Econometric Modelling
Professor. Sujata Kar
Department of Management Studies
Indian Institute of Technology, Roorkee
Lecture – 13
Multiple Regression

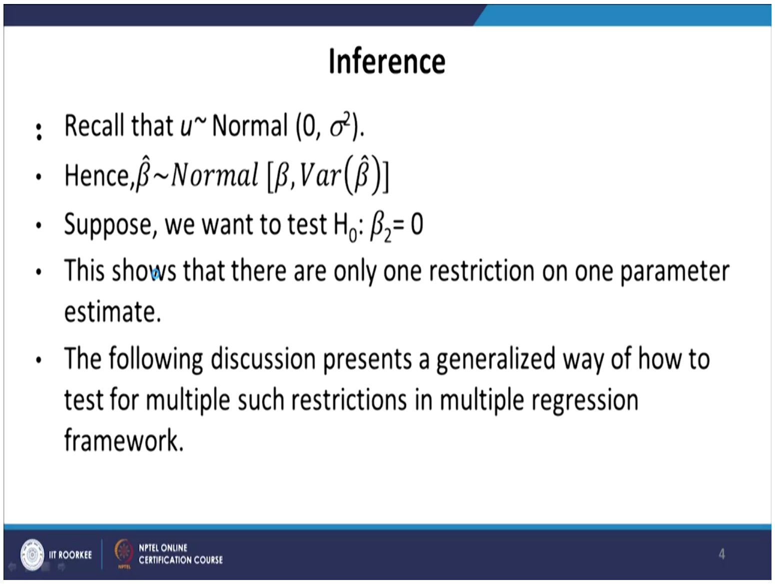
Hello and welcome back to the course on econometric modelling. This is module thirteen and I am going to continue some more discussion on multiple regression.

(Refer Slide Time: 00:40)

So, the way we had introduced inference in the context of simple regression analysis, that is a two variable case analysis, here also I will first discuss inferences, some generic discussions or general discussions in the context of multiple regression analysis. And after that, I will also discuss one goodness of fit measures.

So, far we have discussed r squared, which is applicable in the context of simple regression as well as multiple regression. But we can offer an improvement over this existing measure of r square as a goodness of fit measure. And that is adjusted R squared, which will be discussed at length in this module. So, I continue with some more discussion on multiple regression.

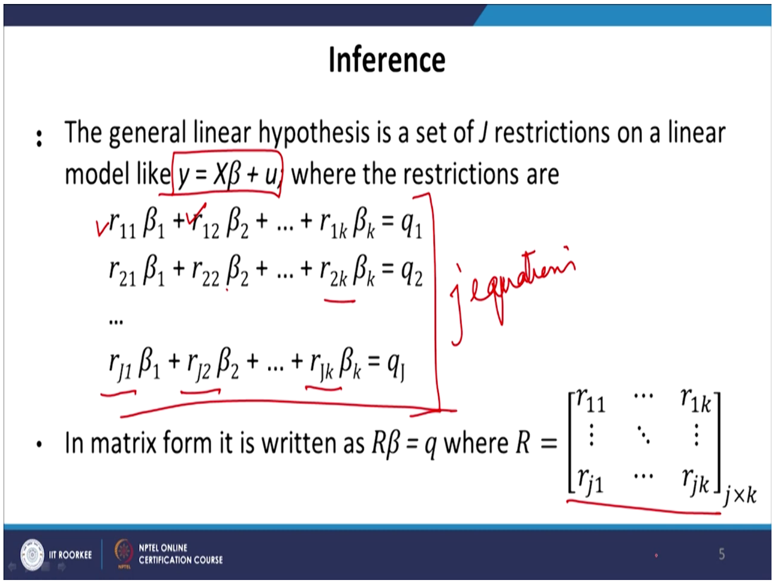
(Refer Slide Time: 01:34)



So, first talking about inferences, we in the previous module, had actually ended with the discussion on the assumption of one of the CLRM a that u is normally distributed with 0 mean and constant variance equals to sigma square. And therefore, the beta hat also follows the normal distribution with a mean beta and variance beta hat. Now, suppose we want to test the null hypothesis, beta 2 equals 0 *(refer slide time: 01:34).* So, it can be beta 2, beta 3, or any beta j in a multiple regression framework.

This shows that there is only one restriction on one parameter estimate. The following discussion presents a generalized way of how to test for multiple such restrictions in a multiple regression framework. So, we will not be specific to any kind of, testing procedure, which will be taken up in a later part, possibly the next part. But for the time being, we are actually presenting a generalized framework for hypothesis testing.

