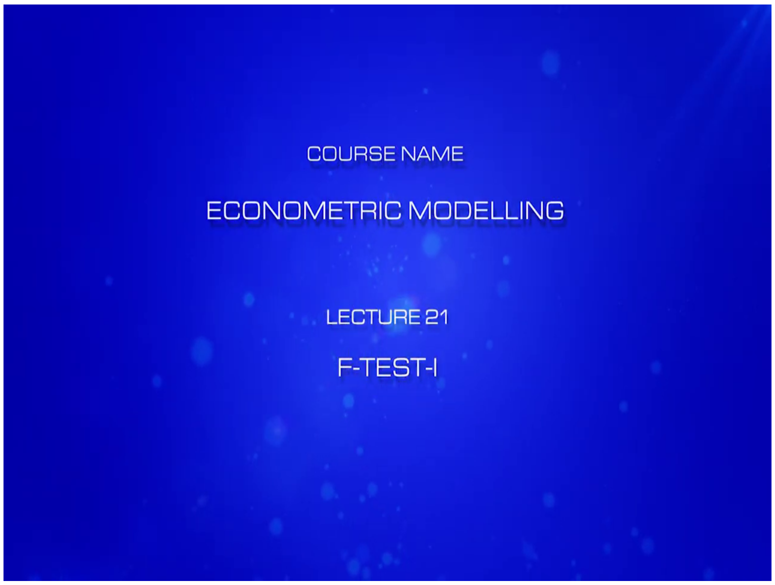
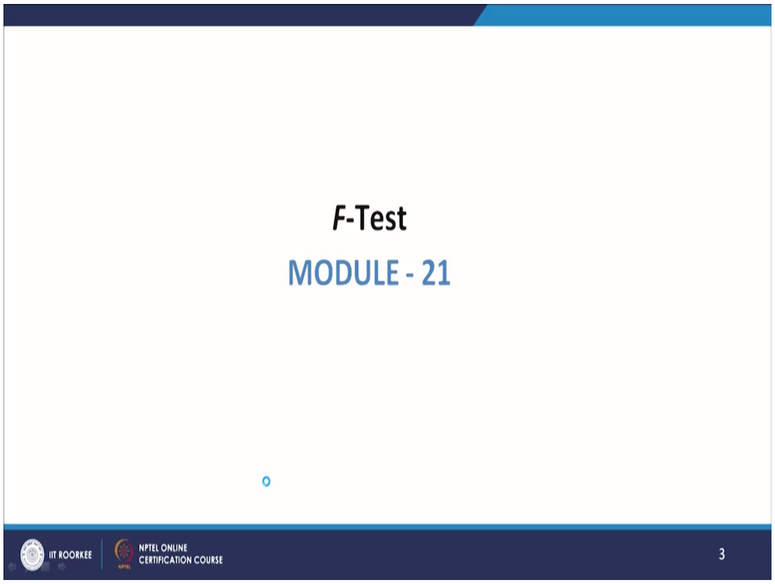
Econometric Modelling
Professor Sujata Kar
Department of Management Studies
Indian Institute of Technology, Roorkee
Lecture 21
F-Test-I

(Refer Slide Time: 00:11)



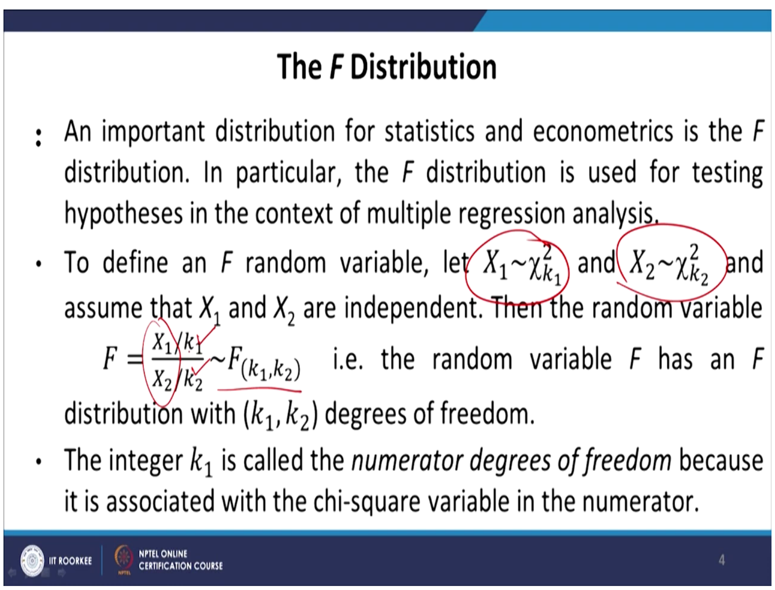
This is Module 21. We are now discussing statistical inferences. So under this, we have discussed so far, t-test and Wald test. Next, I am going to deal with F-test which is one of the most commonly used tests.

(Refer Slide Time: 00:50)



In this module, first I will discuss the basic theories related to F distribution, and then in the next module, we will be presenting some examples and applications of F-test.

(Refer Slide Time: 01:10)

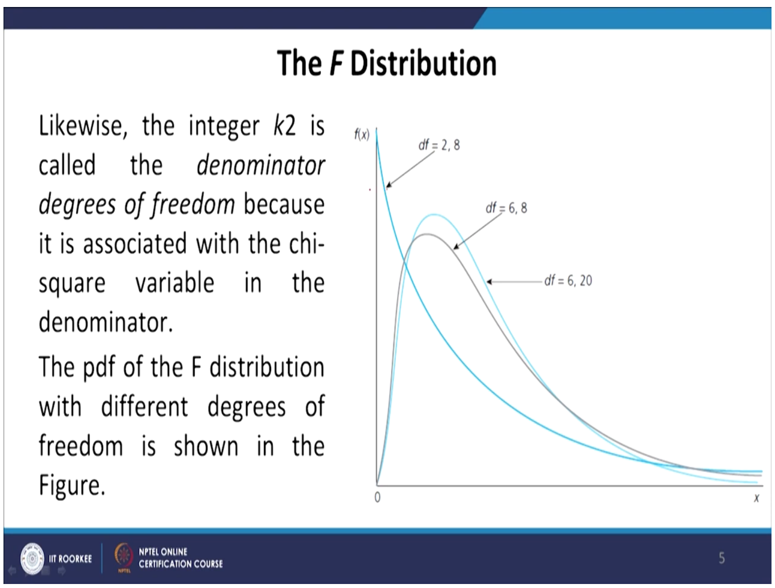


So F distribution is an important distribution for statistics and econometrics. In particular, the F distribution is used for testing hypothesis in the context of multiple regression analysis. This is primarily because we actually test multiple hypothesis simultaneously.

To define an F random variable, let X_1 follow the chi-square distribution with k_1 degrees of freedom, and X_2 follows a chi-square distribution with k_2 degrees of freedom, and X_1 and X_2 are independent. Then, the random variable, which would be obtained as a ratio of X_1 divided by its degrees of freedom, X_2 divided by its degrees of freedom will follow an F distribution with k_1 and k_2 degrees of freedom.

