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

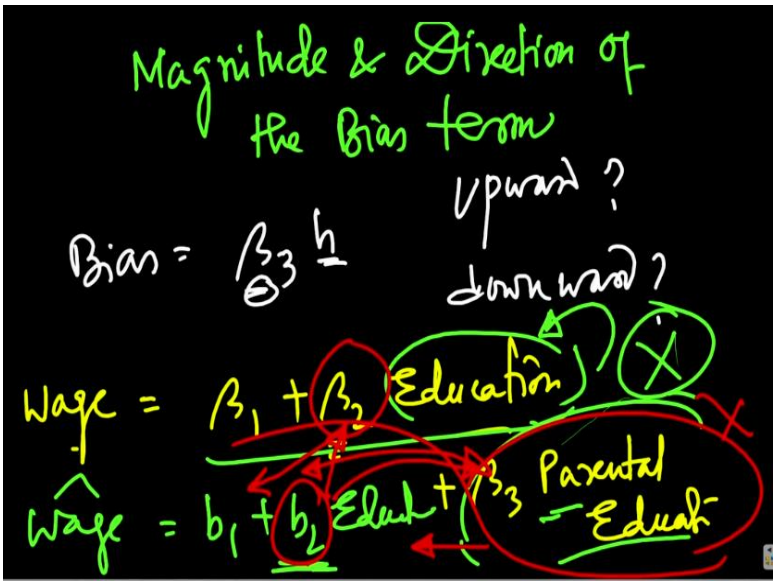
Lecture – 85
Model Specification - Continued

Hello and welcome back to the lecture on applied econometrics. We have been talking about model specification. Within model specification, we have been talking about omitted variable bias and within omitted variable bias we are talking about exclusion of a particular variable, particular relevant variable and we try to see what happens if the particular relevant variable is excluded and we have kind of estimated the extent of bias that we will see if that particular relevant variable is excluded.

Of course, you need to remember here we are basically dealing with a very simple toy model. We are only talking about two variables, two explanatory variables in our model, the true model and whereas the after the omission of variable we have a simple OLS. So because the simplicity of calculation are relatively easy, ease of calculation we are using this kind of model.

But if we have let us too many explanatory variables then it is really difficult for an econometrician to calculate this bias and all other estimations we are going to see. So it is basically to get an idea what happens if we actually exclude a particular relevant variable.

(Refer Slide Time: 01:30)



So, le tus say in the previous lecture, we spoke about the extent of bias and bias we said is beta 3 into h, if we run a regression between the two explanatory variable h is the correlation coefficeint or the regression coefficient and beta 3 is the regression coefficient in actual model between the true X variable and the Y, the dependent variable. Now, we need to understand what are the different cases.

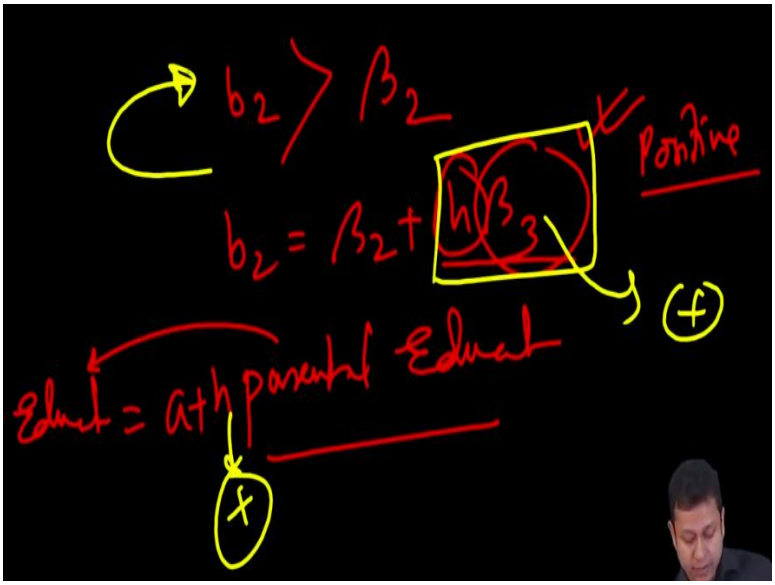
So whether the bias is going to be upward or downward. and what is the extent of this bias, upward bias or a downward bias? Wheat is the extent of this bias? And we will write down all the different possible examples, with some examples we will try to illustrate that. So let us say we are talking about wage and let us say we have this equation beta 1 + beta 2 education, so we know that education has a positive correlation. So wage and education, they are positively correlated.

So education, beta 2 is going to be a positive, it will be a positive regression coefficients, the signs will be positive. And let us say we have parental education. So, we know because of parental education because our parents had education, so they knew the importance of education and they actually invested in our education. So, actually that helped us to get more education and that actually helped us to get a higher wage.

