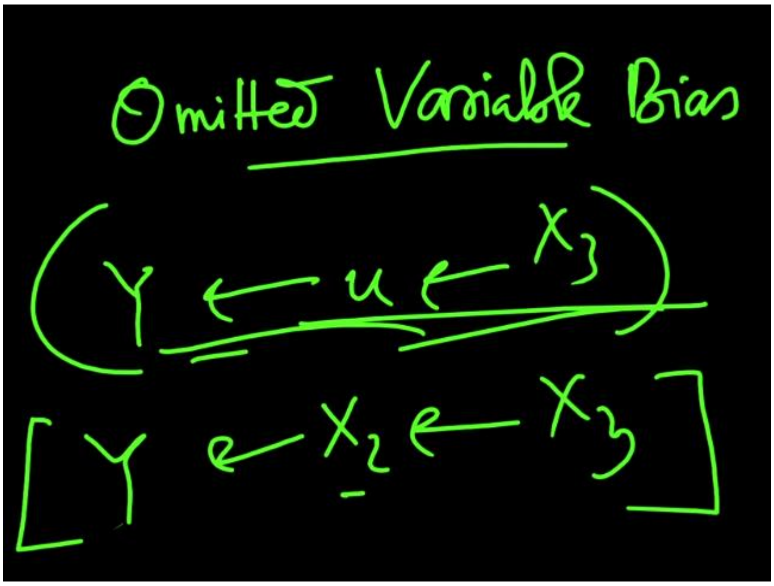
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology - Kharagpur

Lecture – 84
Model Specification - Continued

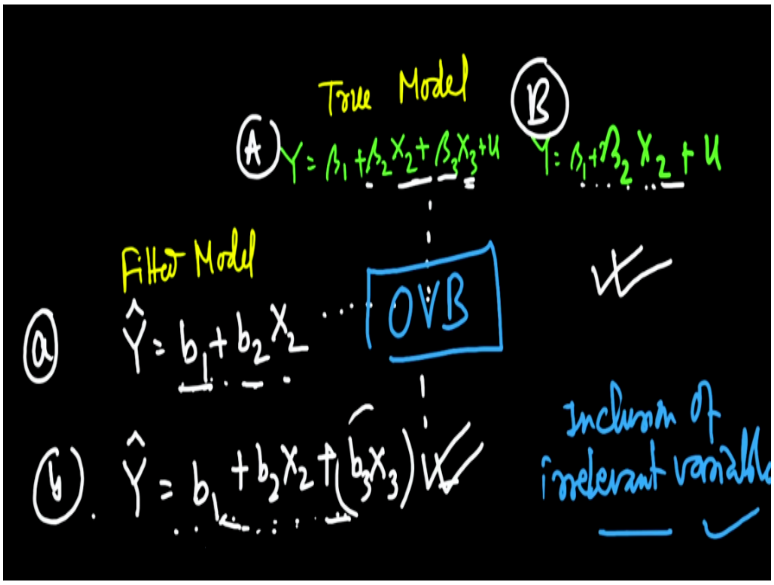
Hello and welcome back to the lecture on applied econometrics. We have been talking about model specification, and specifically we are talking about omitted variable bias. So there are at least two types of model specifications we talked about to get the right variable. And second is to measure the value of the variable. Now we have been talking about the first problem to get the right variable and the first part of the first problem is the problem of omitted variable bias, so where do we actually exclude a relevant variable, now what happens there?

(Refer Slide Time: 00:59)



So when we are talking about omitted variable bias, just to recap we said that it actually influences my Y variable of interest, the dependent variable through two routes; one is through error term and second is through the other variable present in the model. And we already talked about this one, this essentially the problem of heteroscedasticity and autocorrelation, which I have explained in detail previously. Now, we will talk about this particular problem, this particular problem when X 3 is actually influencing Y through the route of X 2.

(Refer Slide Time: 01:44)



Now before I do that, let me actually explain the problem of omitted variable bias a little bit in detail so that you understand the relationship with the true model and the fitted model in case of omitted variable bias. And this is something I have taken from Christopher Dougherty, true model and you have fitted model. First let me write down the true model. So let us say the true model is Y is equal to let us say beta 1 + beta 2 X 2 + beta 3 X 3 + some u.