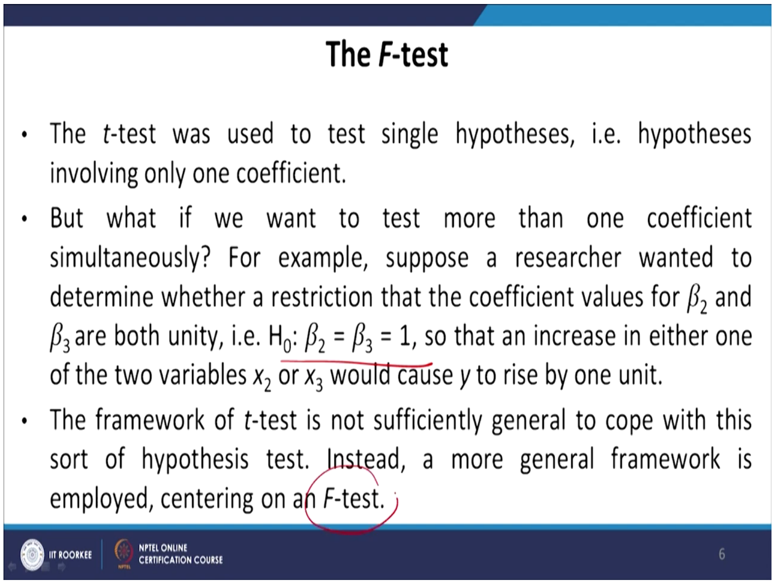
The integer k_1 is called the numerator degrees of freedom because it is actually associated with the numerator.

(Refer Slide Time: 02:12)



And similarly, integer k_2 is called the denominator degrees of freedom because it is associated with the chi-square variable in the denominator. The pdf of the F distribution with different degrees of freedom is shown in the figure above where you can see that unlike the t and normal distribution, the F distribution actually does not have a nice bell-shaped curved. Rather, it is a skewed distribution and the skewness is actually more, it almost takes the shape of a j distribution with the decline in the numerator degrees of freedom.

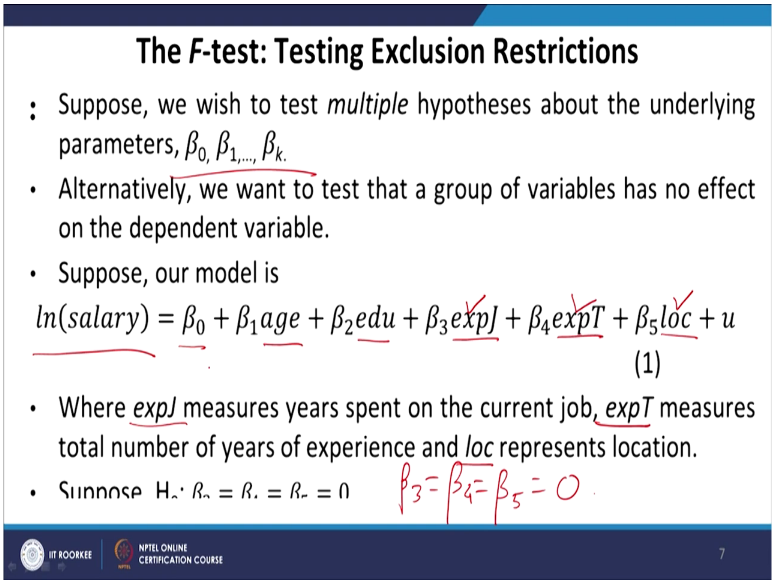
(Refer Slide Time: 02:50)



The t-test was used to test the single hypothesis, that is, hypothesis involving only one coefficient. But what if we want to test more than one coefficient simultaneously? For example, suppose a researcher wanted to determine whether a restriction that the coefficient values for beta 2 and beta 3 are both unity. That is, my null hypothesis is beta 2 equals beta 3 equals 1, so that an increase in either one of the two variables, that is X_2 or X_3 would cause Y to rise by 1 unit.

The framework of the t-test is not sufficiently general to cope with this sort of hypothesis test. Instead, a more general framework is employed, centering on an F-test.

(Refer Slide Time: 03:37)

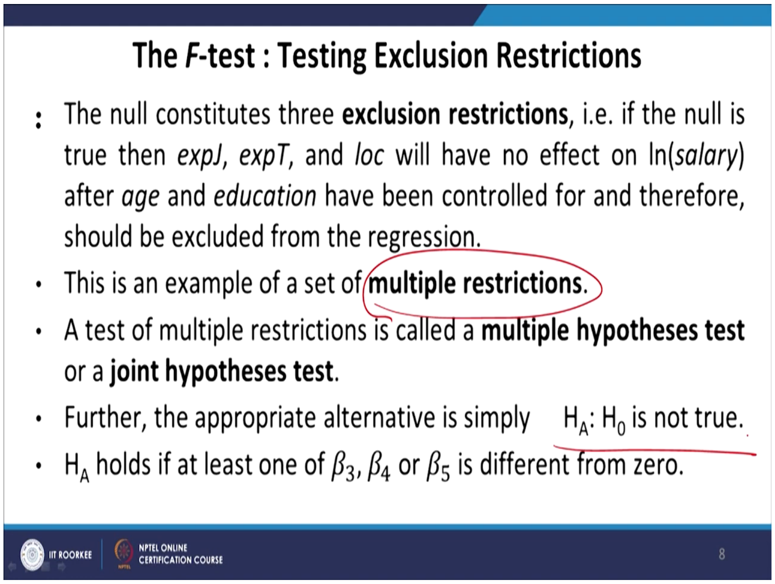


Now, suppose we wish to test multiple hypothesis about the underlying parameters beta naught to beta k. So, this is a general specification where we have k variables. Alternatively, we want to test that a group of variables have no effect on the independent variable simultaneously. So, we are basically examining the impact of a group of variables, simultaneously on the dependent variable.

So, we take an example. Suppose our model is the logarithm of salary is equal to beta 0, the parameter associated with the constant term, beta 1; age, beta 2; education, beta 3; experience J, beta 4; experience T and beta 5; loc plus u, where experience J measures years spent on the current job, experience T measures the total number of years of experience and loc presents the location.

Now, suppose our null hypothesis is beta 3 equals beta 4 equals beta 5 equals 0. So, we are going to test whether the null hypothesis states that these three variables do not impact the dependent variable which is the logarithm of salary.

(Refer Slide Time: 05:00)