Now, if in our model, we actually end up excluding that lets say I have excluded that part in our model, so what we will have? We will have only beta 1 + beta 2 into education. Now because it is excluded, the parental education that component will be sort of captured in the education component. So, the education variable actually take some part of the parental education, it will mimic some part of parental education.

And the bias that we will have that will actually if the b 2 or this is a true model and if our estimated model is this b 1 + b 2 into education. So, what will happen this b 2 is actually taking into account this B 2 and at the same time this component here because in this true model, I actually did not take into account parental education in my actual mode or the estimated model. So, what is happening in the student model this b 2 is actually representing this as well as this. So, then I say the b 2 is actually upward estimation of B 2.

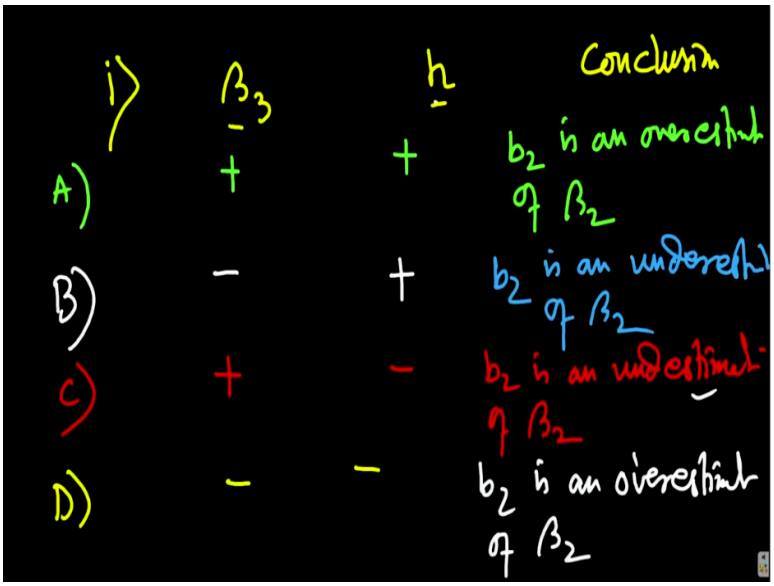
(Refer Slide Time: 04:34)



So, essentially it means that my b 2 is actually greater than B 2. Why because b 2 here is representing B 2 + h B 3 and here by h B 3 in this case, now this b 3 we are saying this is a positive quantity, but there is also this age, we need to see what is the sign and magnitude of age. So, since my parental education is if I look at my parental education and children's education, let us say the education of the children I basically do a regression and let us say I have a + h parental education.

So this parental education is also influencing the children's education positively. So then the h is going to be positive, alright. So then I have B 3 positive, I have h positive, so essentially this whole component is going to be positive and that is why my b 2 is going to be bigger than beta 2. So that is case 1. So I will write down this case is now.

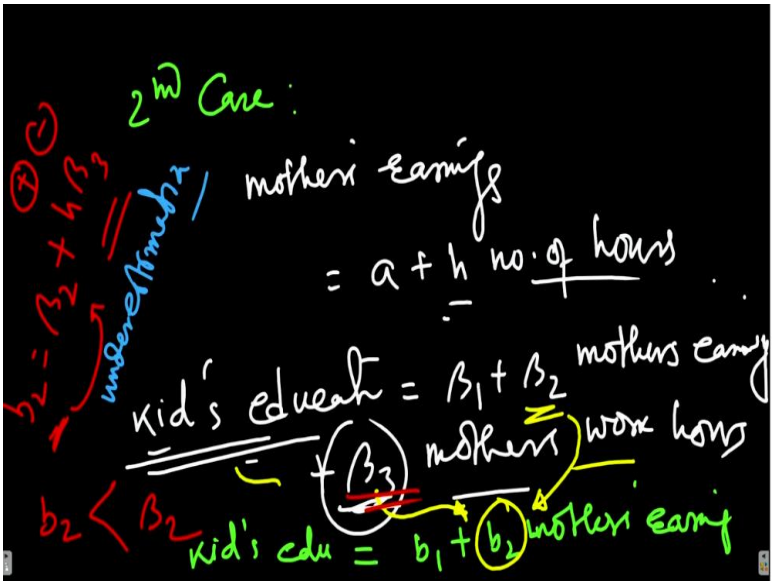
(Refer Slide Time: 05:45)



Case 1. Let me write down all the different cases. So let us say I have my, here I will write down the value of B 3 and then I write down the value and direct sign of Y and the conclusion what happens because of that. So in the previous example, that is our case 1, we have seen B 3 to be positive, h to be positive. So b 2 is an overestimation of beta 2. Now we will talk about the second case and I will again explain it through some example.

