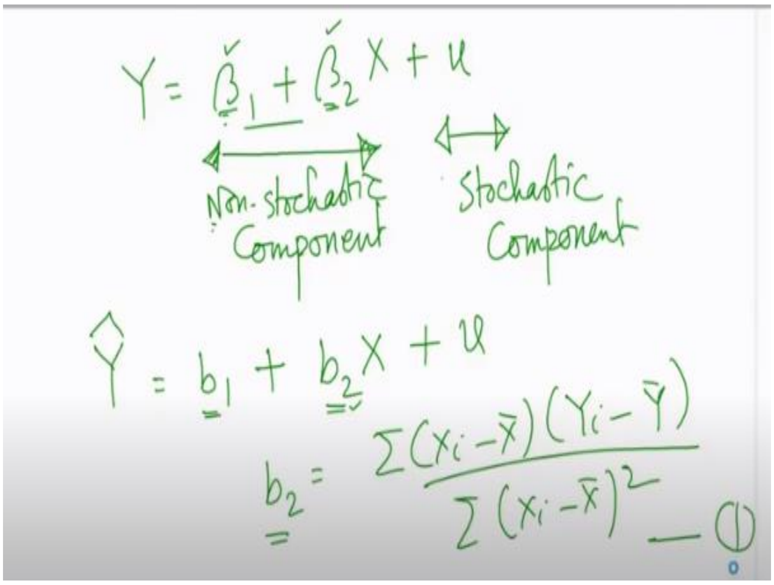
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology-Kharagpur

Lecture - 43
Error Term, Coefficient of Determination, Regression Coefficient (Contd.)

Hello and welcome back to the lecture on Applied Econometrics. We are in module II and we are talking about econometric modeling. And we are talking about different concepts in this lecture. And in this particular lecture, we are going to talk a little about regression coefficient. Now when we talk about regression coefficient, let me first write down the very simple regression equation.

(Refer Slide Time: 00:51)



Where we write say Y is equal to beta 1 plus beta 2 x and say some error component, right? Now we have explained that a regression equation you can always write a component which is stochastic and which is not stochastic, okay. So this part the error term it is the stochastic component and this term with the variable and coefficient, this is the non-stochastic component.

Now what do we mean by stochastic and non-stochastic component? So in stochastic component you always have some randomness. You do not know what the value of u is. Whereas, the non-stochastic component is basically the deterministic component. You have some idea, you actually know the value of this coefficients, okay. So that is why it is non-stochastic component.

So when you draw a regression line, they are basically the slopes of that line, slopes and intercepts. So you know the value. So the moment you know the value it becomes non-stochastic component, right? Now what you have to see is that this beta 1, beta 2 these are all true population parameter. So for a given population they are constant. And there is no doubt about you know it being a stochastic component.

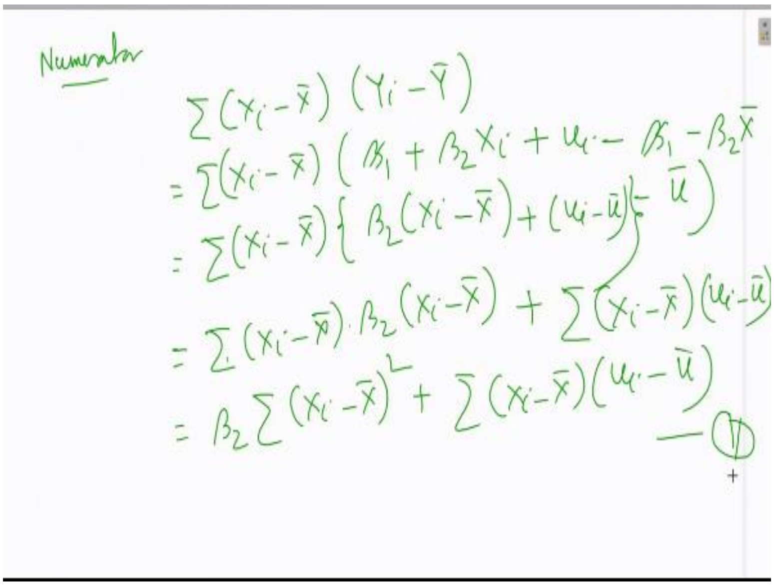
But when it come to data that you, you know have in a sample and you try to estimate beta 1, beta 2, do we still have non-stochastic component there and how that actually looks like. So that is something we are going to see in this lecture. We will see when we actually do a regression table we normally use, for example, t test, f test, and we create confidence interval.

Now all these things if you want to do, you really cannot do if you have a fixed constant value right, it has to have a distribution. Now where from the distribution is coming? So that is what we are going to see, when it you know how this population parameter beta 1, beta 2 is estimated from sample parameter b 1, I will to differentiate I will say b 1 and b 2, okay.

