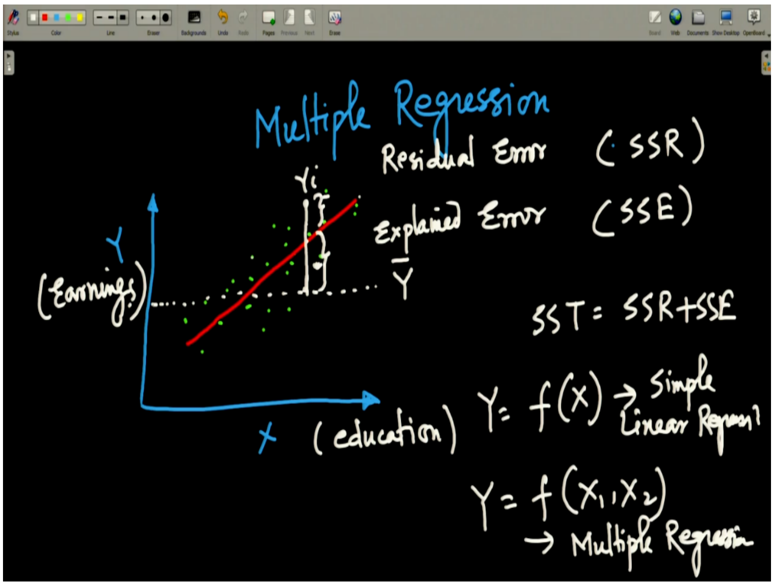
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology - Kharagpur

Module - 6
Lecture - 48
Multiple Regression

Hello and welcome back to the course on Applied Econometrics. In this lecture, we are going to touch upon the multiple regression. Now, in the previous lecture, we have seen how simple regression works, and we have plotted Y and X, and we have seen the different components of error. Now, let me recap; let me actually draw it.

(Refer Slide Time: 00:43)



So, we had this coordinate system, where I had explanatory variable X and my dependent variable Y. And I got lot of different points, which are actual observations for X and Y. And now, what I have to do here is, I have to actually draw a regression line. And we have seen that we can actually draw a regression line by minimising the sum of square of error terms. Now, so, that we know.

And we explained that, if I have a, say a line here, say this is the mean value of all the Y's, and I get it Y bar, so, then, I need to divide the error terms into different components. And we have seen that we can actually divide this error term. So, one error term is the line. So, this part of, let us say we are talking about this point. And then, the total error from this point, total distance of this, let us say Y i from Y bar is this, this length.

Now, if I divide this total distance into different components, I have this component and I have this component. Now, the first component here, it is explained by my regression line. And I can call it, this is the explained error. We have defined it as explained error. And the other part of the error is not explained by my regression line, and we call the residual error; this part is called residual error.

Now, when I, for all the different points, in the same manner if I get the error terms and if I actually divide the error terms into 2 components, one is the residual error and another is explained error. So, we can actually get the sum of the different error terms. And we have seen that we actually square all the different error terms. So, essentially, what we get is sum of square for this different components, residual and explained. So, how we express them?

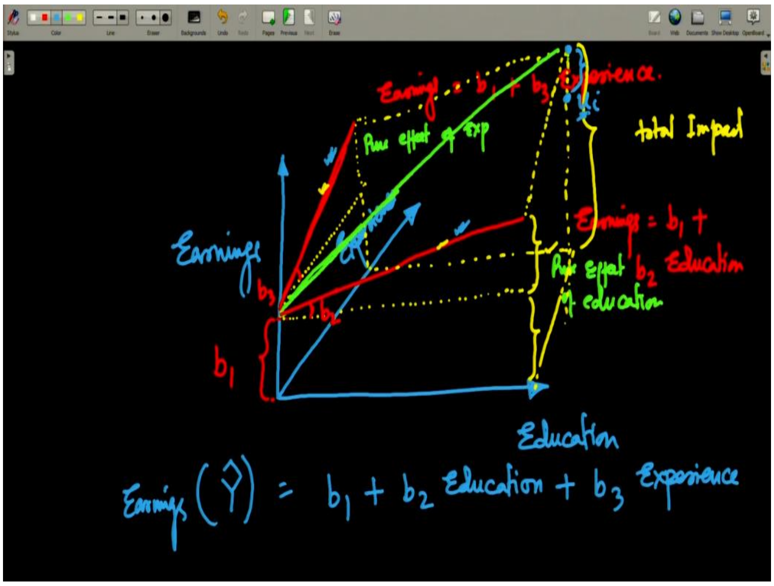
We express them SSR, sum of square for residual error; and we term explained error as SSE, sum of square explained. And the total error SST = SSR + SSE. So, we have seen it already. Now, this is for the simple linear regression. And if I want to have more than 1 variable; so, here I have only 1 variable. Let us say this X variable here is my education. And my Y variable here is, let us say my income or earnings.