Let us say from some exam case, let me actually not write down. There is the first case, we will talk about a second case later. Let me actually first illustrate the example before I actually write the second case or make let me use the next page to explain the second case, then I will come back here.

(Refer Slide Time: 06:45)



The second case. Let us say I take this example where I have let us say I am talking about working class mothers, they are working. Now what happens if let us say my mom is a working class mother and she needs to work. Her earnings is basically proportional to the her number of hours she is working. So number of hours, so let me write down like mother's earnings is equal to a + h into number of hours, where h is going to be the positively correlated.

Now, let us say I am talking about kid's education. Kid's education is equal to I know that this is going to be, let us say the equation is going to be beta 1 + beta 2 into mother's earning because if the kid's mom is actually working and she is earning, so she will invest that money for the kid's education. So, they are positively correlated, whereas there is another component

if the mother is actually working for long hours outside and she is outside the home, so she will not be able to pay attention to the kid's education.

So, what will happen is that the beta 3 that is going to be negatively correlated so that that will have a negative sign basically, not beta 3 will be negatively correlated but mother's work hour let us say, mother's work hour let us say, so mother's work hour is going to be negatively correlated with kid's education. And that is why this beta 3 will have a negative sign. So, this is a case where in the previous example I have.

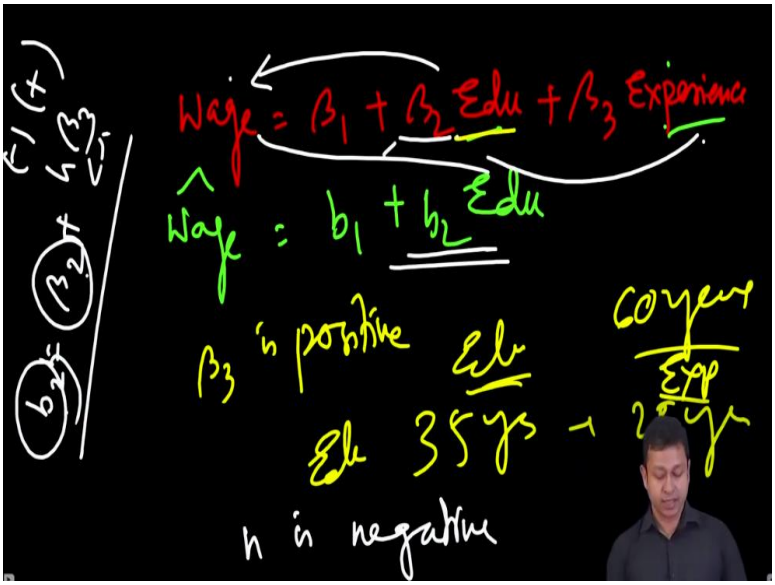
In example two I have beta 3 negative because number of hours the mother is working is actually negatively correlated with kid's education, but h is positive because of what? Because of the fact that the mother's income and mother's number of work are positively correlated. So, in this case what will happen if my in my regression equation, I actually omit my estimated regression equation, so kid's edu is equal to b 1 + b 2 into mother's earnings.

So then in this case what will happen to the b 2? So in this case the b 2 and beta 2; so beta 2 is of course b 2 is capturing beta 2, but at the same time it is capturing this component also, but because this is having a negative impact, so b 2 is actually going to be less than beta 2, here the result is going to be b 2 is less than beta 2 and why is that because of inclusion of this negative component.

So, essentially here b 2 is again beta 2 into h beta 3, but because one is negative and one is positive, this component is going to be negative. So, then beta 2 two is an underestimation, sorry b 2 is an underestimation of beta 2, this is an underestimation, correct? This is something we need to understand. This is an underestimation of beta 2. So, it is going to be I will write down multiple colors, b 2 is an underestimation underestimation of beta 2.

Now let us try to get another case, the third case let us say. And in this third case, we will have let us say, this one is positive and this one is negative. So by now we have an intuitive understanding what sort of relationship it should have and it is going to be again b 2 is an underestimation of beta 2 and what would be good example of that.

(Refer Slide Time: 12:02)



So let us say we talk about take the example of the same wage equation. But in this case, let us say wage is dependent on this. So the actual regression equation is his beta 1 + beta 2 into education and beta 3 into experience, alright. Now how let us say this is the true model and my estimated model wage hat = b 1 + b 2 into education. So I have basically omitted experience whereas experience is related to education.

