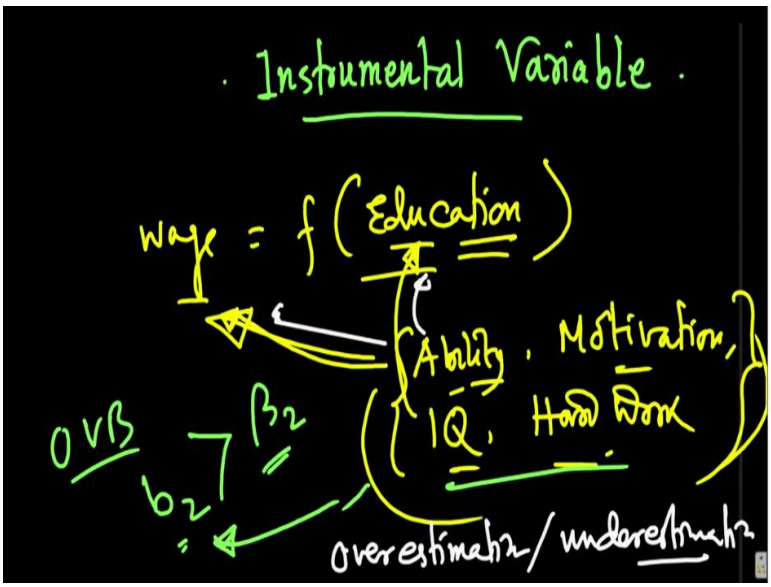
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology – Kharagpur

Lecture – 98
Instrumental Variable

(Refer Slide Time: 00:35)



Hello and welcome back to the lecture on applied econometrics. Now, today we are going to talk about a very important concept in economics and it is very dear to economists and that concept is called instrumental variable. And why it is so dear to economists because with instrumental variable, it is something that actually comes from a natural experiment and we economist really cannot do any experiment because everything is in the natural settings and you cannot do an experiment.

It is not something you can do it in the lab or like in other science disciplines. But instrumental variable that something comes from the natural experiment that actually mimics almost what we do in the lab and that too in the natural setting, so that has a really significant value in economics research. But the problem with instrumental variable is very difficult to find a good instrument.

I will explain all of these attributes of instrumental variable, but first let us go back to the previous lecture when we talked about model specification and we talked about the error term and measurement error. And the measurement error was we said that one way of actually

addressing the measurement error is to use proxy and we said that there is something called ideal proxy, which can actually help us to get rid of some of the errors, measurement errors.

But in most of the cases we do not get an ideal proxy, we have imperfect proxies. Now, instrumental variable is actually something if we can satisfy some criteria that could be excellent proxy to address the measurement error. And I will actually talk about instrumental variable with example that we have been continuing with. There is a relationship between education, wage, and other variables like experience.

And other dummy variables that we actually use in the Mincerian wage regression equation. Now, this example I have already explained, now let us say my wage is a function of my education and there could be other variables also. Now, the thing is that we are using education because we do not have any data on many other related variables, which actually can influence wage and let us say which could be determined by things like my ability or it could be by motivation.

I am really motivated guy; motivation let me correct my spelling motivation or it could be determined by something like let us say, I do not know, let us say my IQ or let us say my hard work; there are so many different things that can actually influence my wage. So by nature I am a hardworking guy, by nature I am a sincere person, by nature I am extremely diligent to my work; no matter what I do.

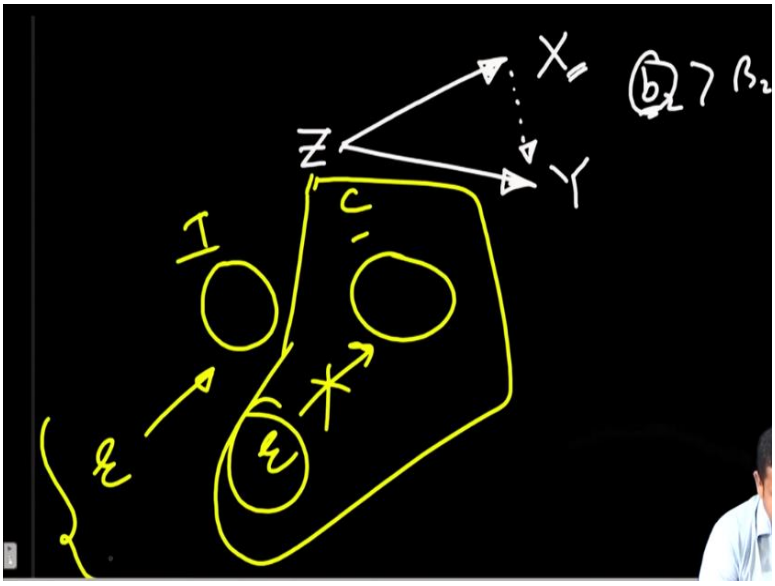
So, I can have so many different attributes. Now, when it comes to education, these factors are also playing a role. So the fact that I have done my primary schooling, the fact that I have done my secondary schooling, my class 12, the engineering and so forth. So all of you, many times you might have felt what is the point in this education and so forth, but you still continued your education.