And in the second case U = beta 1 +; let me actually reduce the size a little bit, beta 2 X 2 + error term. Let us say I have fitted a model and that model let me use a different color for fitted model. Let us say this is Y = beta 1 + so I have these or I would rather write so in case I am fitting a model, so I write b instead of bets, we are estimating that b 1 + b 2 X 2. I do not have an error term here and Y let us say this is the one case I got this model. In the other case I got this model b 2 X 2 and b 3 X 3.

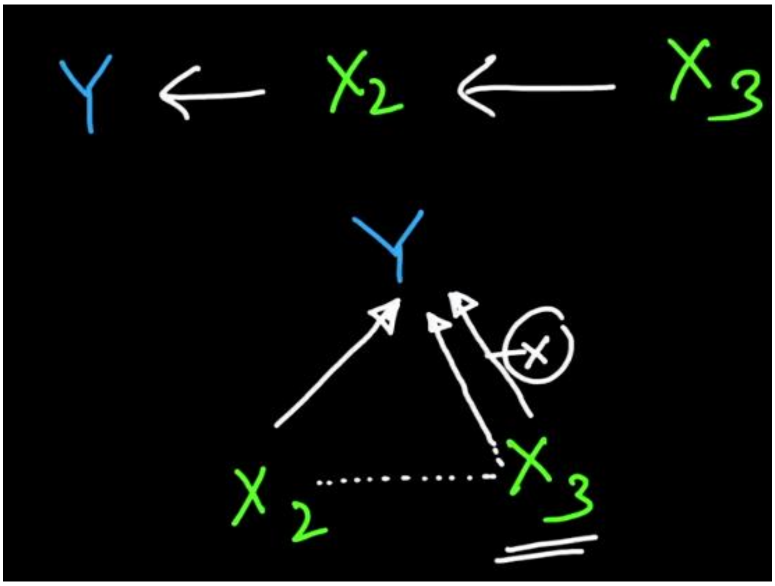
So, if these are the possibilities, so if this is the true model and let us say 2 students have got two different fitted models. So, this is fitted model 1, fitted model 2. So, if this is what we get, so then in this case, what I am getting I actually have omitted an important variable which is X 3. I did not include X 3 in my model. So, this is a problem of omitted variable bias, let me write with a different color, this is a problem of omitted variable bias.

Whereas, this one if the student has actually fitted model b, this b 1 and b 2 X 2 and b 3 X 3 whereas the true model is this, so, then this is a true model. So, I will write this is a correct fit, we have got the right model. On the other hand, if let us say the true model is this, actually Y is only explained by X 2 and there is no X 3 component present and the student has actually got the model, he actually got b 1 + b 2 X 2, so exactly he has only one explanatory variable which is X 2.

So in this case, the model is true model, here the model fitting is correct. On the other hand, if this is the true model and a student has actually ended up fitting this model, so he has actually included this component, which was irrelevant, so this is a case of where inclusion of an irrelevant variable. So, these are the two problems that we see when we are talking about model specification and when particularly talking about inclusion or exclusion of a particular variable of interest.

So, we are talking only this one right now, so this is something we will talk about later. And within omitted variable bias, we are talking about error route, u route and X route.

(Refer Slide Time: 05:22)