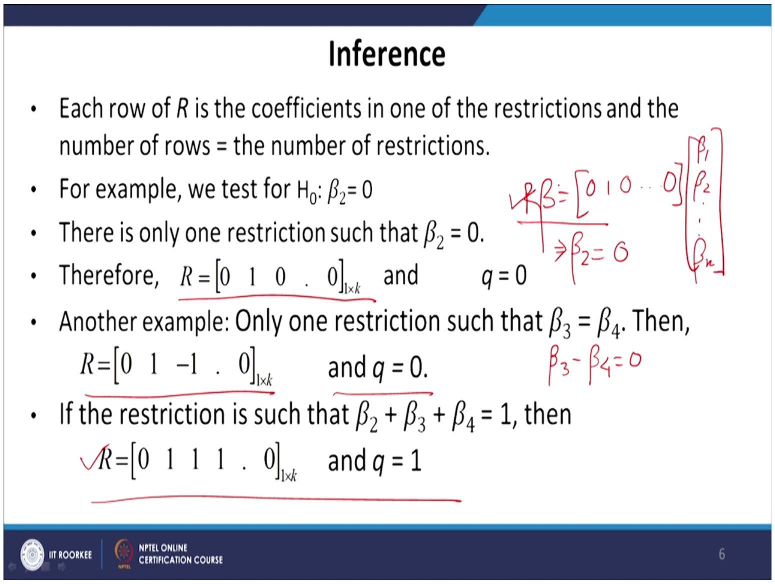
(Refer Slide Time: 02:50)

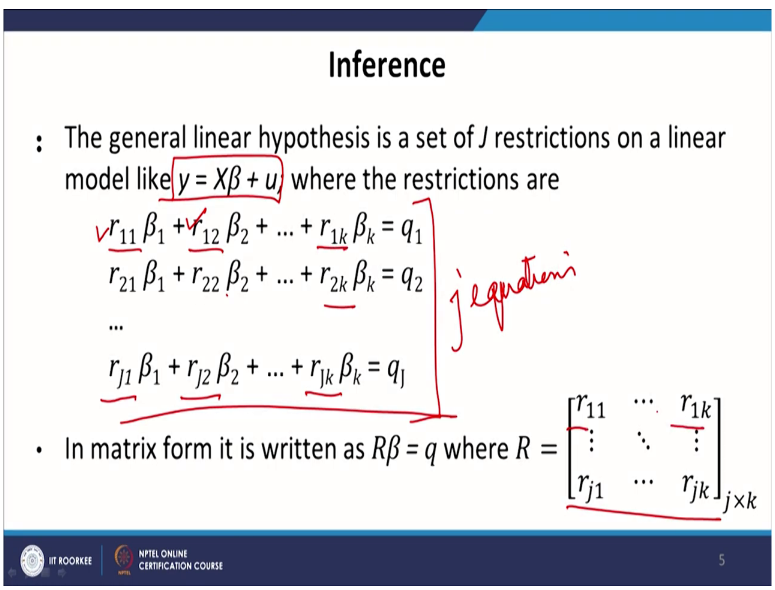


So, the general linear hypothesis is a set of j restrictions on a linear model, like y equals X beta plus u. So, if you remember, this is our population model y equals X beta plus u where y is an n by 1 vector x is n by k, or n by k plus 1 matrix, beta is k plus 1 by 1 vector, and u is n by 1 vector *(refer slide time: 02:50)*. Now, I am defining certain restrictions. Now, these restrictions actually relate to the hypothesis that we need to test or we probably want to test.

This is how *(refer slide time: 02:50)* we write the set of restrictions, suppose there are j restrictions. So, in that way we have j equations here because there are j restrictions. In the matrix form, it can be written as r beta equals q, where r consists of these coefficients *(refer slide time: 02:50)*. So, r_{11}, r_{12}, r_{1k}, and similarly r_{21}, r_{22}, r_{2k}. And finally, I have r_{j1}, r_{j2}, r_{jk}. So, this is how we present it in terms of a matrix form. And beta is the original beta vector having beta 1 beta 2 to beta k. And that is how r is a j by k vector, beta is a k by 1 vector, and q is a vector of j by 1 dimension.

(Refer Slide Time: 04:25)



