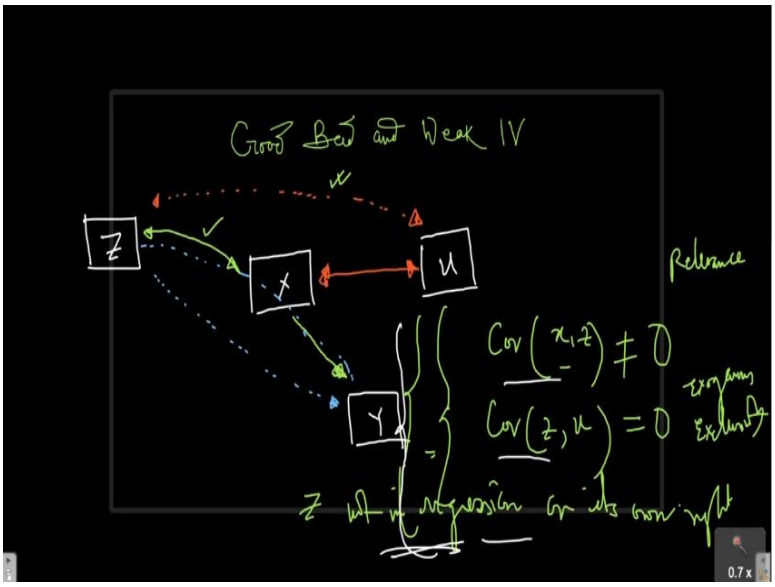
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology – Kharagpur

Lecture – 104
Good Bad and Weak Instrumental Variable

Hello and welcome back to the lecture on applied econometry.

(Refer Slide Time: 00:28)



We have been talking about instrumental variable and in the previous lecture we spoke about endogenity problem and how instrument can actually come to rescue any problem of endogenity. Now, we were careful about the fact that an instrument could be good, bad or a weak instrument and we need to be very choosy about an instrumental variable, if we are really planning to use an instrumental variable in the research.

And in a given context of research, an instrument could either be good or bad or it could be a weak instrument. Now, we need to be able to identify, if an instrument is good or bad or weak in a particular research. And we will see there are certain ways of course we have some idea by now. But we will outline the theoretical framework to identify, if an instrument is good bad or a weak instrument.

So, let us try to understand graphically the relationship between x y z and the error term u. So, if I have my x here and let us say I have my y here and I have my error term u here and let us say I have a z instrument which is here. So, I will try to identify the relationships which

are desired which are not desired. Now, I want let me use different colors to notify that. So, I want x and y to have strong relationship.

So, I am using a green color. And let us say I do not want so, you know I do not want any relationship between x and u. So, I am basically marking it as red but it is present, if I have endogenity. So, I want to actually avoid this. Now, I want to use an instrument. So, when I want to use an instrument so, there should be strong relationship between this z and x. There should be a strong relationship between z and x when I am using them as an instrument.

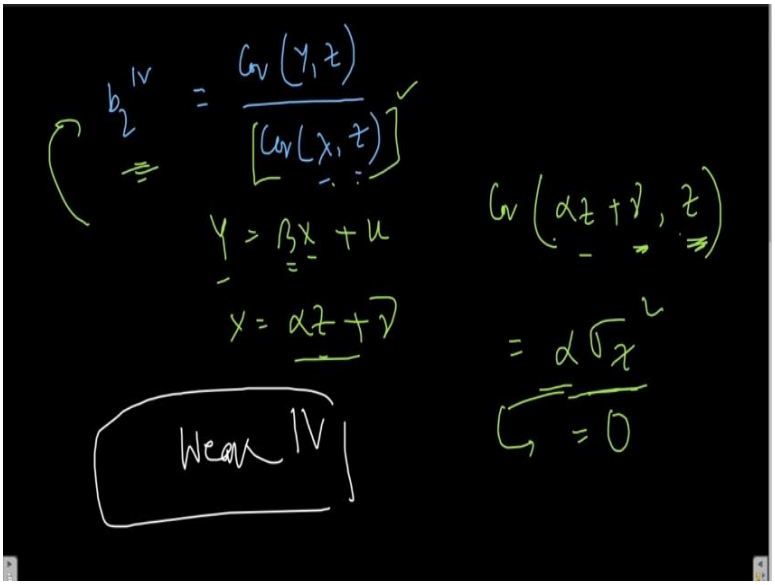
Now, what? I do not want any relationship to exist between z and u. Now, I use a red dotted line because this relationship does not exist. So, I do not want. So, it is a dotted line but it is red because that is undesired. Now, so far is the relationship in y and z is concerned. So, I know that z is going to influence y but through x. It is the entire effect of z is to be channelised through x. That is the desired condition.