And that is because of these factors you already have, you already want to do so many good things in life, so you continue with your education. Now, these factors are it is quite likely not just the degree, the education degree that you get, but the factors responsible for achieving those degrees are actually the same factors which are actually giving you a wage, better wage. I do not need to draw this arrow, I already have.

So the same factors are actually helping you to get a better wage. So, education in a way then actually masking the effect of all the different variables and we have already talked about this when he talked about omitted Variable bias. We said that your how, your b, the estimated coefficient of education let us say b 2 could be actually be overestimate beta 2 because your b 2 is actually capturing all these influences.

So we said that depending on what are the factors, the strength, the direction we can actually identify the cases of overestimation and underestimation. We had a detailed discussion on that. But the point is that there could be some other variable, which are actually responsible for both of these.

(Refer Slide Time: 05:23)



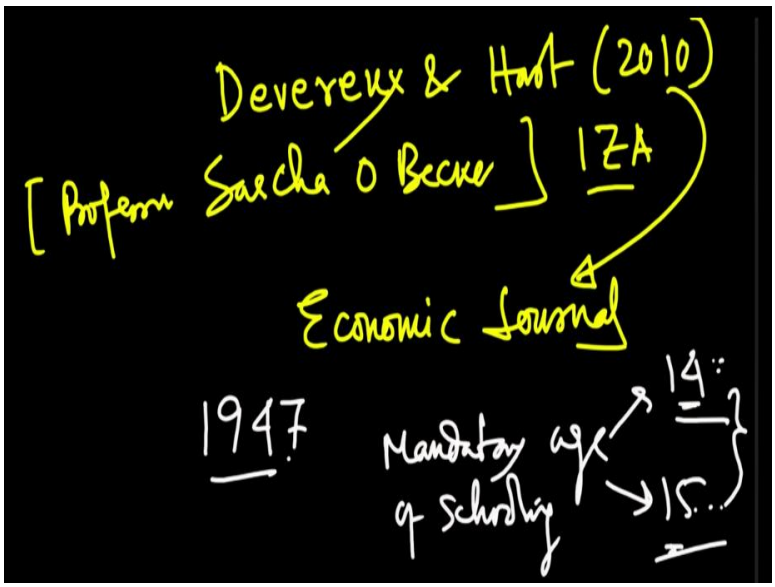
So, if I can write it like this let us say this variable is known as X, this ability, motivation, hard work sincerity in all these areas let us say this is Z. And let us say this leads to my variable X which is my education or it may also directly influence Y. And because I do not have Z, so when I measure the effect of X on Y, some part of Z is being actually captured by X and this is basically a case of overestimation of the coefficient b that determined there is an overestimation of actual beta 2, b 2 is an overestimation of beta 2.

Now, this is the problem. Now how do we really sort of address it? Now one way to address it is actually that is the best way perhaps is to you decide you basically choose people random. The moment you choose people randomly, you actually take care of their other attributes and we will talk about that in next few lectures. Now, when I say randomly, so basically you have two groups; a treatment group and a control group.

And you have those people selected randomly and you give education to one group and you do not give education to other group, you simply do not give education to other. But then that is a problem, you cannot do that on ethical grounds. I mean, it is everybody's life, right. So you cannot take them back to past and again you can change their decision on education, so once done is done.

So, these people will be deprived of education and if you want to see the lifecycle earning just to understand the impact of education that is like too high cost to pay. So we cannot do that. So if we have a good instrument, actually that can mimic this whole experiment that we actually want to do and that example I am going to give.

