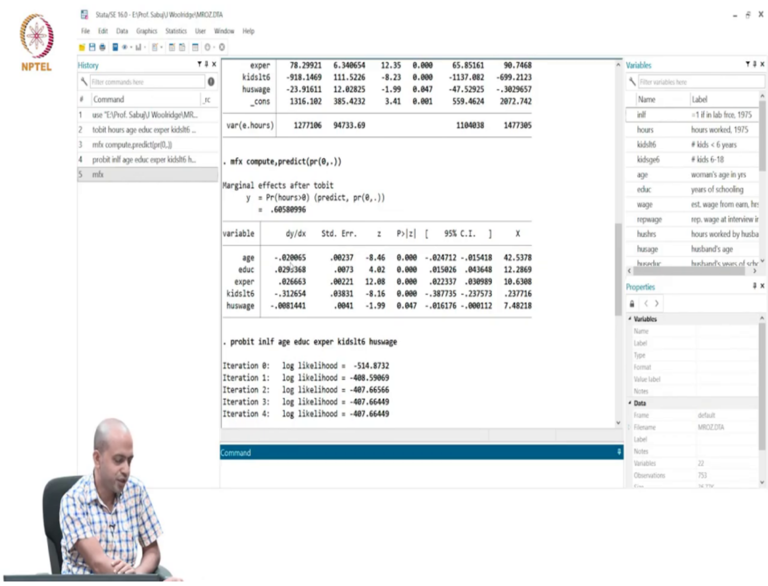
That why I predict pr probability and in the bracket 0 comma dot, it means greater than 0 that means positive labor supply. So I am asking Stata to first predict what is the probability of labor force participation and then you can actually calculate a change in probability from that base point. So, this is the command for estimating first type of marginal effect and if you put into this is the command, this is the output. So now I can say that for unit change in age, probability in labor force participation goes down by 0.02 units.

(Refer Slide Time: 22:15)




NPTEL
. probit inlf age educ exper kidslt6 huswage
Iteration 0: log likelihood = -514.8732
Iteration 1: log likelihood = -408.59069
Iteration 2: log likelihood = -407.66566
Iteration 3: log likelihood = -407.66449
Iteration 4: log likelihood = -407.66449
Probit regression Number of obs = 753
LR chi2(5) = 214.42
Prob > chi2 = 0.0000
Log likelihood = -407.66449 Pseudo R2 = 0.2082
inlf Coef. Std. Err. z P>|z| [95% Conf. Interval]
age -.0593833 .0077706 -7.64 0.000 -.0746133 -.0441533
educ .1255977 .0246222 5.10 0.000 .0773391 .1738562
exper .0710077 .0073495 9.66 0.000 .056603 .0854125
kidslt6 -.8817595 .1161426 -7.59 0.000 -1.109395 -.6541241
huswage -.023102 .0125228 -1.84 0.065 -.0476463 .0014422
_cons .822506 .4377519 1.88 0.060 -.0354719 1.680484
Command
mfx


NPTEL
Log likelihood = -407.66449 Pseudo R2 = 0.2082
. mfx
Marginal effects after probit
y = Pr(inlf) (predict)
= .58397637
variable dy/dx Std. Err. z P>|z| [95% C.I.] X
age -.0231637 .00303 -7.64 0.000 -.029107 -.01722 42.5378
educ .048992 .0096 5.10 0.000 .030176 .067808 12.2869
exper .027698 .00286 9.70 0.000 .022101 .033296 10.6308
kidslt6 -.3439487 .04541 -7.57 0.000 -.432942 -.254955 .237716
huswage -.0090114 .00488 -1.84 0.065 -.018584 .000562 7.48218