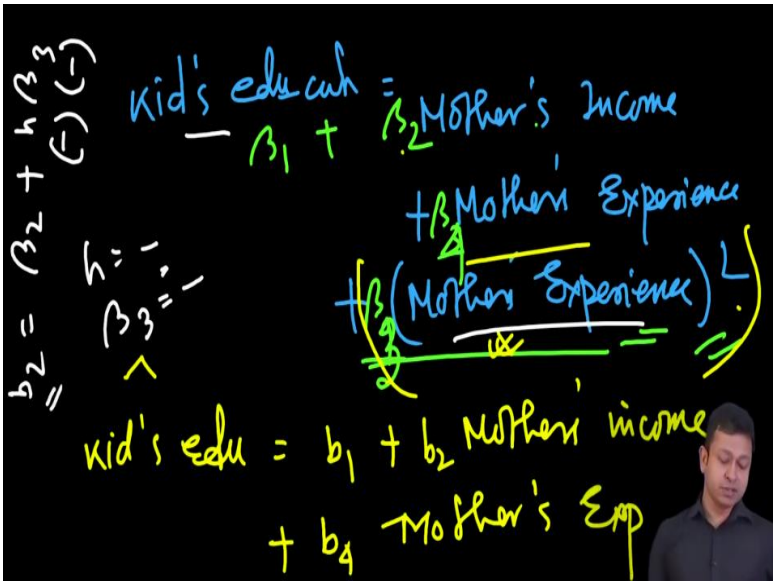
So more experience means more wage. So like more education means more wage, more experience means more wage. So that means beta 3 is positive. Now how about the relationship between education and experience? So, education; remember that so the more education we have, so let us say I have a lifespan of 60 years and let us say I am in school for whatever reason so let us say 35 years of edu, then my experience is going to be 25 years.

So education and experience they are actually negatively correlated. So the more education I have, the less experience I will have. I will just show you in an example after this lecture in the next lecture, how we can actually interpret, how we can actually see that in reality it is negatively related. So then I can say in this case h is negative. So, exactly the third case that we talked about, so h is negative and beta 3 is positive.

So this is an example where education and experience so they are negatively related whereas both education is positively related to wage or wage is positively related to the education and wage is positively related with experience. So, then basically if we have that, so how would b 2 and beta 2 will be related. So, b 2 two here again is basically beta 2 + h beta 3. Now, h is negative, beta 3 is positive.

So essentially the whole term 2 b 2 is going to be an underestimation of beta 2 because beta 2 is higher but when you get a b 2, so that actually since it is capturing this negative component also it is going to be an underestimation. So, this is exactly what we said. Now, coming to the last part and we know what is the last part by now and that is basically both negative, and I need to find an example for that. What would be an example for both b 3 and h negative?

(Refer Slide Time: 14:47)



The third example we will write kid's education and again we have the working class mother. So kid's education we will have let us say mother's income. We had mother's income, then we will have mother's experience and then we will have mother's experience squared. So, I will write le tus say this is equal to an equation form, I will just add the equation is beta 1 + beta 2 into mothers income, beta 3 into mother's experience and beta 4 into mother's experience square.

So I looked into this component and why I have included this component actually this is based on Mincerian wage regression. So we know that income and experience, they are positively related. But income and experience squared they are not positively related because we know that with age, basically as we age or experience increases or income and experience relationship gets plateaued.

And that is because at a higher age, we actually do not get the increase in our income proportionately and that component is captured in the experience squared. So, essentially this experience squared and income they have a negative relationship. So essentially that would

mean, so I would rather say this is just to maintain the convention, let us say this is beta 3 and this is some beta 4.

So if in my regression equation, so let us say I have in my estimated regression equations, kid's edu let us ay hat I have b 1 + b 2 mother's income and let us say I have omitted this one and I have only have this b 4, I mean it really does not matter if I want to include it or not because I want to see the relationship between b 3 and b 2, so let us say I have it, mother's experience. So, I have actually missed out on this term, I have actually missed this relevant term here.

Now, when I miss this relevant term, I know the relationship between basically b 2 and b3, **three** so essentially these two are negatively correlated. So h is going to be negative because mother's experienced square and mother's income they are negatively correlated. So h is negative. And what about kid's education and mother's experience square? Well, since mother experience is actually; the more the mom is experienced that means the longer time the mom is working.

And if the mom is working for a very long time, very many years, so she is basically not giving the time for her kid's education and that is where what is happening is she is actually, the relationship between kid's education and mother's experience squared is actually going to be negative. So my beta 3 is actually negative. So then the beta 2 b 2 + h B 3, and here h is minus and b is minus, so double negative will make my case a positive.

So what will happen, my b 2 is going to be an overestimation of beta, just like the first case. So it is going to be b 2 is an over estimation of beta 2. So these are the four cases that you have seen the relationship between h, beta 3 and beta 3. So the next lecture, whatever you have seen now, we are going to do some hands on to actually explain that and we will also talk about the relationship between this beta terms and h terms and and the R square. So with this, we will end this lecture here. Thank you.