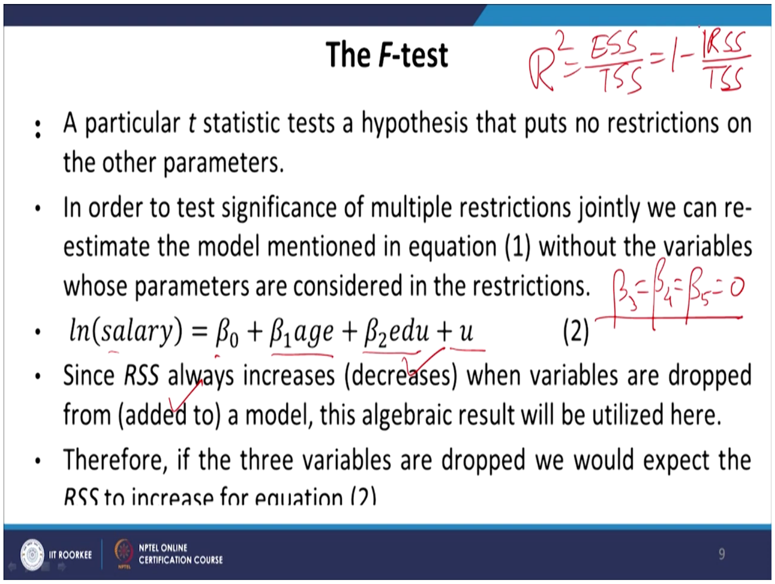


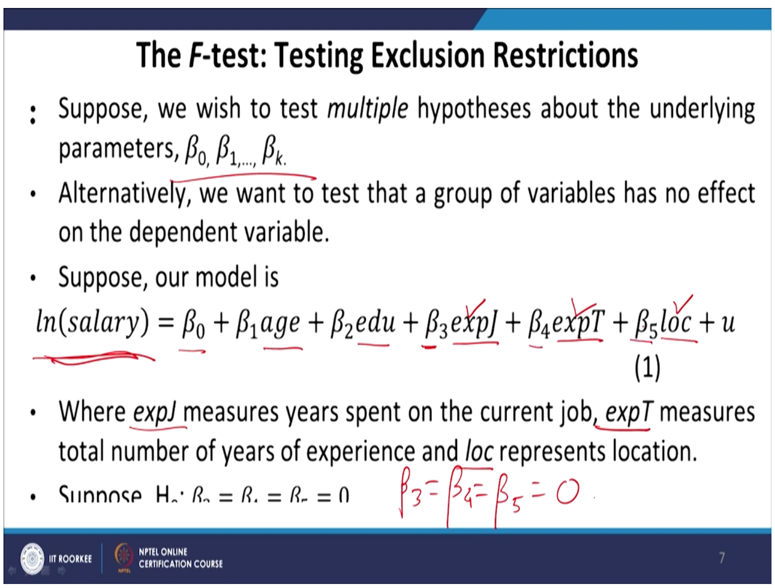
The null constitutes three exclusion restrictions. So, these are called exclusion restrictions because through these restrictions, we are excluding some variables. That is, if the null is true, then expJ, expT and loc will have no effect on log salary after age and education have been controlled for, and therefore should be excluded from the regression. This is an example of a set of multiple restrictions. We are having multiple restrictions, restrictions on three parameters, simultaneously.

A test of multiple restrictions is called a multiple hypothesis test or a joint hypothesis test. Further, the appropriate alternative is simple H_A:H_0 is not true. H_A holds if at least one of beta 3, beta 4 or beta 5 is different from zero.

So in that joint hypothesis, test of joint significance of beta 3, beta 4 and beta 5, even if any one of them is different from 0, then we will actually not reject the alternative hypothesis. That is, the alternative hypothesis becomes true.

(Refer Slide Time: 06:15)





A particular t statistic tests a hypothesis that puts no restrictions on the other parameters. So, when we say that t statistic or t-test is used for only one hypothesis, then that implies that there are no restrictions on other parameters.

In order to test the significance of multiple restrictions jointly, we can re-estimate the model mentioned in equation 1 without the variables whose parameters are considered in the restrictions.

So what is my restriction, or what are my restrictions? My restrictions are: what is stated in the null hypothesis, that is, beta 3 equals beta 4 equals beta 5 equals 0.

Now, when we incorporate the null hypothesis or the restrictions under the null hypothesis in the original model, then that becomes our restricted model. So, if these are incorporated, then you can see that equation 1 will not have beta 3; 0, beta 4; 0, beta 5; 0. So these three variables will be dropped from the equation.

And what I am left with is simply log salary equals beta 0 plus beta 1 age plus beta 2 education plus the error term. Since, RSS, that is, residual sum of square always increases when variables are dropped from or they decrease when variables are added to a model, this algebraic will be utilized here.

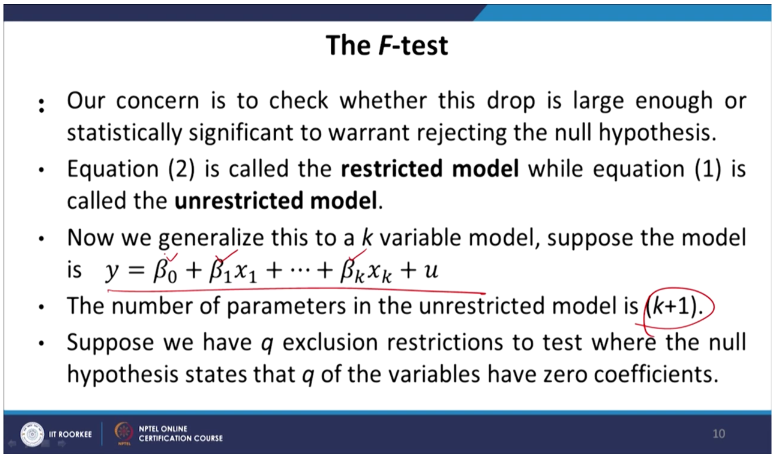
So, therefore, if the three variables are dropped, we would expect the residual sum of the square to increase for equation 2.

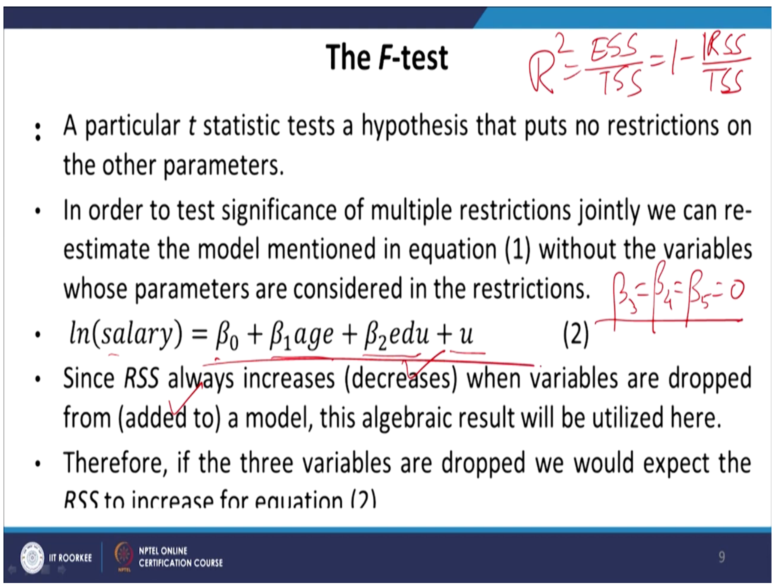
If you remember, we had discussed that R square, the concept of adjusted R square was introduced because the problem with R square is that R square always tends to increase the moment we increase the number of independent variables regardless of whether they are statistically significantly contributing to the explanation of the dependent variable.

So, unless and until the coefficient value is exactly equal to 0, R square would always increase whenever we increase the number of the independent variables. And since R square is defined as the explained sum of squares divided by the total sum of the square, alternatively 1 minus the residual sum of a square divided by the total sum of the square.

So this is something important, which implies that RSS actually decreases whenever there are variables that are dropped from an equation.