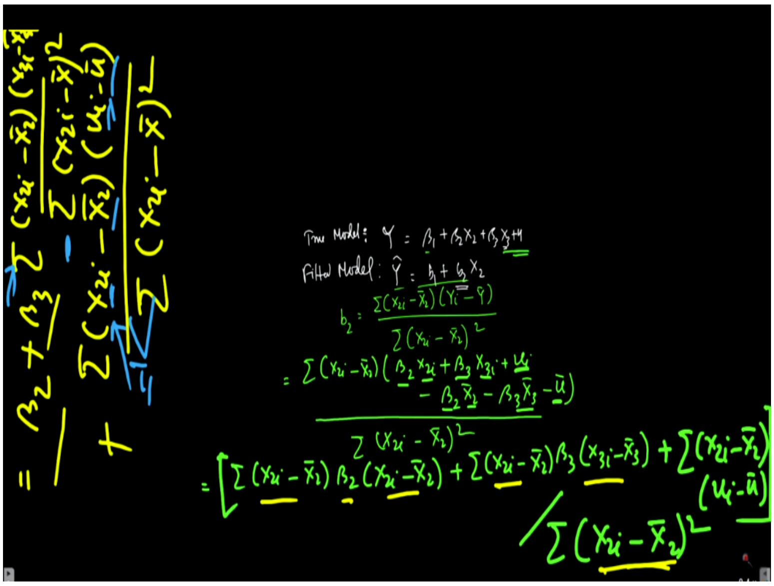


So, let us now; we have explained the u route, let us now talk about the X route. When Y is getting influenced by X 2 whereas X 3 is omitted, so X 3 is actually influencing Y through X 2, so that is a problem that we will talk about. Now, let me actually draw another diagram to illustrate it even in a better way. Let us say this is my Y alright, this is variable of interest and this is my X 2 and this is my X 3.

So, how exactly can draw it? So X 2 is influencing Y directly and then there is a direct influence of X 3 on Y too, but because I am not having X 3 in the model, so what is happening is that this X 2 is actually mimicking the effect of X 3 and that is being present in the Y model. So, whereas this particular component is actually absent. So, this is how we can actually explain the influence of, the basically try to understand the problem of omitted variable here.

So, let me actually explain mathematically. This is a graphical illustration, we have got an intuitive understanding and this is a graphical understanding and let us now talk about the mathematical illustration of it.

(Refer Slide Time: 07:01)



So, let us say in this true model Y = beta 1 + beta 2 X 2 + beta 3 X 3 and u and fitted model is Y hat = b 1 + b 2 X 2. Residually, I have omitted this variable, alright. So, let us now try to understand what is happening to my regression coefficient. So, what beta 2 is actually representing here? How much it is able to explain and what component? So, let us try to get that.

So we know by definition b 2 is equal to we know the definition of the coefficient that is basically X 2i X 2 bar into Y i – Y bar and then in denominator I have X 2i –X 2 bar whole square that is how actually estimate coefficient in a simple OLS, this is a simple OLS. Now, let us substitute the value of Y i and y bar from the real model. So then it will be is equal to X 2i – X 2 bar into Y i is going to, I am going to get the value from this model.

So essentially this component is going to go vanished beta 2 X 2i beta three 3 X 3i + u i – beta 2 X 2 bar – beta 3 X 3 bar – u bar. So, that is what you have in your numerator and denominator is going to be the same. That is X 2i – X 2 bar whole square. So, it has become pretty lengthy, I mean a little daunting also, but nothing to be is actually very simple here. So, let us actually get the terms separately.

I wish I could do it in the same page, let us see. So, if we can let us say I take X 2i – X 2 bar and then here I have let us say there is a beta X 2i and there is a beta, sorry beta 2 X 2i and beta 2 X 2 bar. So I can take this beta 2 here and I can have an X 2i – x 2 bar and then I will have another term that is beta 3. So here for X 3i and X 3 bar so if I take these, so that will be X 2i – x 2 bar into beta 3 three and then I have X 3i – X 3 bar.

Iti s really getting a little cluttered, but let us see. And then I have this u i and u bar. So all I have is X 2i – X 2 bar into u i – u bar. Therefore, all the terms I actually divide by this denominator, so we will see what we get from here. So, let me actually use this space on this page. So let me actually write it. So here I have you see this term and this term actual results into this term and all you are left with is beta 2.