So, essentially let me draw it. So, I can draw something like this. It is a blue line so that it basically indicates that there will be correlation but that effect will be channelised through this. Essentially that is going to be channelized through this. So, with this we sort of get the criteria. The criteria we have for a good instrument is that covariance of x and z should not be 0. That is this relationship has to be strong.

And covariance of z and u that should be equal to 0. There should not be any relationship between u and z. And the third criteria that you said is that z should not be in the regression equation on its own. So, we will write z 0 in regression on its own. Now, the first criteria here when I say that x and z has to be strongly correlated so, this is basically a criteria for relevance and the second criteria is basically the criteria of 0 exogenous.

And also, it is exclusivity also because the entire the effect of z is not channelized through u but this analyzed through x only. So, it is exclusivity also you can say. Now, let us write down the expression for the instrumental variable.

(Refer Slide Time: 04:48)



So, let us say beta 2 I V here is going to be covariance of y z divided by covariance of x z. Now, let us say the relationship between covariance y z is very weak and let me actually write down the very simple equation for instrumental variable. Let us say my y, let me use a different color, my y is equal to beta x plus some error term u and my x is equal to some alpha z or some another error some gamma.

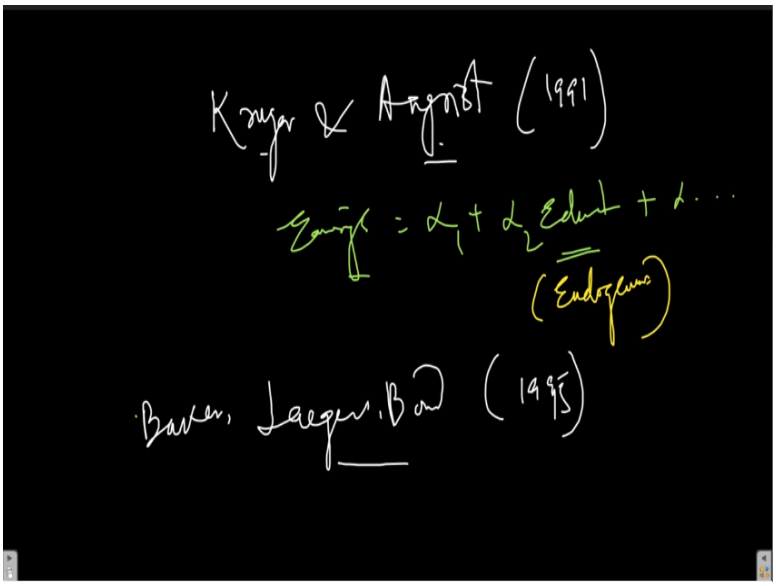
Now, here z is the instrument and x is the independent variable that explains y. So, in this context in this equation context I can write this beta 2 I V or which is basically an estimate of beta is basically this. Now, if I expand this, let me actually expand this covariance x z. So, covariance x z is going to be, if I substitute x with this alpha z gamma, z. So, what I will get is basically alpha sigma z square.

Now, if this alpha term is equal to 0. So then what I am going to get is that the covariance this term, if I expand it I will this covariance between gamma and z that is going to be 0. So, essentially I am going to have only this term and if my alpha coefficient is 0. So, it is going to give me a 0. So, what I am going to have in the denominator is 0. So, if I have my denominator 0. So, b 2 I V is going to be undefined so that is something I do not want.

Even, if I have my b 2 I V is actually very small. So that is going to give me even, if this covariance x z is going to be very small so that is going to give me a very large value of b 2 I V. So, which essentially would mean I am going to get a large variation of b 2 I V which I do not want essentially. So, this is the problem when I have my the problem when I have this relationship between x and z weak. So, this is an example of weak I V.