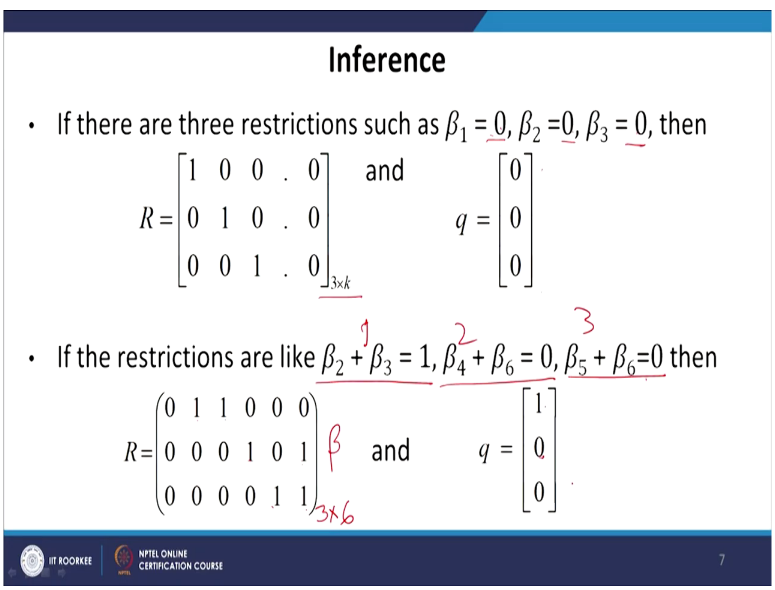


Now, each row of R is the coefficient in one of the restrictions, and the number of rows is equal to the number of restrictions. So, you can see that each row that is r_{11} r_{12} to r_{1k}. So, each row of R is the coefficient in one of the restrictions, and the number of rows is equal to the number of restrictions. For example, we test for hypothesis beta 2 equals 0. So, there is only one restriction such that beta 2 equals 0. Therefore, we write it as r equals 0 1 0 0, and so on.

Because in that case, my r multiplied by beta will be 0 1 0 multiplied by beta 1, beta 2, beta k. And if I multiply these two things, then you can see that this leads to only beta 2, and given that q equals 0, I have beta 2 equals 0. So, that is how one restriction can be written in this format. Similarly, if we have restrictions, such that beta 3 equals beta 4, which can alternatively be written as beta 3 minus beta 4 equals 0, then R is written in this format, again, we have only one restriction.

And that is why only one row is in R and when I write again, this R multiplied by beta in this format, then I will be having beta 3 minus beta 4 equals to 0. Similarly, if the restriction is such that beta 2 plus beta 3 plus beta 4 is equal to 1, then again, I have only one restriction. And this can be written like this because again, by multiplying R with a beta, we can obtain beta 2 plus beta 3 plus beta 4 equal to 1 where q equals 1. So, this is how we present the restrictions in matrix form.

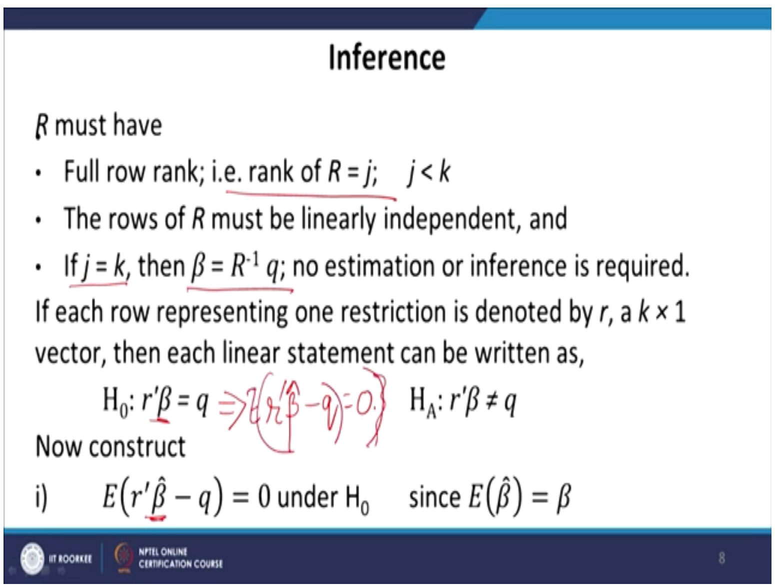
(Refer Slide Time: 06:30)



Now, suppose there are three restrictions, such that beta 1 equals 0, beta 2 equals 0, and beta 3 also equals 0. In that case, I have three restrictions. So, we have a 3 by K matrix, and a q equals 0, of course, because for all these restrictions, on the right-hand side, I have zeroes *(refer slide time: 06:30*). If the restrictions are like beta 2 plus beta 3 equals 1, beta 4 plus beta 6 equals 0 and beta 5 plus beta 6 equals 0, then you can see that again, there are three restrictions one, two, and three. So, R is a 3 by, and at the max, we assume that there are only 6 parameters.

So, this is R by 6 matrices. First of all, if I now multiply r with the beta, then you can see that we will be having beta 2 plus beta 3 that is equal to 1 beta 4 plus beta 6 equals 0 beta 5 plus beta 6 equals 0. So, this is how we can actually write multiple restrictions in matrix format. So, a single restriction can be written in matrix format. Similarly, multiple restrictions can also be written in matrix format.

(Refer Slide Time: 07:52)



Now, what are the conditions that are required, R must have full row rank. Now, we know that the rank of a matrix is defined as the maximum number of linearly independent rows or columns of the matrix. So, if we have the maximum number of linearly independent columns, that gives us the column rank of the matrix, and similarly, considering the maximum number of linearly independent rows gives us the row rank of the matrix.