And I will write Y hat instead of Y, because I am estimating using my sample observations. So now I have to get this b 1 and b 2. And I have to see how they are related with beta 1 and beta 2. So we will start with b 2. In fact we will do, we will focus on the b 2 part. So the coefficient for x. Now how do we, you know by definition b 2 is, by definition, b 2 is summation of x i minus x bar into y i minus y bar by x i minus x bar whole square, okay.

Now what we will do is we will simply replace the Y values in terms of the x values okay. So what I will do? So it means summation of x i or actually what we can do is we can only take the numerator like we have been doing.

(Refer Slide Time: 04:19)

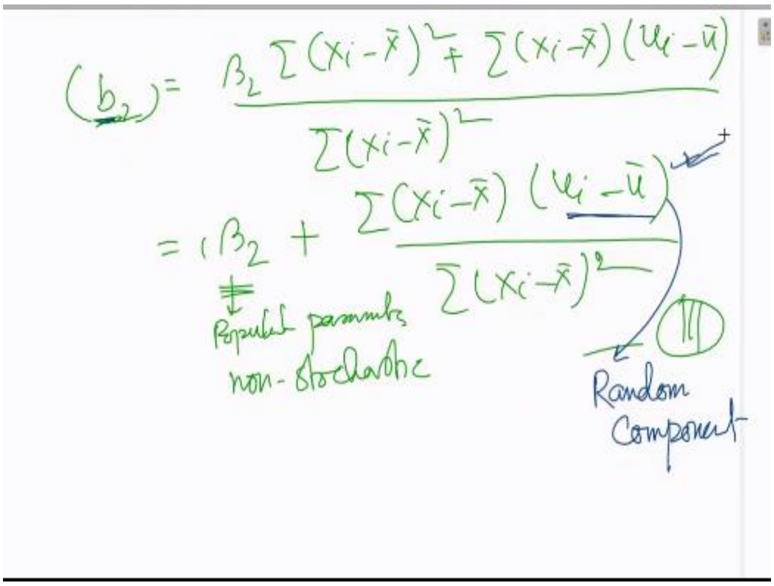


So summation of, so the numerator is summation of x i - x bar into summation of y i - y bar. Now my Y equation is nothing but this. So I will replace my Y with this. So essentially this means summation of x i - x bar. Y i is beta 1 + beta 2 x i + u i - beta 1 - beta 2 x bar - u bar. So essentially I have I did not use another comma another bracket there.

So if I further calculate it, I will have this beta 1 beta 1 cancelled and what I will have is a, let me use a second bracket, beta 2 x i - x bar and u i - u bar. So this is what I will get. Let me close it also. If I further expand it what I will get is x i or x i into x bar into beta 2 into x i minus x bar plus I will have summation of x i minus x bar into u i minus u bar, okay.

Now I take beta out. So what I will have, beta 2 into summation of x i minus x bar whole square and here I will have x i minus x bar into u i minus u bar, okay. Now recall the first equation we had. The beta 2, the expression for beta 2 is this. Now if I replace, I should have given some number 1, let us say this is 2.

(Refer Slide Time: 06:25)



If I use the one, equation 1 and equation 2 together what I will get is B 2. So we have to remember, we must not confuse B 2 and beta 2. So b 2 is equal to B 2 is equal to beta 2. Then I had summation x i minus x bar square and then I had summation of x i minus x bar u i minus u bar, okay. And then I have to divide the whole thing with summation of x i minus x bar whole square.

Which means, what I am left with beta 2 and then I have summation of x i minus x bar u i minus u bar by summation of x i minus x bar whole square. Now this is something very interesting. And this is something we need to keep in mind when we are going to do all these regressions and hypothesis testing of the regression coefficients and so on and so forth. So note it carefully you have your beta.

Sorry you have your b2. This is something you need to be very mindful about. And here you have your beta 2. So beta 2 is the population parameter, okay. This is the population parameter. And you estimate this population parameter using, you never know this population parameter or you can do is estimate this population parameter as b 2. And b 2 actually consist this population parameter which is non-stochastic.

And whereas, this component you have this error term here. The interesting part here is that you have this error term, which is basically the random component, which is the random component. And because you have this random component, this is a stochastic component. And the moment you have this random component b 2 will not have a specific value, specific slope value.

But rather it will assume a distribution according to the distribution that the error term assumes. So that is why it is very important to understand that the regression coefficient is actually it is a distribution and it is actually the distribution, the properties of regression coefficient distribution is determined by the properties of the distribution of the error term, okay.

So this is something that is very important for us to sort of keep in mind when we actually understand the error term. And that is why whenever we are you know considering any hypothesis testing or any other thing, we always get a distribution for the coefficient, b 2.