So, I kind of presume that my; or actually, what I see from the data is that my earnings is actually increasing with the level of education. And that is what we see from the data. But suppose if we want to have more than 1 explanatory variable, let us say I want to have say experience; because I know from my knowledge that people who have more work experience, they earn more.

So, it is not just the education, but also the experience which matters when it comes to the earning. So, now, how do I actually capture the impact of other variables. So, only when I, consider more than 1 explanatory variable, then it becomes a case of multiple regression. So, here, it is a simple regression, this one. So, here we write, say f X, Y is equal to, so, simply f X; and where my X is education.

But when I have more than 1 explanatory variable, so, Y is equal to, let us say f of X 1, X 2. It is a case of multiple regression. So, here I write, simple linear regression, where we are only talking about linear regression. And here we call it multiple regression; this is also linear regression; multiple regression. Now, we will try to explain graphically how a multiple regression is different from the simple linear regression.

(Refer Slide Time: 05:04)

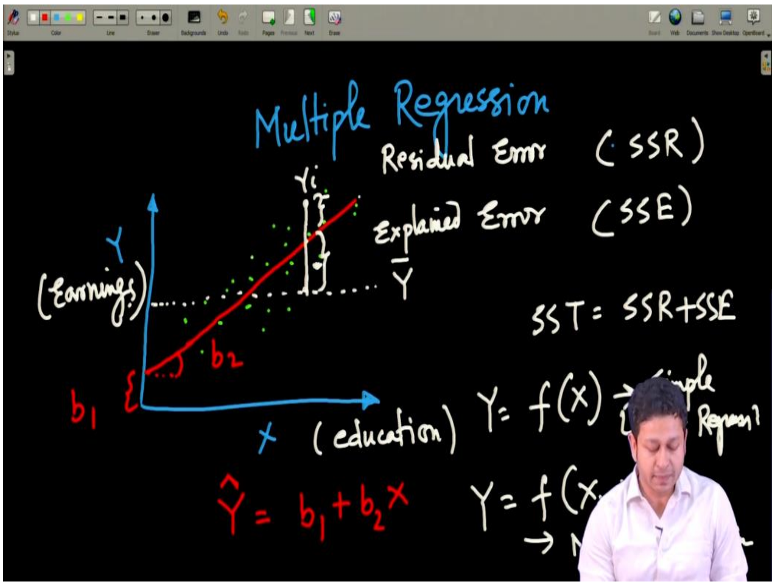


So, let us go back to the same example, education and earnings, but here I am going to add another line and that is for my experience. So, let us say I have drawn another line that is experience. So, this is my education, this is my earnings and this is my, say experience. So, here, I am just considering 2 explanatory variables; not more than 2 because we know that we have limitation in terms of how many dimensions we can plot.

So, here is a 3-dimensional system, where we can plot at max 2 explanatory variables. If you ask me whether we can plot like 3 explanatory variables, we cannot; or 4 explanatory variables, we cannot. So, our limitations is this, that we have to plot in a 3-dimensional plane. So, that is the maximum we can do. But the concept goes like this; you can actually think that there are multiple planes and you can actually keep on adding the different dimensions.

So, here we have 2 dimensions, in one dimension, we plot education; in another dimension, we plot experience. And the outcome variable or the dependent variable is earnings; so, as we had in the previous case. Now, how do we really get the regression line? That is what we have to see here.

(Refer Slide Time: 06:27)



So, just like in the previous case; we have seen previous case, we had this line, let us say we extend this line up to here. And if we extend this line up to here, we have this intercept term. Let us say we term it as a b 1. And then you have a slope, and this slope is, let us say b 2. So, then, I will write my regression equation, the out, the regression line. So, I can write Y hat is equal to, say b 1 is the intercept term plus b 2 into X.