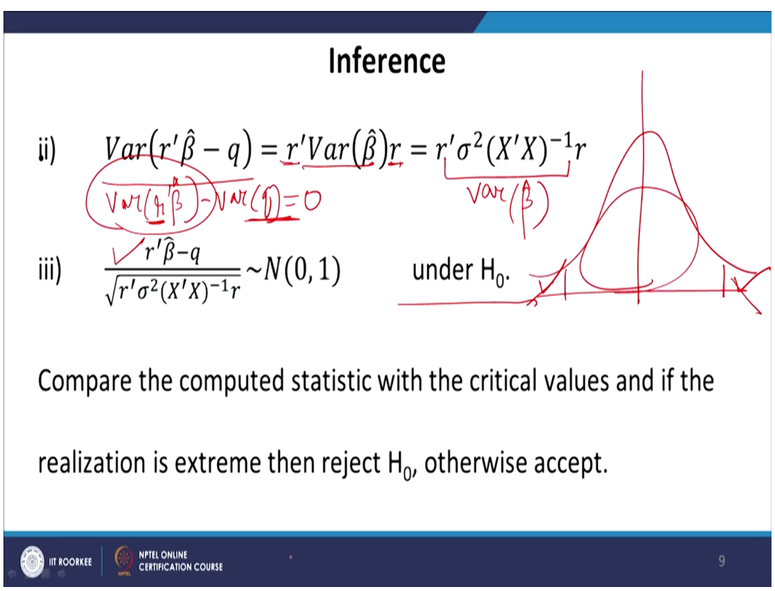
So, R must be a full row rank matrix. So, that rank of R equals to j and j must be less than equal to k, j is less than equal to k is a requirement, but again, j equal to k is not probably desirable. So, we prefer j less than k. Now, the rows of R must be linearly independent that is implied by the full row rank of R itself. And why do we must not have j equals to k? Because, theoretically or mathematically, if j is equal to k, then beta equals to R inverse q given that we know the value of R because R simply gives us the matrix of the restrictions and we also know the value of q.

We can always have an estimated value of the beta hat. So, there is nothing to infer. So, no estimation or inference is required if j equals to k. So, j less than k is the requirement. If each row representing one restriction is denoted by R or k by 1 vector, then each linear statement can be written as r prime beta equals to k that is our null hypothesis and the alternative hypothesis is r prime beta naught equal to q.

So, now we construct the expected value of r beta hat minus q under equal to 0 under the null hypothesis because expected value beta hat equals to beta, this has also been mentioned previously that all the null hypothesis and alternative hypothesis are framed in terms of the population parameter. But the population parameter is not observable.

So, while going for testing of hypothesis, we work with the sample estimate. So, that is how the population parameter beta is actually replaced with its sample estimate. So, r beta equals to q is my null hypothesis, which is equivalent to r prime beta minus q equals 0. And this beta is replaced with the beta hat.

(Refer Slide Time: 10:55)



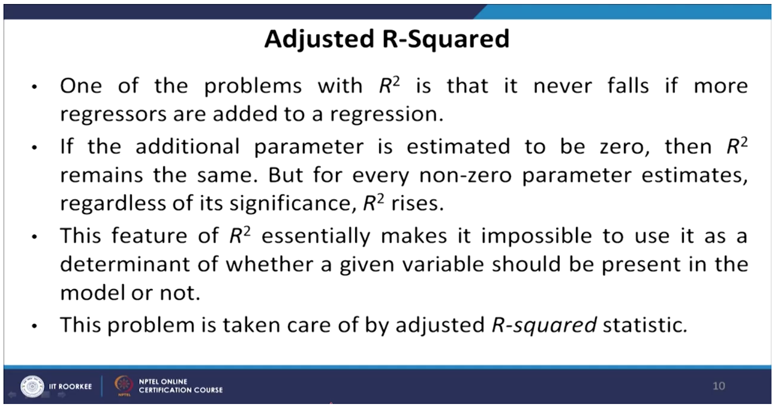
The variance of r beta hat minus q will be r prime variance of beta hat r equals to r prime sigma squared x prime x inverse r *(see the slide above*). Now, this is the variance of the beta hat, that has been already been derived for you in the previous modules. And by considering this expression, we can see that we will be writing variance of r prime beta hat minus variance of q, there should be a covariance term also. But of course, r prime beta hat and q, q being a constant there cannot be any covariance.

So, as a result of which we have the only variance of q subtracted from the variance of r prime beta hat, q again being a vector of constant terms cannot have any variations. So, the variance of q equals to 0, as a result of which I only have a variance of r prime beta hat. Now, when something constant r prime is also a matrix of, constant numbers, as a result of which there is no variations, and it should come out of the variance operator, when it comes out of the various operator, it should be a squared term.

So, since it is in matrix format, what we have is that r prime r, r prime is pre multiplied and r is post multiplied, and then in between, we have variance of beta hat. So, r prime sigma square x prime x inverse R is the expression for the variance of r prime beta hat minus q. And that is how, we have r prime beta hat minus q divided by the standard error of root of r variance of r prime beta hat minus q following a normal distribution with 0 mean and 1 standard deviation.