So, we do not want any weak I V in the regression, if I am really going to use an I V. So, I do not want to weak I V and one example of weak I V could be.

(Refer Slide Time: 07:22)



So, one particular that is actually a very famous research by Kruger and Angrist. So, these two were the and many other econometricians but Angrist has a lot of contribution, Kruger also has a lot of contribution in so, far as I V regulation is concerned. So, they published a paper in 1991 where they actually tried to estimate the effect of schooling on earning but only this time they thought that the education or schooling is an indigenous variable.

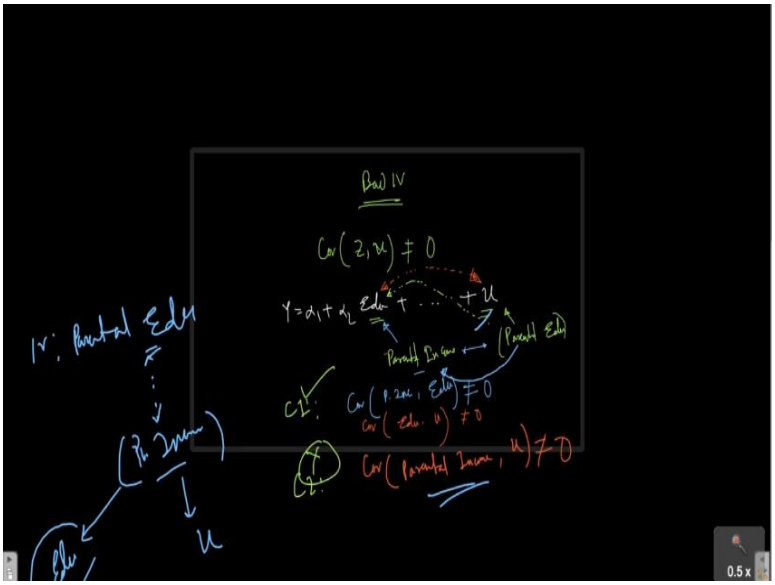
That is not an exogenous variable that we have been doing all throughout. So, if earnings is a function let us say alpha + alpha 1 + alpha 2 education and many other variable. But you know that is how, it is related but if this education is not exogenous. So, they said their endogenous variable. So, what they did? They actually took an instrument and the instrument is the year of birth or quarter of birth rather. So, the idea was that depending on the quarter of birth.

So, people will have the students will have or the children will have a specific age to cut off age to join a particular class. So, essentially they will have different age group differences like people who are you know born in the first quarter, second quarter, third quarter. So that way they wanted to see, if they can actually find the true effect of education. Unfortunately so that was not considered to be a very strong I V.

Because that yeah that quarter apart explains very tiny proportion of the variation of schooling. That was a research that was done later on by the name is Baker Regina, Baker Laeger and I think so that was in 1995 paper. So, this paper actually sort of brought out the fact that the I V they have used that is a weak I V because the quarter of birth actually explains very little in the variation of schooling.

So that is a example of week I V. So, perhaps we will try to avoid the cases of weak I V. Now, I will talk about a bad I V.

(Refer Slide Time: 09:41)



A bad I V is something where the relationship between covariance of z and u is not equal to 0. Essentially, if this relationship is this part of relationship that is actually there is some sort of relationship. So then that is not that is actually a bad I V. So, we will again try to avoid that. Let me just give you an example of bad I V. Let us say, just like the previous example education, we know that education is a endogenous variable.

And let us say I have this equation again y = alpha 1 + alpha 2 education and let us say some other variable and there is some u term. Now, I know that education is an endogenous variable instead of quarter of part I have chosen let us say parental education or parental income. Let us say parental income. Now, I know, if my parents are rich so, they are they are going to send me to good schools.

Now, parent of income may appear to be a good instrument because parental income has a strong relationship with education. So, I can write covariance of let us say parental income

and education is not equal to 0. So, the first criteria is made and I can consider that this is to be a good proxy. But there is a problem here. The problem is parental income is just not related with my education but there are so many other factors that parental income is related with.

And let us say, it is parental education. So, parental income is basically an outcome of parental education. So, parental education and parental income are related and then what is happening is the fact that the, I am not including parental education. So, parental education is getting into the error term. At the same time parental education is something that is influencing my education.