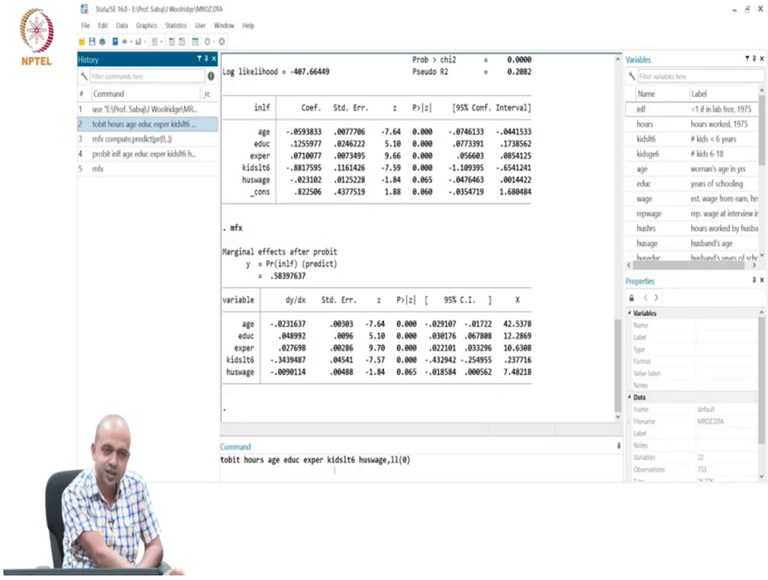

NPTEL
var(e.hours) 1277106 94733.69 1104038 1477305
. mfx compute,predict(pr(0,.))
Marginal effects after tobit
y = Pr(hours>0) (predict, pr(0,.))
= .60580996
variable dy/dx Std. Err. z P>|z| [95% C.I.] X
age -.020065 .00237 -8.46 0.000 -.024712 -.015418 42.5378
exper .026663 .00221 12.08 0.000 .022337 .030989 10.6308
kidslt6 -.312654 .03831 -8.16 0.000 -.387735 -.237573 .237716
huswage -.0081441 .0041 -1.99 0.047 -.016176 -.000112 7.48218
. probit inlf age educ exper kidslt6 huswage

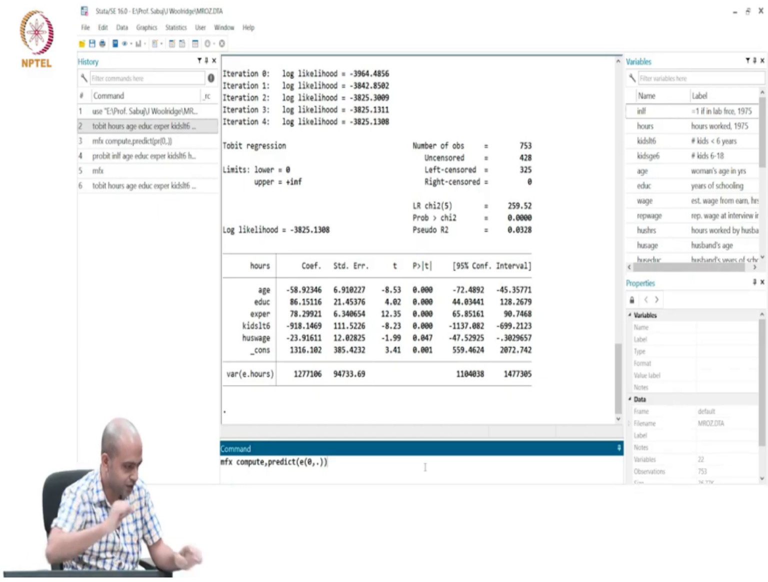
Now, if I compute the same model using Probit what will happen? So, the command would be Probit in labor force, then age, education, experience, kidslt6, and huswage. This is your Probit model and dependent variable of course, here it is hours but in the context of Probit it is in labor force and then this is the output.