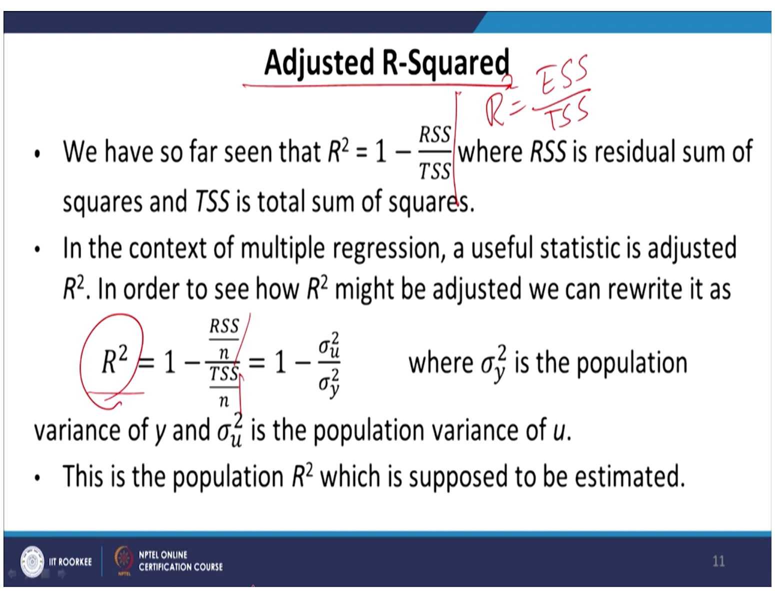
This is under the null hypothesis. Now, compare the computed statistic with the critical values and if the realization is extreme, then we will reject the null hypothesis otherwise accept the null hypothesis. So, this is a very similar thing once I have a computed or calculated test statistic, then I go for rejection or acceptance on the basis of whether it falls within my acceptance region or in the rejection region. So, if the values are very large, or extremely small, that is, for extreme values, they would fall in the rejection region and we will be rejecting the null hypothesis. Otherwise, I will be accepting the null hypothesis.

(Refer Slide Time: 13:28)



So, having talked about inferences next we talk about adjusted R squared. One of the problems with R square is that it never falls if more regressors are added to a regression. If the additional parameter is estimated to be 0, then R squared remains the same, or I would rather say it might remain the same. But for every non zero parameter estimate regardless of its significance, that is whether it is statistically significant or not, r square inevitably rises this feature of r square, essentially makes it impossible to use it as a determinant of whether a given variable should be present in the model or not. This problem is taken care of by adjusted R squared statistic.

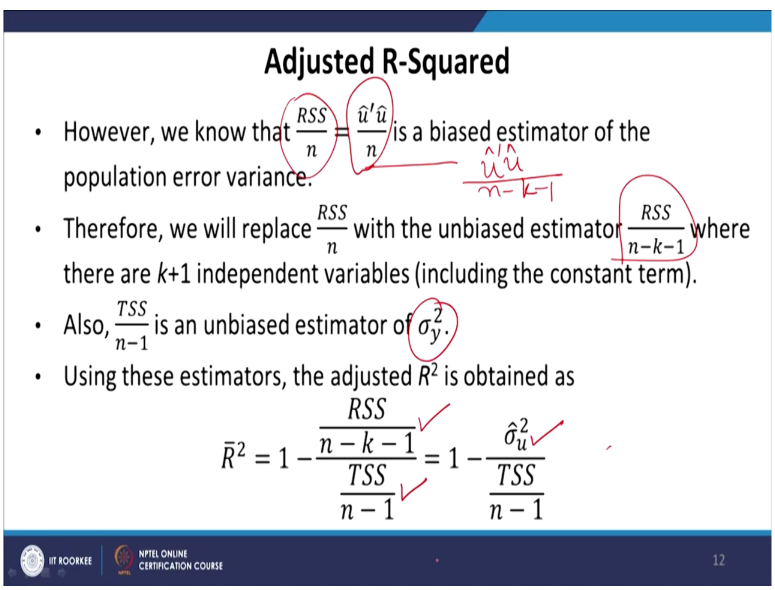
(Refer Slide Time: 14:15)