(Refer Slide Time: 08:55)



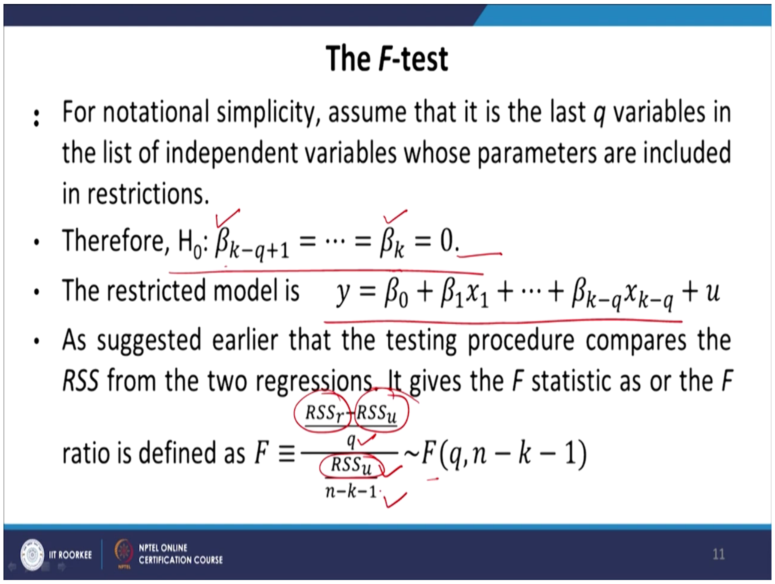


Our concern is to check whether this drop is large enough or statistically significant to warrant rejecting the null hypothesis. So equation 2 is called the restricted model. While equation 1 is called the unrestricted model, the original model is the unrestricted model, where we do not include any restrictions.

Now we generalize this to a k variable model, suppose the model is, the k variable model, the way we write it, y equals beta 0 plus $\text{beta}_1 X_1$ plus beta $_k$ X $_k$ plus u. The number of parameters in the unrestricted model is of course k plus 1, that is, these k parameters plus one are associated with the constant term.

Suppose, we have q exclusion restrictions to test where the null hypothesis states that q of the variables have 0 coefficients.

(Refer Slide Time: 09:50)



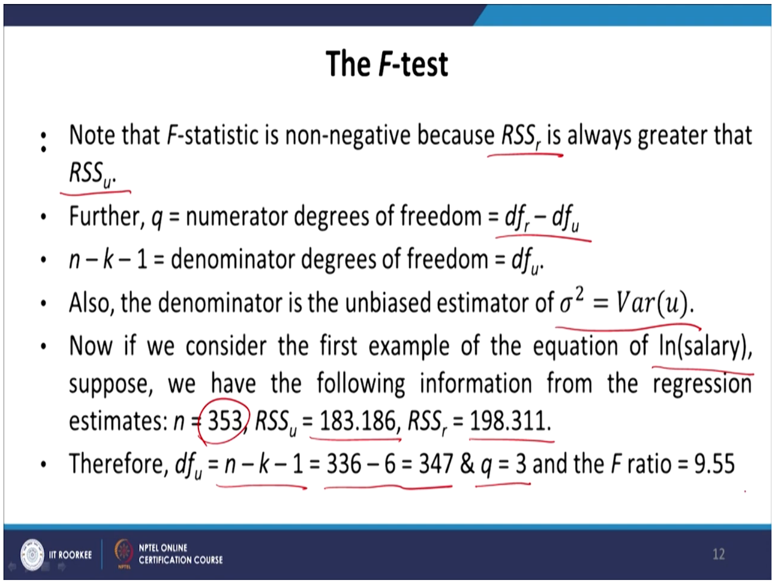
For notational simplicity, assume that it is the last q variables in the list of independent variables whose parameters are included in the restriction. Therefore, our null hypothesis is, beta k minus q plus 1 equals beta k minus q plus 2, and so on up to beta k and all of them are equal to 0 is my null hypothesis.

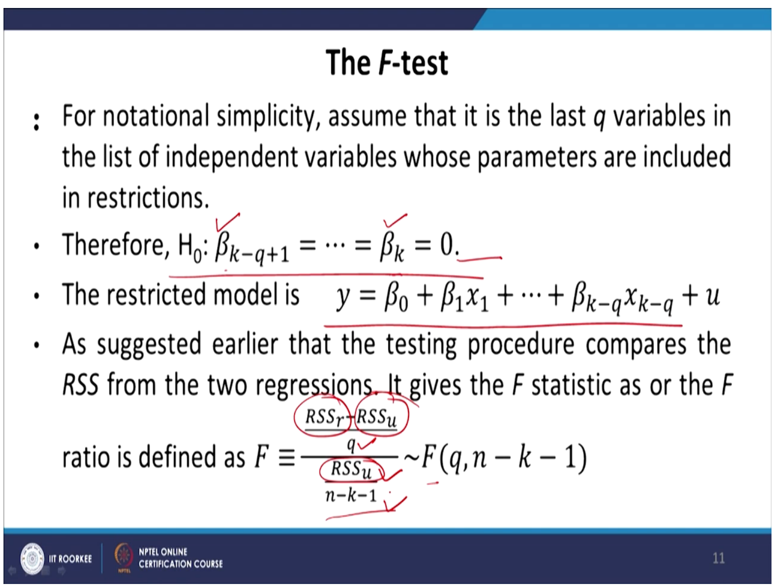
The restricted model is y equals beta 0 plus beta $_1$ X $_1$ up to beta k minus q X k minus q plus u. So, the restricted model does not include these parameters. As suggested earlier that the testing procedure compares the RSS, that is the residual sum of squares from the 2 regressions.

It gives the F statistic as, or the F ratio defined as RSS from the restricted regression minus RSS from the unrestricted regression divided by q, that is the numerator degrees of freedom, the number of restrictions divided by RSS, the unrestricted regression that is, RSS from the unrestricted regression divided by n minus k minus 1 degree of freedom associated with the denominator.

And it follows a q square distribution with k degrees of freedom and n minus k minus 1 degree of freedom.

(Refer Slide Time: 11:20)





Note that F-statistic is non-negative because RSSr is always greater than RSSu. We have just discussed that when we drop variables from an equation, or from a model, then the way R square falls down, R square which is 1 minus RSS, so RSS is supposed to go up. The residual sum of the square goes up.

And as a result of which we would always have RSSr greater than RSSu, that is RSS obtained from the restricted regression greater than RSS obtained from the unrestricted regression.

Further q equals the numerator degrees of freedom, which is again, the degrees of freedom associated with restricted regression minus the degrees of freedom associated with the unrestricted regression.

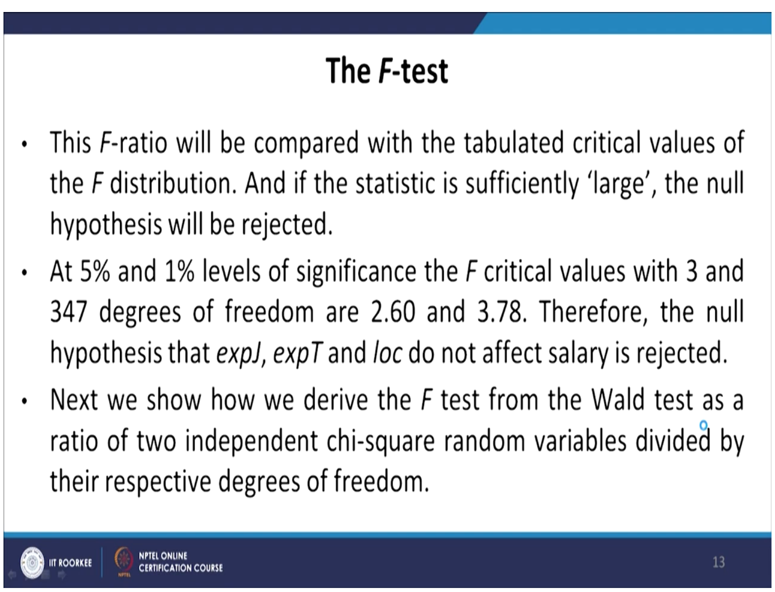
And finally, n minus k minus 1, which is the denominator degrees of freedom, that is equal to the degrees of freedom associated with the unrestricted regression. Also, the denominator is the unbiased estimator of sigma square equals to the variance of u.