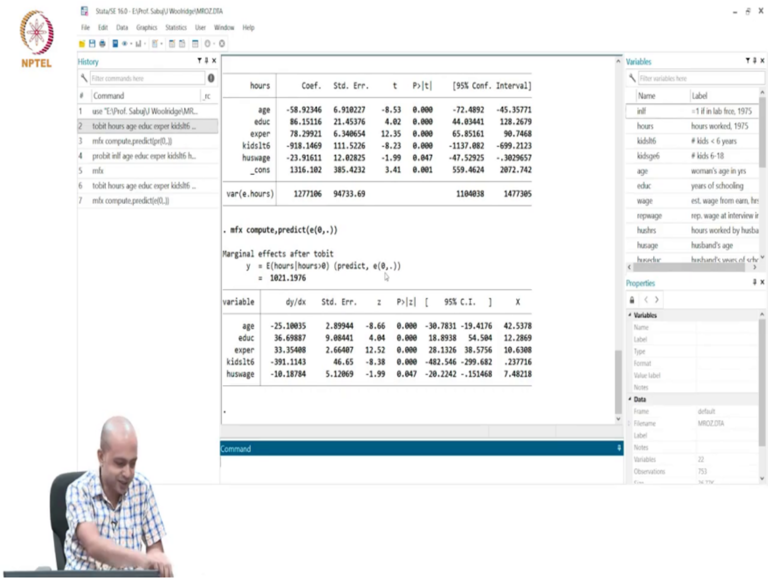
And after estimating these if you put mfx, look at the marginal effect of age. The marginal effect of age is -0.02 and the marginal effect you got from Tobit is -0.02. That means what I said what we discussed that, Tobit is basically Tobin's Probit, that means, the output that you got from the Probit model and the marginal effect is exactly matching with the marginal effect that we got from that Tobit model as well.

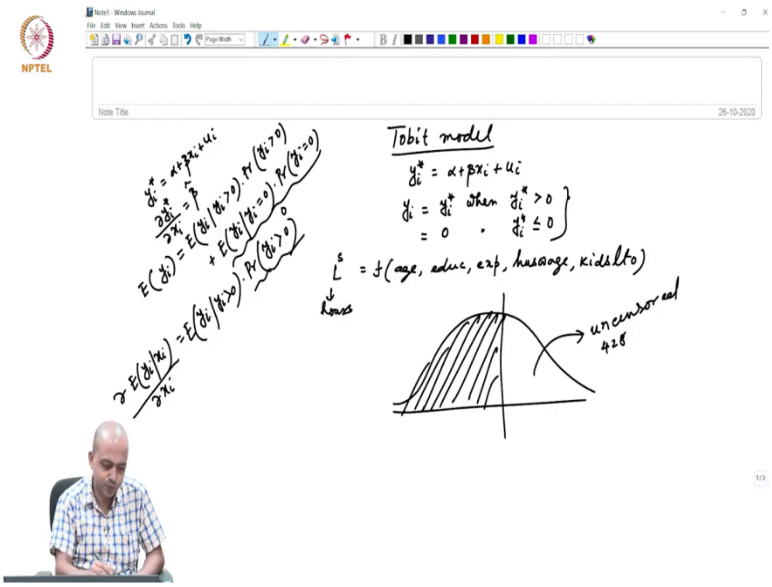
That is why we say that the first stage is like estimating a Probit model, whether the married woman will participate in the labor force or not, once they participate, then we can actually observe the positive amount of labor supply and then we can get the elasticity of that amount of labor supply with respect to any of your explanatory variable.

(Refer Slide Time: 24:12)









I will once again estimate the Tobit model. For that I am putting this command. The advantage of this Stata is that whatever command you are using, it will save here and if you simply click on that, it will directly take the command, you do not have to type the command again and again. So, this is the output once again and you have to say that mfx compute.

Now, I am not interested in estimating the change in probability rather change in labor supply. So, for that you have to say that predict, in the bracket you have to say that instead of pr you have to mention e, e for elasticity, this is the command. So, I am asking Stata to first predict that the labor supply amount is greater than 0 and then you estimate the change.

Look at here what Stata is first calculating, expectation of hours given hours greater than 0 and multiplied with the probability. That is exactly what we are discussing in our theoretical model because estimation is basically, sorry, yeah, look at this, expectation of y given y greater than 0 and multiplied with the probability. Hence Stata is also first calculating this is expectation of y given y greater than 0 and then this is the probability. So, that is how you can get the marginal effect of the Tobit model and you should now directly use the coefficients from the Tobit model as the marginal effect. So this is basically the Tobit model what we have discussed.