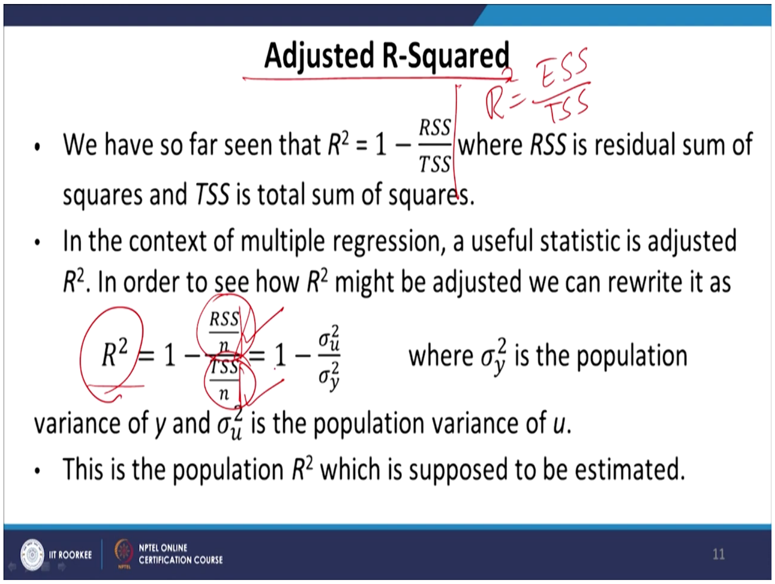


So, we have so far seen that R square is defined as 1 minus residual sum of squared divided by the total sum of square or R square is explained sum of squares divided by the total sum of squares *(see the slide above 14:15)*. So, this is one expression and the alternative expression is R squared is explained sum of squares divided by the total sum of squares. In the context of multiple regression, a useful statistic is adjusted R square. In order to see how R squared might be adjusted, we can rewrite the R square value as RSS by n and TSS also by n, both dividing, we are dividing both the numerator and the denominator by n. So, RSS by n is sigma square u that is the variance of the error term, and TSS by n is the variance of the dependent variable that is y, so sigma squared y. So, sigma squared y is the population variance of y and sigma squared u is the population variance of u.

Now, in this context, just let me tell you very briefly that, when we talk about adjusted R square, then actually the original R squared value is adjusted for degrees of freedom. Now, how do we adjusted for degrees of freedom? So, this is the population R square which is supposed to be estimated or that is generally estimated along with, a sample regression analysis.

(Refer Slide Time: 15:54)



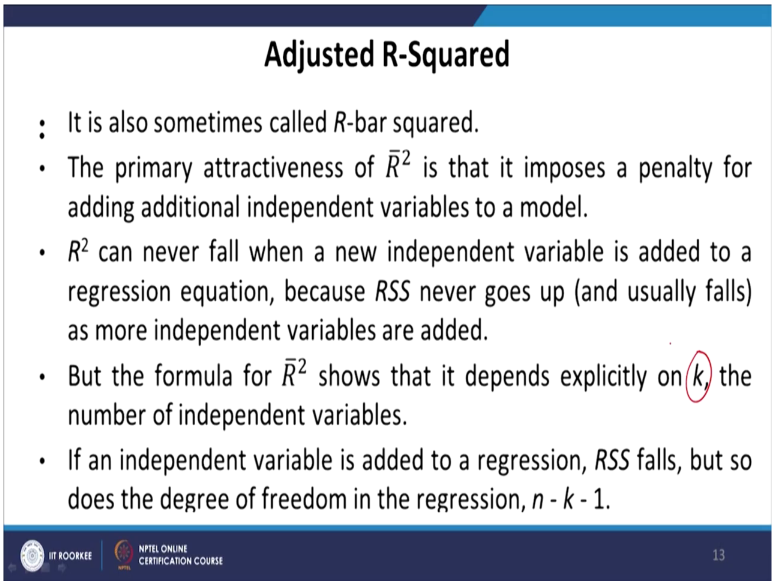


However, we know that RSS upon n is equal to u hat prime u hat divided by n is a biased estimator of the population error variance. This is something we have already proved that this thing measuring the population error variance is not actually equal to the residual variance. The residual variance is this or this could be this, but then this is actually not an unbiased estimator of the population error variance, what is the unbiased estimator? We have already discussed and proved that the unbiased estimator is u hat prime u hat divided by n minus k for k variable case n minus 2 for two-variable case or n minus k minus 1 in case I have k plus 1 independent variable including the constant term. So, basically n minus the number of independent variables gives us basically the unbiased estimator of the population error variance. Therefore, we will replace RSS by n with the unbiased estimator which is RSS upon n minus k minus 1 where there are k plus 1 independent variable including the constant term.

Also, TSS upon n minus 1 is an unbiased estimator of sigma square white for the same reason that we are providing, dividing it by the number of degrees of freedom. So, using these estimators the adjusted R squared is obtained as RSS upon n minus k minus 1 divided by TSS upon n minus k. This is what we had for the R squared. Now, that these expressions are now replaced with unbiased estimators of these expressions and these unbiased estimators are basically sample estimates adjusted for degrees of freedom.

So, sample estimates adjusted for degrees of freedom. So, this is what is R residual error where residual variance is a sigma hat square u or sigma hat u square divided by TSS upon n minus 1 the entire thing subtracted from 1 gives us the adjusted R square.