So, all these different X variables if I plot, I will get this estimated predicted Y. So, predicted Y is always represented as Y hat. So, this is basically the line that we have created. So, that is what is this equation representing. Now, here also we will have, say we want to understand the impact of education. So, let us say there is an intercept term here, let us say it is some b 1. It is not necessary that the two b 1's are the same; we are just maintaining the same notation, so that we do not have any confusion later on.

So, this is the intercept term, like we had in the previous case. And then, you actually try to estimate the impact of say education, let me use a different colour here. So, you want to see the educational impact of experience. So, impact of education would be measured by the slope coefficient. And let us say the slope coefficient for the impact of education is going to be b 2.

Now, if that is the case, and if I draw my regression line here; so, just the same way I am; or let me use the same colour, the red colour that I used in the previous case. So, if I draw the regression line here, I am just considering the impact of education. And let us say this is the angle here. And let us say this is b 2. So, this is b 2. And then, this line is basically representing the impact due to education only. So, the regression line here would be, for this line, is going to be earnings.

I can actually write it as earnings is equal to b 1 + b 2 into education. So, this is pure effect of education. We can say that this is the pure effect of education. This is the intercept term. This is the intercept term; this is the pure effect of education. This part is a pure effect of education. Now, we will also be interested to actually understand the pure effect of experience, because all these different components do matter when it comes to earnings; because we have already specified in our model that it is both education and experience that matters, so far as my earnings is concerned.

So, if I try to also estimate the impact of experience on earnings, so, I will again get the pure effect due to experience. So, in this, when I talk about pure effect, it is not mixed up with the effect of education or the intercept terms. Intercept term is already there. Then I measure the impact of education on earning. And then, I also measure the impact of experience on earning. Let us say the line goes like this.

So, this is the impact of experience. And the angle, let us say is b 3; this angle is b 3. Now, if I try to understand what just has happened is that; so, let me actually complete the diagram. So, we can basically draw the projection here. This is the projection. And we can actually get a resultant of both these 2 lines. So this line is actually representing the pure effect of earnings. So, here, if I just consider this particular regression equation with b 1 and b 3, what I will have is, earnings is equal to b 1 + b 3 into experience.

And this part is pure effect of experience. And this part is pure effect of education. Now, I want to, at the end of the day, since I am considering a multiple regression, I have different explanatory variable, I need to get the final regression line that I want to get from all these different impacts. So, what I will do? I will basically draw a resultant line. So, I have this line due to my pure effect of education.

I have a line due to pure effect of experience. And then, I will get a resultant line. How do you get that? It is simple. We will get the resultant line. And let us say I again draw some projections here. And lets us say this line goes like this. Actually this line should be parallel to the previous one, right? This should be parallel to this one. So, it will be something like this. And I will have another line which is parallel to this line, so, which will be something like this.

And then, if I draw a line from the origin or from this point where my intercept ends, so, what I get is this line. Or essentially, this kind of represents a plane, because I am considering more than 1 dimension here. So, here I will have the final regression equation, where I will have, I will actually consider all the impact of education and experience on earning. So, this is what is the impact that is going to be. This is the impact.

So, if I draw it, so, this is the total impact due to education and experience; so, total impact. Now, that is basically my Y value. That is the predicted Y, right? Now, there is another component that is the error term. So, we need to consider the error term also. So, the error can go; so, this is the predicted Y value. Now, the error can go actually above that. Error could be positive, error could be negative.

So, anywhere you can consider the error. Suppose my error is negative, u i for the particular Y i; so, then, in that case, you will basically subtract Y by your, depending on which side the Y is, we will add or subtract the error term. So, now, we have graphically understood how it will be like when we draw a multiple regression. And when we actually draw it, do a visual depiction of the multiple regression; so, we have seen how it would look like.

But at the end of the day, what we will want is, we want an equation like this, right? So, here, for multiple regression, my equation is going to be, again my earnings or Y hat is essentially nothing but b 1 plus I will have my b 2 into education plus b 3 into my experience. So, this is my predicted regression equation which I am basically trying to derive from the different explanatory variables. So far, that goes, that is the basic illustration of multiple regression.

And going forward, we are actually going to use some data to show how multiple regression works. And we are actually going to show you how these different lines that we have just drawn, these actually matters in terms of explaining the data. So, we will see with that example. And also we will talk about a regression table, like when we do a multiple regression, there are different elements in a regression table. So, we need to understand each of these different elements in a regression table. They are very important. And we will do that in the next lectures. Thank you.