(Refer Slide Time: 07:31)



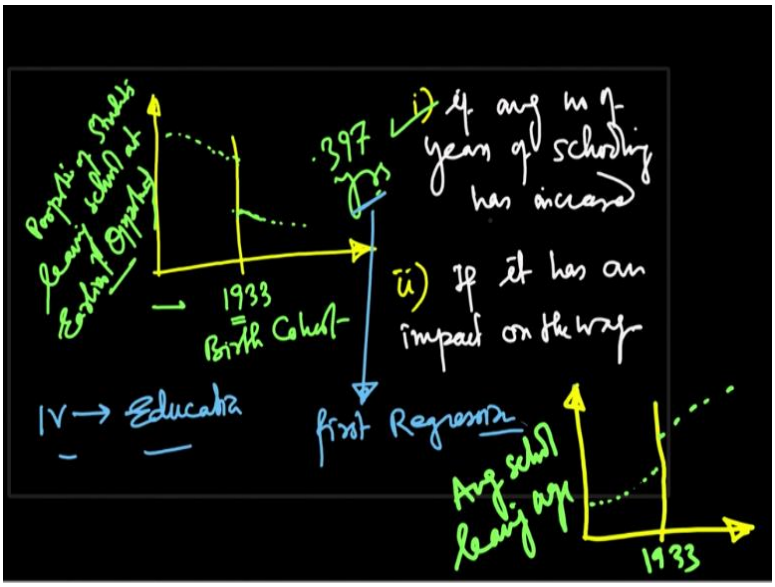
Now the example I have taken from this paper by Devereux and Hart that is really a nice paper and there is a very good sort of blog by Professor Sascha O Becker. You search this, you can find it in the internet. This is a blog posted in IZA website, so you can search it, we can provide you the link. So coming back to the paper, this is published in economic journal and this paper actually talks about reform that was undertaken in the UK in 1947 and that reform was mandatory age of schooling.

So mandatory age of schooling initially it was, but after this reform the mandatory age of schooling became 15. Now, that is sort of a way to you if you increase the mandatory age, now after 1947 people have to stay back in school even if they want to drop out. Let us say in

1947 people dropped out at the age of 14, but after that those who want to drop out they have to stay back in school for one more year.

So, then I can actually see if, so let us say if this factor of legislation is the only responsible factor for one additional year of schooling so that would mean that essentially I will be able to estimate the true effect of education, true effect of one year of education on wage because it is something forced from outside.

(Refer Slide Time: 09:53)



So, to actually explain what happen is that so what they have seen is that they have actually got those results. So, they actually saw this in the context of students stay in schooling, so number one if that mandatory. So basically, we asked two questions if the mandatory, the provision of mandatory schooling has improved the retention of the students in the school or if the average number of years of schooling has increased? So that is one and the second is; so let me write it down.

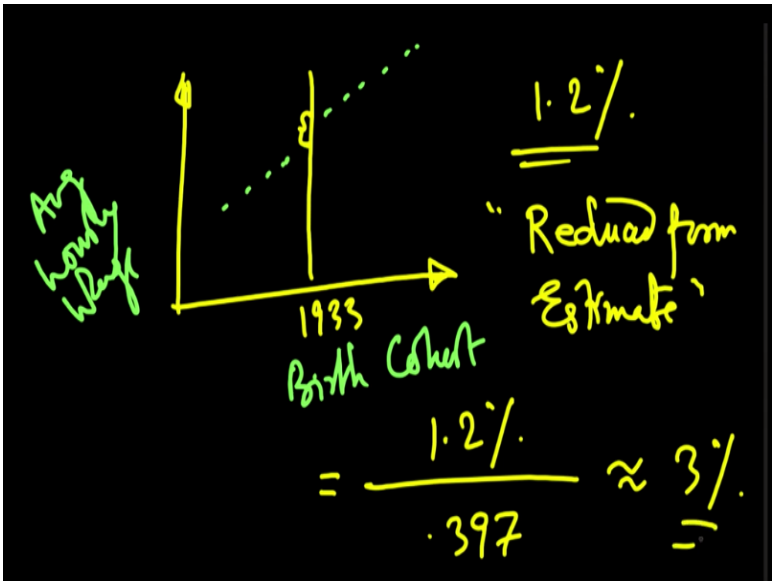
If average number of years of schooling has increased? And number two if it has an impact on the wage? So, to answer these two questions what they have found out is that the fraction of students who are actually leaving the school at the earliest opportunity, so let us say this is my I just explain, so this is a point where I can see this sort of jump here and this is year 1933. Now, note that it is 14 years of mandatory schooling, it is like basically mandatory schooling till age of 14 and mandatory schooling till age of 15.

So, if this legislation was brought in 1947 and if I go back 14 years, so essentially I will get 1933. So, it shows that and if my Y axis is actually proportion of students leaving school at earliest opportunity. This is the proportion of students leaving school at earliest opportunity and this is the cohorts, birth cohort. So essentially, I see immediately after 1933 the school leaving has declined drastically, so that is effect one.