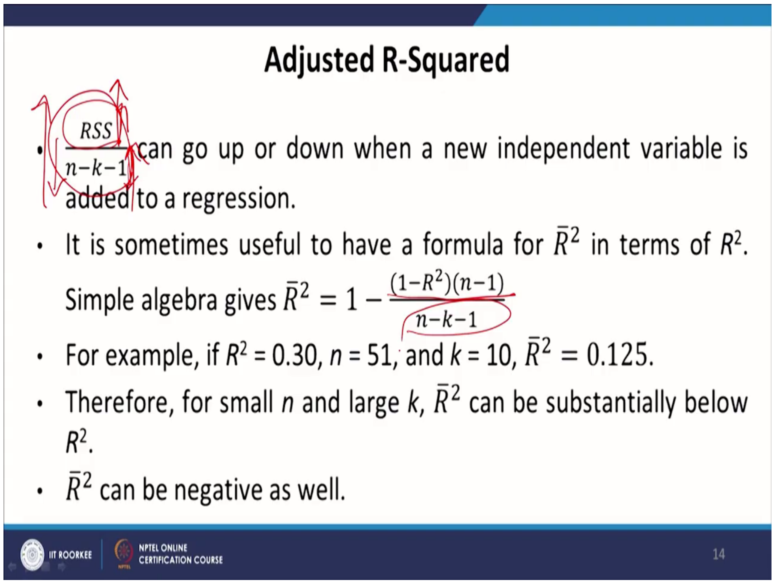
(Refer Slide Time: 18:19)



Adjusted R square also sometimes is called R bar squared. The primary attractiveness of R bar square is that it imposes a penalty for adding additional independent variables to a model, r square can never fall when a new independent variable is added to a regression equation because the residual sum of squares never goes up and it usually falls as more independent variables are added.

But the formula for R bar square shows that it depends explicitly on k that is the number of independent variables. It can be also k plus 1 the number of independent variables whatever you consider k or k plus 1, if an independent variable is added to a regression, RSS falls, but so does the degrees of freedom in the regression n minus k minus.

(Refer Slide Time: 19:16)

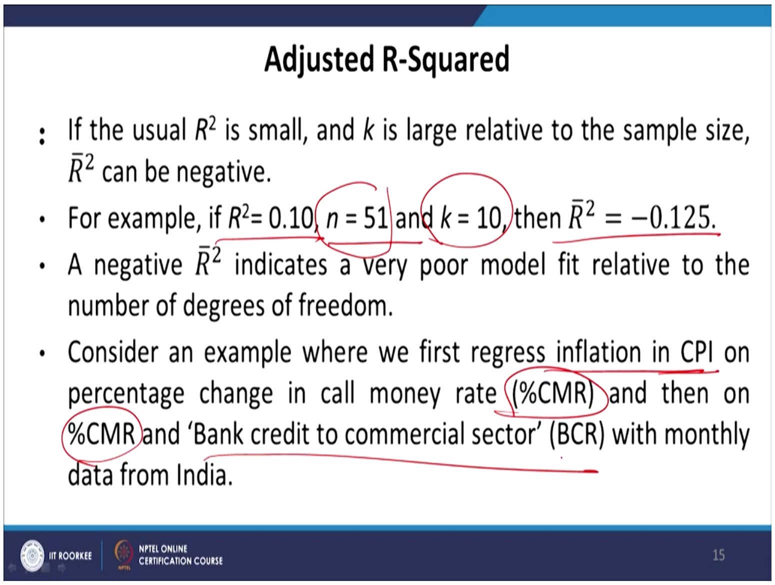


So, RSS upon n minus k minus 1 can go up or down when a new independent variable is added to a regression, whether this entire expression would go up or down, that depends on actually which one is going up or down by a larger proportion. So, if the increase in RSS, because of an addition of a new variable is more than the increase or the decrease in n minus k minus 1 more than the increase in n minus k minus 1, then, the entire thing will go up if the increase in RSS is less than the increase in n minus k minus 1, then the entire thing will go down. So, that is how we actually can have RSS upon n minus k minus 1 either going up or going down. So, unless and until an independent variable or the additional independent variable makes a significant contribution to the residual sum of square, RSS upon n minus k minus 1 will actually not go up.

And as a result of which adjusted R square can fall down. It is sometimes useful to have a formula for R bar square in terms of R square, simple algebra gives R bar square equals 1 minus 1 minus R square multiplied by n minus 1 divided by n minus k minus 1. Now, we take up some examples that if R square is 0.3, and we have a sample size of 51 i.e. n is equal to 51 and k is equals to 10 i.e. 10 independent variables including the constant term, then R bar square can be calculated as 0.125 *(Refer time 19:16)*.

So, therefore, for small n, that is when the sample size is not very large, and large k that is relative to the sample size, the number of independent variables is large, then R bar square can be substantially below R square. R bar square can be negative as well.

