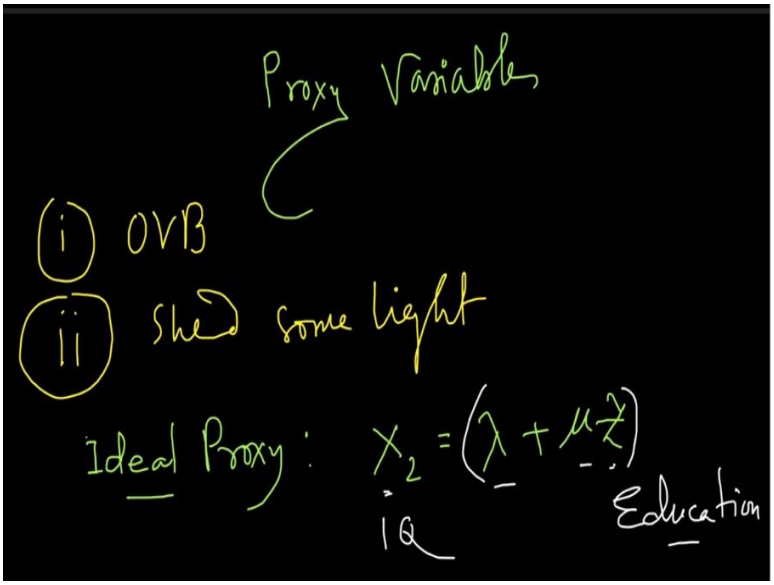
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology - Kharagpur

Lecture – 91
Continuation with Proxy Variable

(Refer Slide Time: 00:27)



Hello and welcome back to the lecture on applied econometrics. We have been talking about proxy variable and its usefulness. So we mentioned that the proxy variables will; there is a risk of having some measurement error, but even then we use a proxy variable because we cannot measure the exact variable of interest. So, if we completely omit the exact variable of interest, the problem that we face is that we might have to have this omitted vriable bias problem in our model.

Which will lead to invalid statistical tests which we actually want to avoid. Secondly, by including a proxy we can actually get some idea about the importance of the variable of interest. It can shed some light on the variable of interest. So, that is why we can actually know that alright so I do not have any measurement of IQ, but I can at least use education as a proxy for IQ. And I can see if the education is actually important variable in explaining wage.

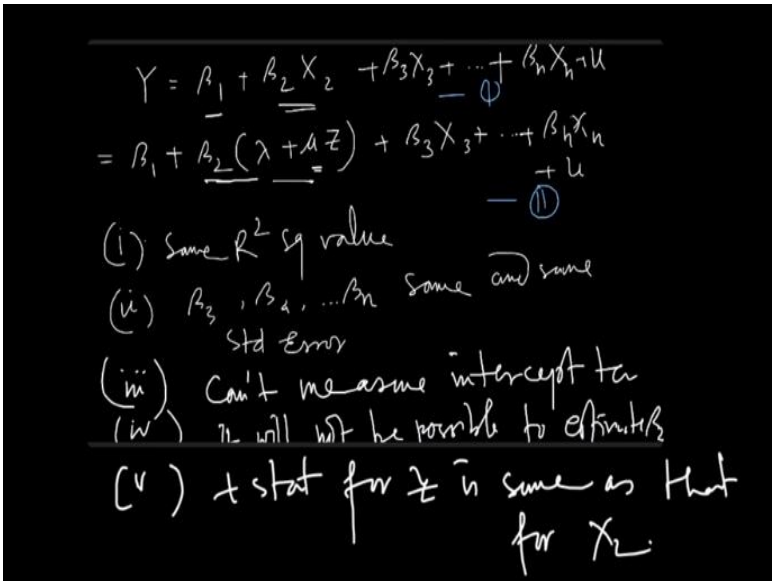
So, then IQ is perhaps a really important variable when I try to measure wage. So, let us talk about all the proxy variable in detail. So let us talk about a particular type of proxy variable

which we can call it as ideal proxy, which is actually very difficult to find, but let us start with the simple one. So the ideal proxy is something where I can let us say I have a explanatory variable X 2 and I want to represent X 2 with some ideal proxy.

And in this case, the ideal proxy will be related to X 2 in some linear fashion, alright. So let us say the Z is the proxy variable. Let us say the X 2 is the IQ and the Z is the education. And let us say the IQ and education are related linearly, which may not be the case in reality, but let us say this is the case where I can stablish a linear relationship. Of course, I do not know lambda or mu because I do not have data about X 2 IQ.

So I only have data on Z, so based on that I cannot estimate really lambda or mu 2. But what you can do is you can actually incorporate the whole equation we can substitute the whole right hand side into the actual regression equation.

(Refer Slide Time: 02:40)



So, let us say my regression equation here is Y = standard regression equation beta 1 + beta 2 X 2 where I cannot measure X 2 here and then I have a beta 3 X 3 and then I have up to let us say beta n X n and there is some error term u. I do not have this, so what I do? I can substitute this X 2 with my proxy. So how I will do that? So beta 1 + beta 2 into lambda + mu Z and then I have beta 3 X 3 and other terms beta n X n and u, the error term.

Now with ideal proxy, we can say the model is correctly specified. If we have this ideal proxy, so what you can have is the following. One is that we can have the same R squared value for this regression. So, let us say I name this regression equation, this is equation 1 and

equation 2. So, it does not matter if I run equation 1 or equation 2, I will have the same R squared value.

Second is that the coefficients and the standard error. So let us say beta 3, beta 4, and other coefficients that you are measuring beta n they will be same and same standard error. We cannot have the intercept term, we cannot measure the intercept term because here the intercept term is going to be beta 1 + beta 2 lambda and now since I do not know lambda, so I cannot measure this intercept term and that is fine.

Point number 4 is that it will not be possible to estimate the value of beta 2 because here what I am going to get is beta 2 into mu, it will not be possible to estimate beta 2 because it is associated with mu which is unknown. It will not be possible to estimate beta 2. It is going to be same as the t stat for X 2 because essentially we do not want the significance test for Z 2 and X 2 to be different.

So, basically t stat for Z is same as that for X 2. So, essentially, if I have an ideal proxy, all these conditions will be satisfied and we do not have any problem with bias or standard error, which could be a problem in case of measurement error. So, the moot point is if we have an ideal proxy, so we do not have to worry about the problem of bias or standard error.

But oftentimes, we really do not have the ideal proxy and that is basically the case in reality. And in those cases what we have is called imperfect proxy. And what is imperfect proxy? An imperfect proxy is something where the proxy is related to the variable of interest in nonlinear fashion, it could be logarithmic way or quadratic way or any other fashion. Now, most of the cases we have imperfect proxy and we need to find out a way to actually address the problem.

And in this cases of imperfect proxy, we can have the problem with bias or standard error. So, we use something called instrumental variable that has really great significance in economics and econometrics research. So, we will talk about instrumental variable in the next lectures. So, with this, we will end this lecture on proxy. Thank you.