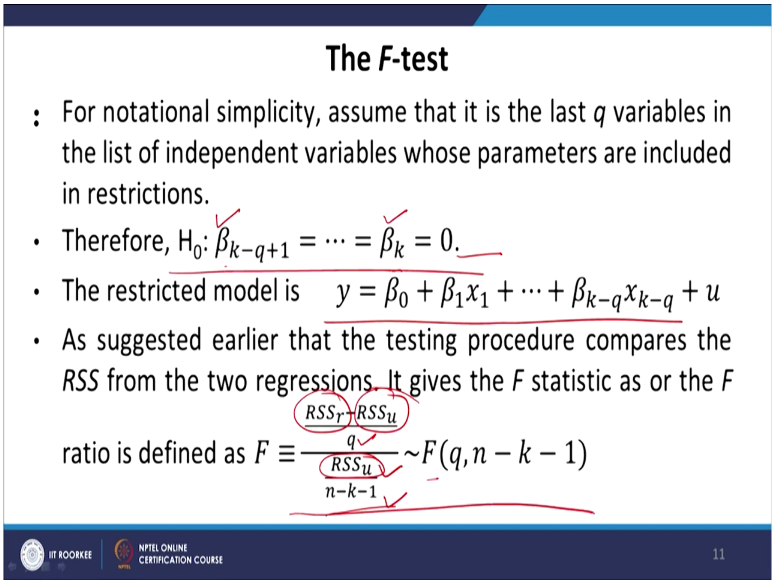
What is the denominator? The denominator is RSSu, which is the residual sum of the square from the unrestricted regression divided by minus, n minus k minus 1. So this is actually the unbiased estimator of the population error variance.

Now, if we consider the first example of the equation of logarithm of salary, suppose we have the following information from the regression estimates. That is, we have a total number of observations which is equal to 353, RSSu is obtained as 183.186, RSSr is obtained as 198.311, therefore, the degrees of freedom would be obtained as n minus k minus 1, which is 347.

We know that there are three restrictions. So, q is equal to 3. And we calculate the F-ratio as 9.55.

(Refer Slide Time: 13:24)



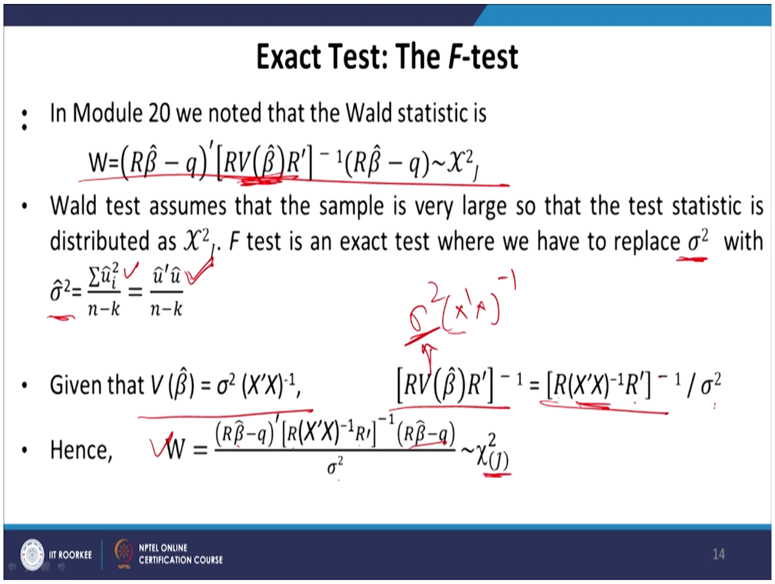


The F-ratio will be compared with the tabulated critical values of the F distribution. And if the statistic is sufficiently large, the null hypothesis will be rejected.

At 5 percent and 1 percent levels of significance, the F critical values with 3 and 347 degrees of freedom are 2.60 and 3.78, respectively. Therefore, the null hypothesis that experience J, experience T, and location do not affect salary is rejected.

Next, we show how we derive the F-test from the Wald test as a ratio of two independent chi-square random variables divided by their respective degrees of freedom. So, where we are getting this relationship, that F is defined like this is now explained.

(Refer Slide Time: 14:14)



In Module 20, we noted that the Wald statistic is this. This is something we derived. That Wald statistic follows the chi-square distribution with j degrees of freedom. Wald test assumes that the sample is very large. We mentioned that it is a large sample or asymptotic test so that the test statistic is distributed as chi-square j.

F-test is an exact test, where we have to replace the sigma square. The sigma square which is embedded here, the variance of beta hat contains sigma square. That sigma square will be replaced with a sigma hat square because F-test is an exact test, test, it is not an asymptotic test. So, we work with samples. The sample need not be a very large sample.

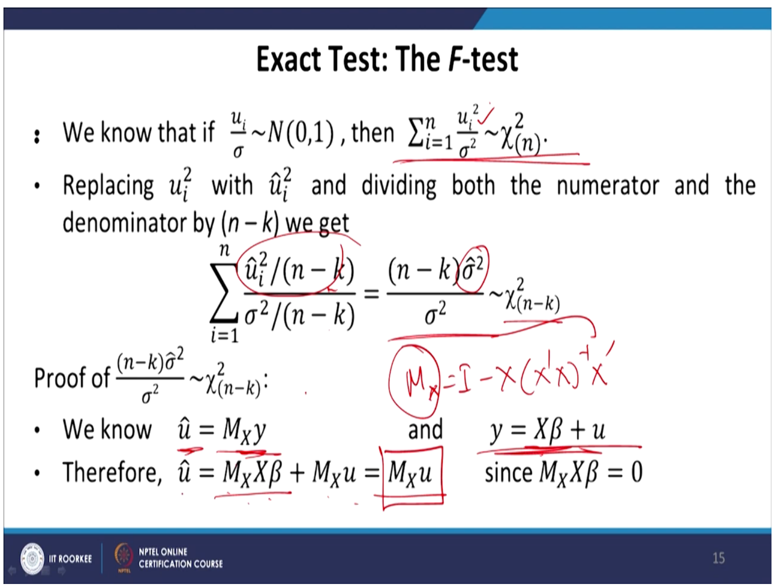
So, as a result of which, the sample residual variance replaces the population error variation. So, sigma hat square becomes a summation, which is actually equal to summation ui hat square divided by n minus k or summation u hat prime u hat divided by n minus k. They are the same thing. This is written by summing individual residuals over ith observations.