(Refer Slide Time: 21:22)

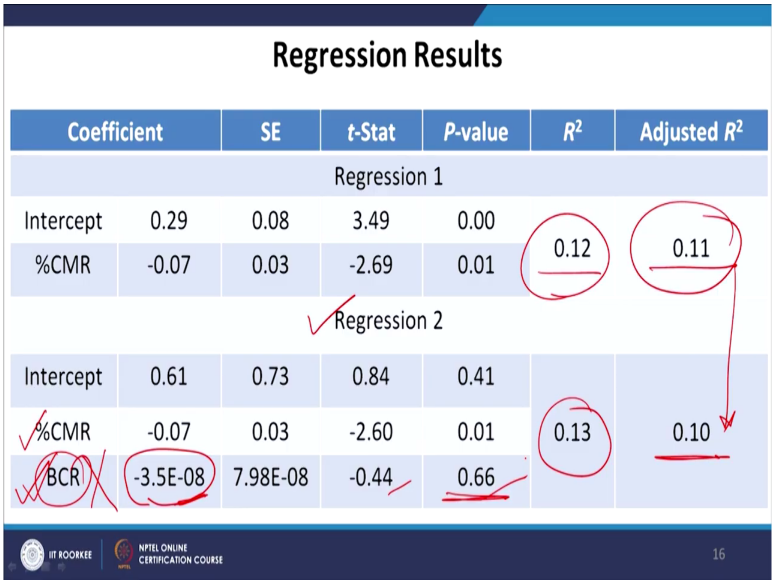


So, if the usual R square is small, and k is large, relative to the sample size, then the R bar square can be negative. For example, if R squared is 0.10, so we actually do not have a good fit anyway. And then we additionally have a relatively smaller sample and the relatively larger number of independent variables, then what happens our R bar square is negative. The Negative R bar squared indicates a very poor model fit relative to the number of degrees of freedom.

Consider an example, where we first regress inflation in CPI on the percentage change in call money rate, and then on the percentage change in call money rate and bank credit to the commercial sector with monthly data from India. So, what we are doing here is that we have one dependent variable to begin with, which is CPI, inflation in CPI, and first we run the analysis or do the regression with one independent variable, which is the percentage change in call money rate, percent CMR.

And in the second step, we do the analysis with two independent variables. One is again percent CMR. And the other one is bank credit to the commercial sector or BCR.

(Refer Slide Time: 22:42)



Regression Results

Coefficient		SE	t-Stat	P-value	R^2	Adjusted R^2
Regression 1						
Intercept	0.29	0.08	3.49	0.00	0.12	0.11
%CMR	-0.07	0.03	-2.69	0.01		
Regression 2						
Intercept	0.61	0.73	0.84	0.41	0.13	0.10
%CMR	-0.07	0.03	-2.60	0.01		
BCR	-3.5E-08	7.98E-08	-0.44	0.66		

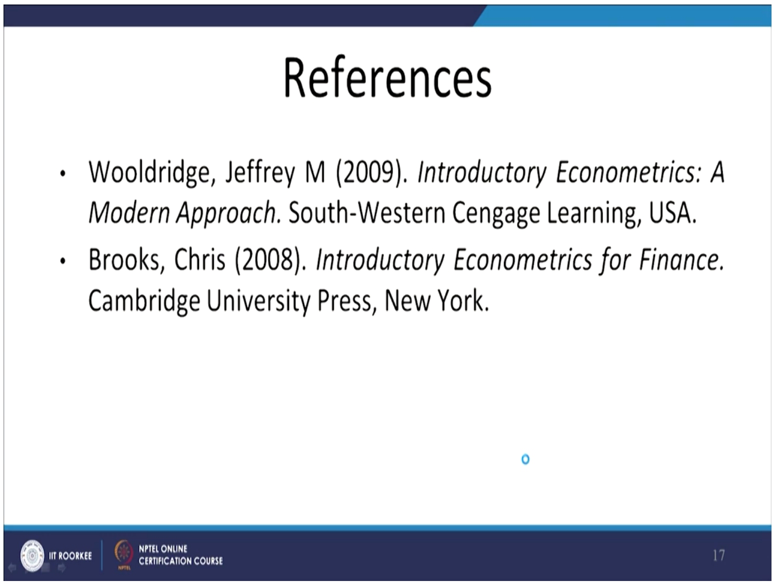
And now we check the results *(Refer time:22:42)*. See, we actually focus on the R square and adjusted R square values. So, you can see that this is the first regression from where we have R square and adjusted R square values at 0.12 and 0.11. Well, this is not actually a very good fit or rather a poor model fit, but I just want to give you some examples of the inclusion of variables and how it is going to impact the R square and the adjusted R square values.

So, in the next regression too when we have two variables, then you can see that this variable BCR that is bank credit to the commercial sector is actually a very small number which is close to 0, its P-value is also very high, which means it is not statistically significant. Nevertheless, our R squared value has gone up from 0.12 to 0.13. But then, this variable is a redundant variable that is reflected by the adjusted R square value or the adjusted R square value that has actually experienced a fall.

So, this indicates that this variable is actually redundant and should not be included in the regression analysis. Of course, there are other criteria that also suggest so, one of them or the most, obvious one is the t statistic or the P values, which indicate that we need not include this variable into the analysis. So, this is about inclusion or using adjusted R square as a measure of goodness of fit.

And that brings me to the end of the basic discussion on multiple regression analysis. Now, we will continue with some of the other features of multiple regression analysis based on some of the assumptions also of Gauss Markov theorem in the next module.

(Refer Slide Time: 24:34)



These are the books that I have followed for preparing the content of this video. Thank you.