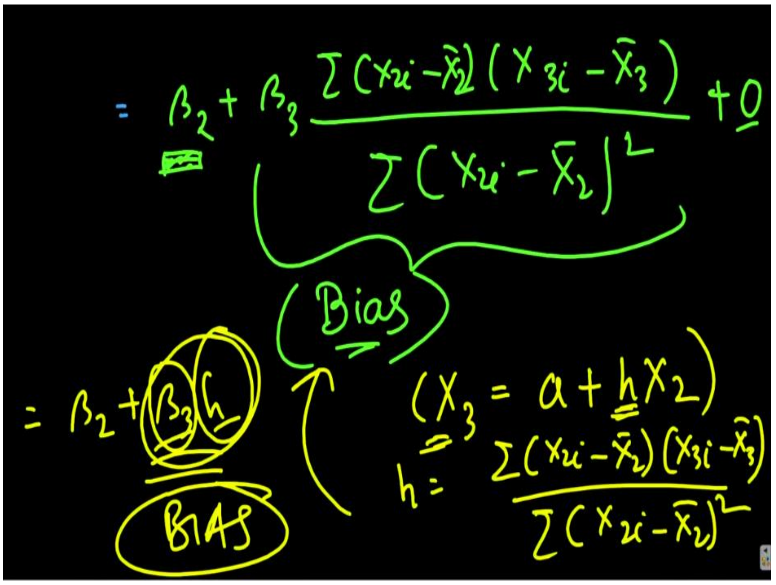
Then in this term, you again get this term and this term, so X 2i X 2 bar into X 3i X 3 bar with a beta 3, but the denominator is different. So, what you get is this beta 3 into summation X 2i X bar into X 3i – X 3 bar by summation of X 2i – X 2 bar. And the third term is X 2i – X 2 two bar into u i – u 2 bar and then you have the same denominator here X 2i – X bar, really cluttered, but let us try to make sense from here.

I did not want to go to the next page because then I had to copy all these things from this page. Actually see so I have a beta 2 here and I have a beta 3 into some component and if you look at it carefully, so it is basically the correlation coefficient between X 2 and X 3. So, you will see that I am basically measuring the covariance in the numerator, this is the covariance between u 2 and u whereas this is the covariance that I am getting between X 2 and the error term.

Now, this is you remember our assumption is known as stochastic regressor and because it is known as stochastic regressor this term, if I take an expectation of this this is going to be 0 because these are the deterministic components, these are already fixed. So, if the coefficients are already fixed, they cannot have any correlation with the error terms, but this term here this is something else we need to consider that carefully.

It is a correlation coefficient between X 2 and X 3. Now, when we have a correlation coefficient between X 2 and X 3; let me actually take a new page.

(Refer Slide Time: 13:16)



And that is let me write down what we got previously. So, beta 2 + beta 3 X 2i – X 2 bar into X 3i – X 3 bar and in denominator I had X 2i – X 2 bar square and the third term was 0. Now, this was supposed to be the true, we are getting beta we are estimating b 2, so b 2 should be I mean b 2 should result in beta 2 if I take expectation. So, from this equation we should get the value of b 2 ideally should represent beta 2, but then we are getting additional component of beta 3 into this term.

And this is effectively representing the bias. So, this part is a bias in the model when I get this whole beta 3 into this term. So, what is the quantity of this bias? So, let us say if I have X 3 is equal to let me write down a + let us say h X 2, so then this correlation coefficient between the value of h is going to be this because this is the correlation coefficient between X 3 and X 2 and it is going to be X 2i X 2 bar intop X 3i X 3 bar by square of X 2i X 2 bar whole square.

So, essentially this is a correlation coefficient or this is basically the the coefficient of X 2 when I do the regression between X 2 and X 3. So, this is h and then if I impute the value of h into the previous equation it will be beta 2 + beta 3 into h, so this beta 3 into h is going to be the bias term. So, we need to understand these bias terms, so the extent of the bias in a regression equation from this value.

So, essentially you see there are two components here, one is the beta 3. So, beta 3 is nothing but that basically represents the actual strength of relationship between X 3 and our Y term. The beta 3 is actually representing that, whereas h is representing the strength of relationship between X 2 and X 3. So, we will see in the next lecture how these relationships, the strength of relationship and the direction of relationship between X 2 and X 3 and Y actually will help us to understand the extent of bias in the regression equation. Thank you.