So, I have seen that the average years of schooling has, retention of schooling has increased and that number is actually 0.397 years. And in fact, this is called first regression. So, what the first regression is basically? If my education is the treatment, my instrument variable is the legislation, so if the instrument variable has an impact on my education? So, I can see that it does have an impact on education that the retention to school has increased by average for the male students.

It increased by 0.397 years, so that is my first thing that I got some answer on. And also corroborating evidence to this is that let me have on the same page, so when I look at the average school leaving age so what I found average school leaving age it actually increased after 1933. And I have the same 1933 here that it has increased. So, clearly see the impact of instrument variable on the treatment variable that is my education.

(Refer Slide Time: 14:09)



Now, coming to the second point, so they actually wanted to understand if the wages has increased, and what they found is actually the wages have also increased. If I draw this, let us say this is average hourly wage and this birth cohort and for this also they have seen something like this whereas you have year 1933. So, there is a jump here. Now, to explain

this part, so essentially average earning and they actually saw this growth is 1.2% average earning and this is called, the second estimate is called reduced form estimate.

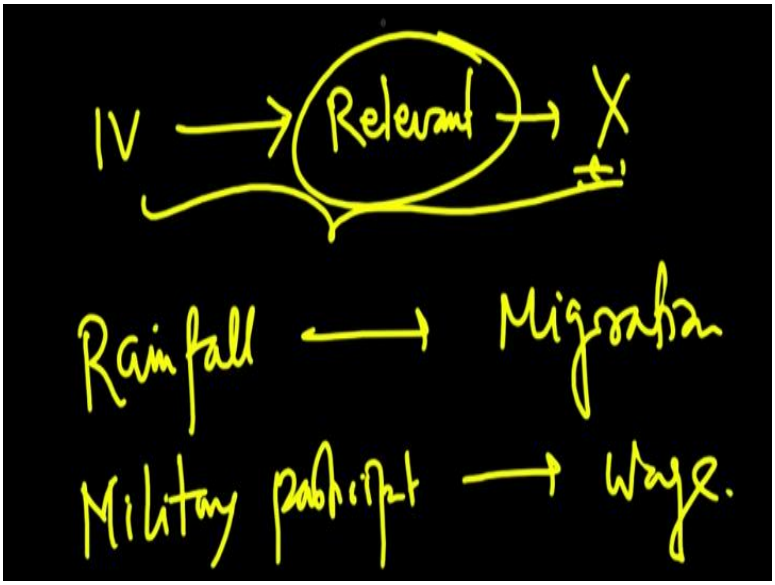
Now, you see because of the instrument, because of this intervention, what has happened? Number one, the average number of years of staying in school has increased and at the same time basically average hourly wage has also increased. Now, I want to understand what is the impact of education on earning that is my research question irrespective of other variables. So, when I actually am looking into the impact of the instrument, I actually consider that the cohorts just before 1947, 1933 and 1934 they are just the same.

So, we do not really consider them to be different, we consider them to be identical. So, these are the some of the assumptions we make. We also consider the labour market conditions to be same. So, we do not take all these into account, but what we take into account is that we see that because of the instrument I actually see the school retention has increased by some X number of years and I also see the wage has increased by some Y percentage.

And my main research question is to understand the pure impact of schooling on my wage. So, what I do is I basically take a ratio of reduce from estimate by the first regression. So basically by 1.2% and here I have 0.397, essentially per unit year of schooling how much is the return, so essentially that gives me something like a value of 3%. So, this is something they have found and this is the beauty of having an instrumental variable.

So, keeping everything else constant, given this intervention, what is the impact and that actually gives me some result that could be quite convincing. Now, basically the intuition here is that given everything else is constant, it is your education that has improved the earning. So, there are a few things that we have to keep in mind and the few things and that is something that I am going to explain in the next lecture.

(Refer Slide Time: 17:27)



And the few things are basically the effect of instrumental variable, it is actually relevant for my X variable of interest, so it cannot be something totally unrelated and I found some result, so then that is not going to work. So, you have to find out something that; here it is the legislation and its impact on schooling.

It could be let us say rainfall, there are some researches, rainfall and its impact on migration and there are so many other instruments where the one very famous study is of military participation on wage and so forth. So, essentially, we need to have a relevant instrument for our model.

(Refer Slide Time: 18:33)

The second point is that exclusivity we call and that is the effect, the impact of the explanatory variable or the impact of the IV has to exclusively be channelled through X, it

should not be channelled through some other route to Y, so it has to be exclusively channelled through X. So, these two conditions; relevance and exclusivity they are very important. And in the next lecture, I am going to actually talk about my instrumental variable with its mathematical form where we can take all these restrictions into account. So with this, we end this lecture here. Thank you.