And this is, we write, u hat prime u hat, we write when we are actually using the matrix or vector notation. Now, given that variance of beta hat equals sigma square into X prime X inverse, this expression, that is, R variance of beta hat R prime, actually, and the entire thing inverse becomes R X prime X inverse R prime whole inverse divided by sigma square.

This is because here variance of beta hat will be replaced with sigma square X prime X inverse. Now, sigma square being a constant would actually come out. And since it will also have a, raised to the power minus 1, it will go to the denominator. And X prime X inverse remains here.

So, that Wald statistic can be now written as R beta hat minus q prime, R X prime X inverse R prime whole inverse. So, this is actually the part. Sigma square comes down, R beta hat minus q, which follows a chi-square distribution with j degrees of freedom.

(Refer Slide Time: 16:42)



We know that ui divided by sigma follows a standard normal distribution. Then, summation ui square divided by sigma square follows a chi-square distribution with n degrees of freedom.

This is what we discussed while introducing the t-test. That, if a variable has a chi-square distribution with one degree, if a variable has a standard normal distribution, then squaring it, we get chi-square distribution with one degree of freedom and by summing it up, summing up the squares over n observations, we actually get chi-square distribution with n degrees of freedom. So, exactly that thing is written here.

Now, replacing ui square with ui hat square, that is, the sample estimate of the population error and dividing both, the numerator and the denominator by n minus k, what we get? We are dividing it by n minus k and also dividing the denominator by n minus k.

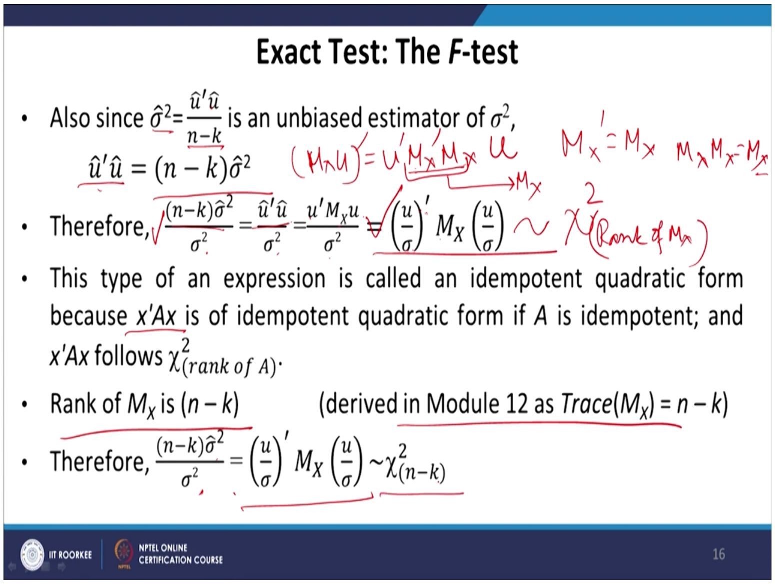
Now, this is sigma hat square. And this, n minus k goes up. So this is what we obtain. And this follows a chi-square distribution with n minus k degrees of freedom. This is something I actually used while discussing the t-test, but did not prove it there. Now, I will be proving that why this expression follows a chi-square distribution with n minus k degrees of freedom.

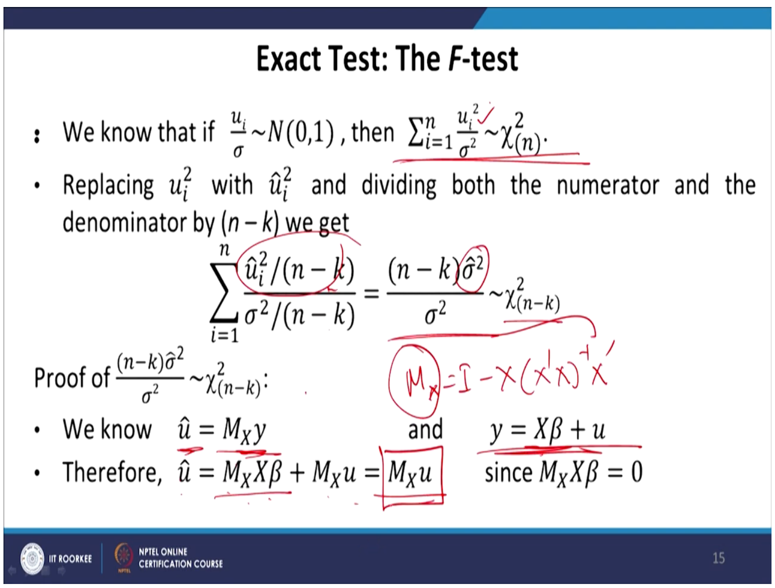
So, we know that u hat is equal to MXy. If you remember, MX was defined as a projection matrix. Why? MX equals to I minus X into X prime X inverse X prime. And this projection matrix was orthogonal to the column space of X, which implies that this MXy, MX multiplied by X would actually give us 0.

Now, we know that y equals X beta plus u. By plugging the expressions, we obtain MX X beta plus MXu. We know that MXX equals 0, as this projection matrix is orthogonal to the column space of X. So, MX multiplied by X becomes 0, which was also proved earlier.

So we are left with only MXu. Now, this is an important expression. That is, u hat equals to MXu, which is going to be used very soon.

(Refer Slide Time: 19:12)





Also, since sigma hat square equals u prime, u hat prime u hat divided by n minus k, and it is an unbiased estimator of sigma square, we can write by just changing sides. So, u hat u prime becomes sigma square multiplied by n minus k.

Therefore, n minus k sigma hat square divided by sigma square can be written as u hat prime u hat divided by sigma square. We just derived that u hat is MXu. So, u hat prime will be MXu prime, which is equal to u prime MX prime. And then u hat will be Mxu *(refer to slide time 19:12)*.

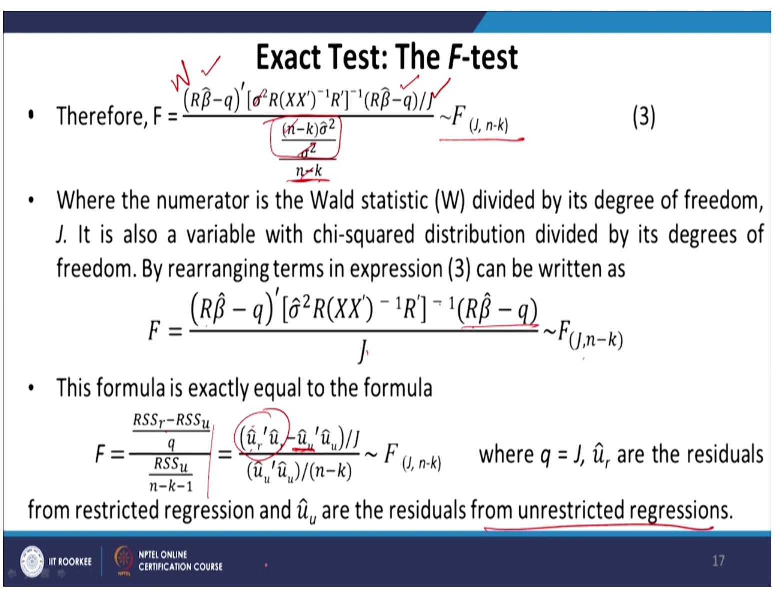
We proved earlier that MX is a symmetric matrix. So, MX prime equals MX and this is also an idempotent matrix. So, MX into MX is actually equal to MX. So, these two terms actually become equivalent to MX and that is how we have u prime MXu divided by sigma square.