Because if my parents are educated I am likely to be more educated because they are going to take care of my schooling, my home education and other aspects of my learning. So, because of that because parental education is now, into the error term and this is actually influencing my education. So, what is going to happen is exactly what we are afraid of the relationship. There is a correlation between error term and the education which I actually wanted to avoid.

Because I wanted to avoid the endogenous problem. So then the equation the condition that covariance of education and error term should be equal to 0 or covariance of here, if I use my proxy. So, parental income. So, this is not equal to 0 of course and of course this is not equal to 0 because of the fact that covariance of parental income and the error term is not is equal to 0.

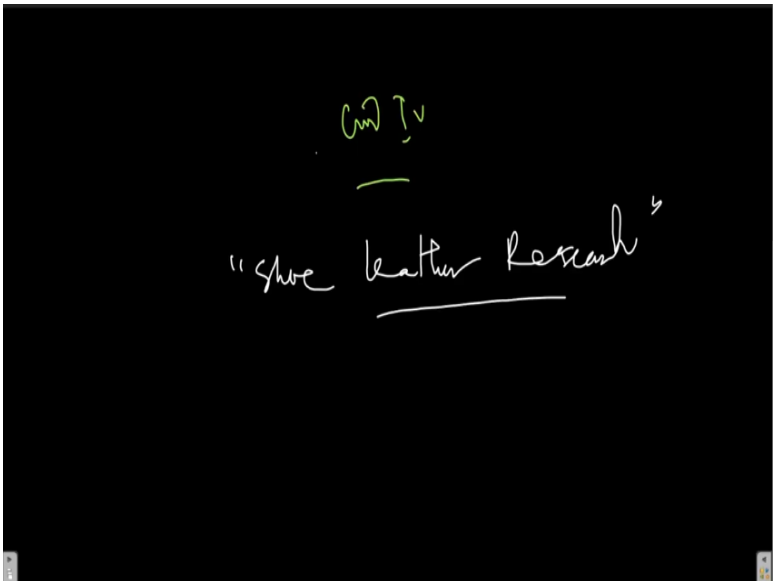
So, my condition one is fine but my condition two is actually violated, this is violated. So then it is a bad proxy. And you can think of what, if I actually take parental education instead of parental income? You can actually think of that. Now, actually let us think about that. So, if I think parental education instead of parental income well. So, parental education is of course influencing my education so that I have seen.

Now, whether that will actually solve the bad I V problem is essentially, it will not solve the bad I V problem. Because parental education is related with other factors and like parental income. So, parental education is influencing parental income here. So, let me write down. So, my I V here, let me write down, let us say parental education. Now, parental education is now, going to be related with parental income P income.

Now, because I am not considering parental income into this equation. So, what is happening is that the parental income is influencing the error term. Now, parental income is another variable that is also influencing my education because my parents can afford my education. So, they can send me to private schools, they can send me to other you know teachings and good they can provide me good books and everything.

So that is basically influence my education. So, the same way like we have a problem with parental income as my instrument, I can have the same problem with parental education as my, if I consider that as an instrumental variable. So, these are the examples of bad I V.

(Refer Slide Time: 14:39)



Now, we can think of good I V whereas when it comes to good I V we essentially need to identify or ensure that these three properties of a good I V are satisfied. These three properties are satisfied. Now, one example of a good I V could be this 1947 reform that was brought in in UK when the minimum age of schooling was legislated. Now, so far this mathematical theoretical knowledge on I V is fine.

But this is not enough to actually do a research with a good I V. You actually need to find out a good I V in the first place. And how which I V could be a good I V to sort of get an intuition behind it? You need to do a lot of institutional, you have to build a lot of institutional knowledge. You need to do a lot of reading, a lot of talking and actually visiting sites and so forth.

To actually identify what intervention was brought and what is possibly the impact of those interventions and in a word you can say finding a good I V is basically doing shoe leather research. You need to do a lot of shoe leather research to actually find out a good I V. So, with this we will end this lecture here and in the next lecture we are going to talk about couple of more properties of an I V regressor that you need to satisfy in order to sort of have a good I V. Thank you.