Now, I am just writing it in this fashion, that u prime divided by sigma, MX multiplied by u prime sigma. This type of expression is called an idempotent quadratic form because x prime Ax is of the idempotent quadratic form if A is idempotent and x prime Ax follows a chi-square distribution with the rank of A equals to the degrees of freedom.

So, u by sigma prime MXu by sigma prime would also follow a chi-square distribution with the rank of MX as the degrees of freedom. Now, the rank of MX is n minus k. This was derived in Module 2, where we derived that trace of MX is n minus k. So, therefore n minus k sigma hat square divided by sigma square, this expression becomes u prime, u by sigma prime.

So this expression MXu sigma follows a chi-square distribution with n minus k degrees of freedom.

(Refer Slide Time: 21:33)



Therefore, our F-statistic, which is basically a ratio of two chi-square distribution divided by, with their respective degrees of freedom, we have in the numerator, the Wald statistic divided by its degrees of freedom which is J, and in the denominator, we have this expression which is again a chi-square distribution and divided by its degrees of freedom n minus k *(refer to slide time 21:33).*

And this would follow an F distribution with J, and n minus k degrees of freedom. So, this thing is written here. Now, by rearranging terms in expression 3, this can be written as, or rewritten as, you can see sigma square and sigma square cancels out because this sigma square is under 1 inverse. They cancel out, n minus k n minus k cancels out.

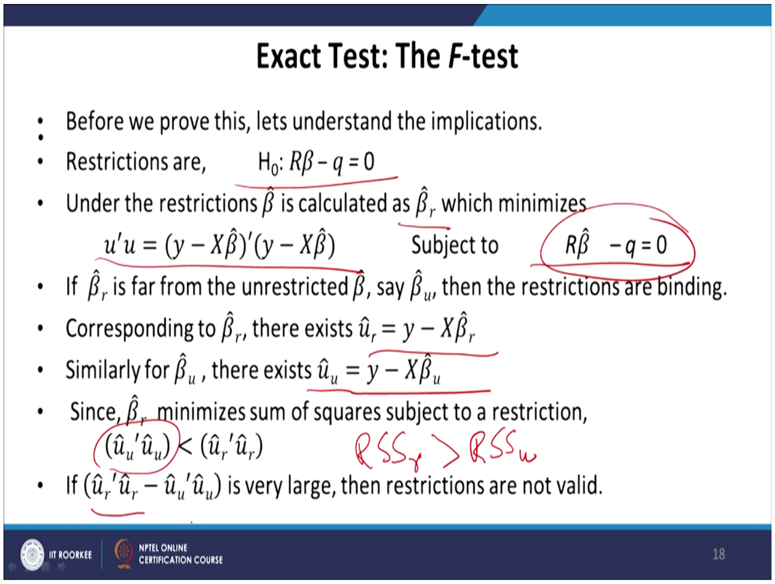
And the sigma hat square, being in the denominator, goes up into the inverse operator. So, what we have is sigma R b hat, R beta hat minus q prime, the term as it is. Sigma hat square, moving up, RX prime X inverse R prime whole inverse multiplied by R beta hat minus q, the term as it is. And this J comes down.

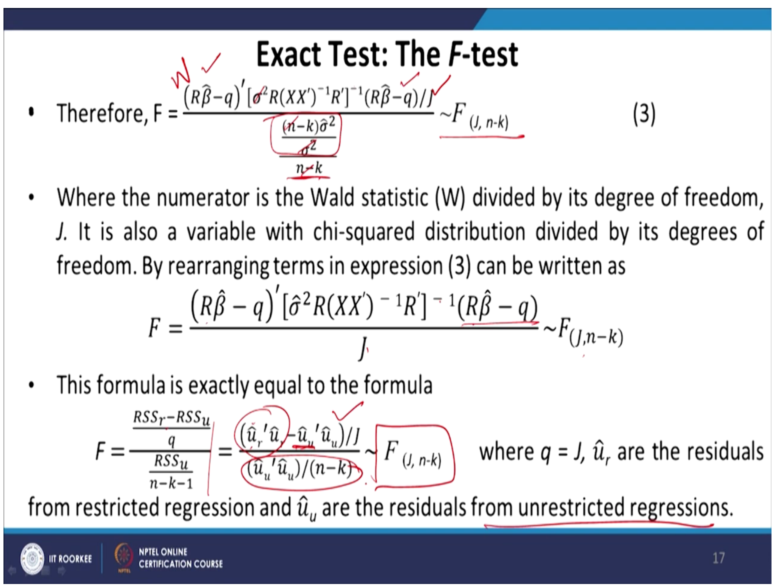
This follows an F distribution with J and n minus k degrees of freedom. And this formula is exactly equal to the formula, which we have earlier used while considering the restricted and the unrestricted regressions.

Now, before we prove that these two are exactly equal, we would actually talk about implications. But before that, I will just let you know that what we are using here is that, a term u hat r, which is basically residuals obtained from the restricted regression, u hat u, which is, or uu hat, which is the residuals obtained from the unrestricted regression. These terminologies would be used. And as you can understand that they are basically the same thing.

RSSr is the residual sum of squares. Residual sum of the square is obtained as, from the restricted regression is obtained as ur hat prime multiplied by ur hat.

(Refer Slide Time: 23:58)





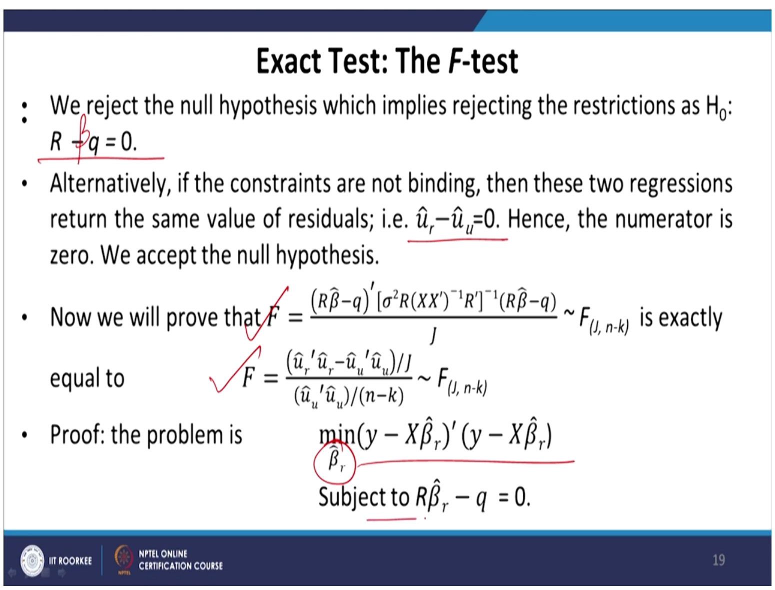
So, talking about the implications, the restrictions are, first of all, H0 equals R beta minus q equals 0. Another restriction beta hat is calculated as beta r hat which minimizes u prime u equals y minus x beta hat prime multiplied by 1 minus x beta hat. This is the usual format and it is a subject, this is subjected to R beta hat minus q equals to 0, which is basically the null hypothesis *(refer to slide time 23:58)*.

If beta hat r is far from the unrestricted beta hat, suppose we call the unrestricted beta hat as beta hat u, then the restrictions are binding. Corresponding to beta r hat, there exists Ur hat which is equal to y minus x beta r hat. Similarly, for beta u hat, there exists uu hat which is equal to y minus x beta u hat.

Since beta r hat minimizes the sum of squares subject to restrictions, we have uu hat prime multiplied by uu hat is less than ur hat prime multiplied by ur hat. Ideally, what it again says, the same thing that RSS from restricted regression will be greater than the RSS from the unrestricted regression.

So if Ur hat prime multiplied by Ur hat minus uu hat prime multiplied by uu hat is very large, then the restrictions are not valid. If this is very large, then you can see that the numerator of the F statistic will be large, which implies that given this value, the F statistic will be large and it will become difficult for us not to reject that null hypothesis. So most often, we would reject the null hypothesis, but the restrictions are not binding.

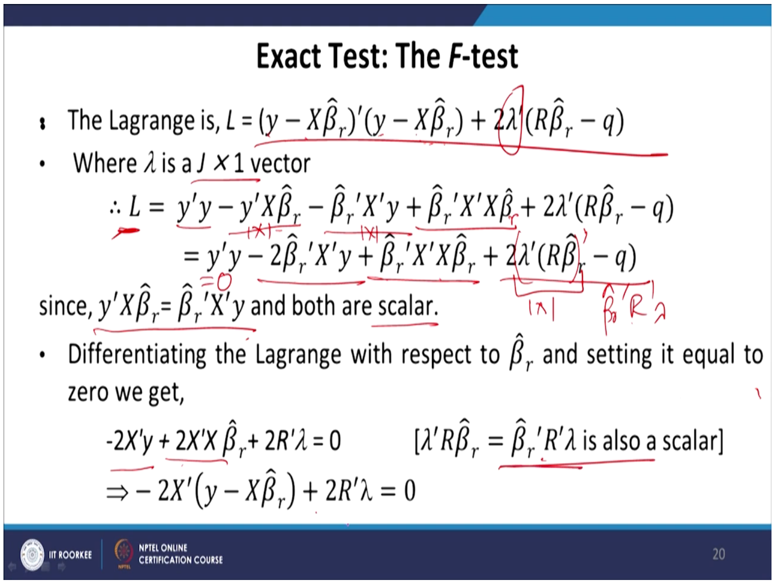
(Refer Slide Time: 25:58)



We reject the null hypothesis which implies rejecting the restrictions as R minus q is equal to 0, R minus beta q is equal to 0. Alternatively, if the constraints are not binding, then these two regressions return the same value of residuals, that is ur hat equals to uu hat, or ur hat minus uu hat equals 0, hence the numerator is 0. We accept the null hypothesis. *(refer to slide time 25:58)*

Now, we will prove that this expression is exactly equal to this expression. So, our problem is that we are minimizing y minus X beta r hat prime multiplied by 1 minus X beta r hat by choosing beta r hat subject to these restrictions.

(Refer Slide Time: 26:53)



Now, since this is a restricted optimization or minimization problem, we would set the Lagrange as this where lambda hat is basically a constant term. Here it is actually a vector of constant terms. So this is a vector J by 1 constant terms *(refer to slide time 26:53)*

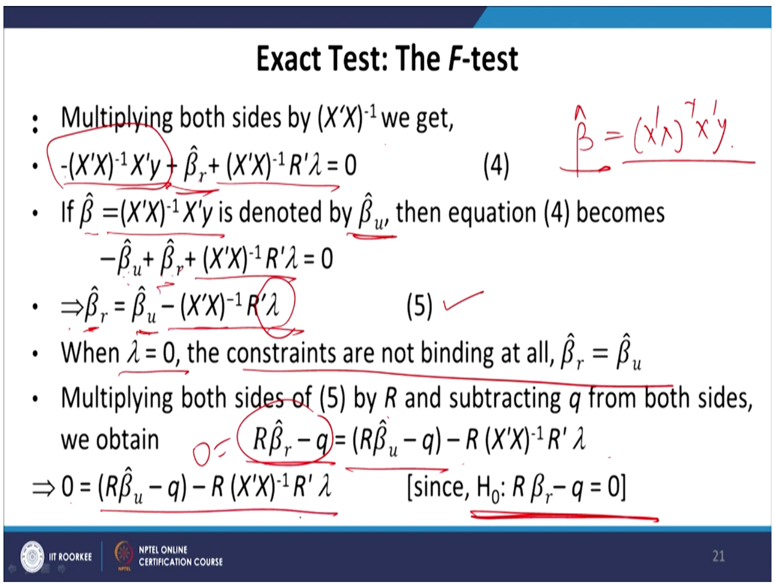
Now, I expand the Lagrange by multiplying these two. So, I have y prime y, then y prime X beta r hat. Similarly, we have beta r hat prime X prime y. And then we have beta r hat prime X prime X beta r hat. So, and then we have the usual thing, plus 2 lambda prime R beta r hat minus q.

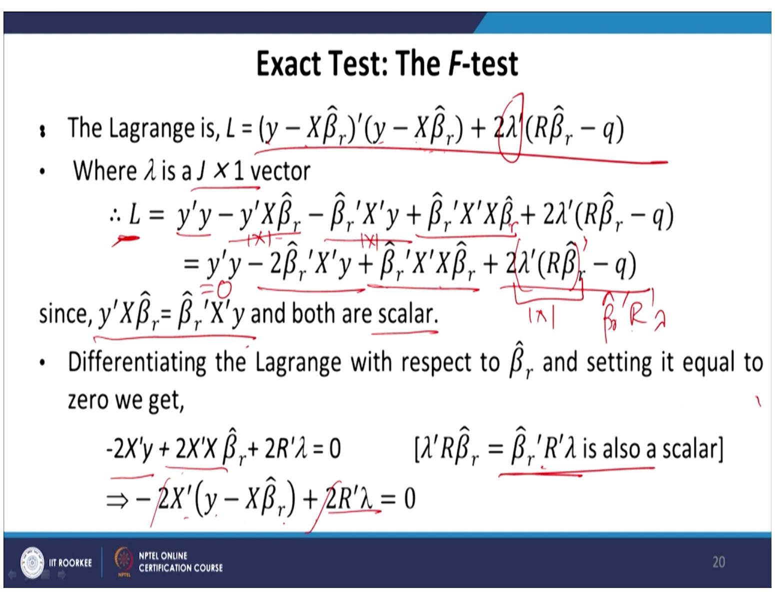
You can see that this is actually a 1 by 1, this is of 1 by 1 dimension, which implies that this is a scalar. Another one is also of 1 by 1 dimension so it is a scalar, as a result of which we can write it like this.

Differentiating the Lagrange with respect to beta r hat, now we are going for minimization. And setting it equal to 0, we get this actually equals to 0, when L is differentiated with respect to beta r hat. So, we have minus 2X prime y plus, from here, we have plus 2X prime X beta r hat. And then we have, again, plus 2R prime lambda.

The reason is that this is actually lambda prime R beta hat r. This is also of 1 by 1 dimension. That it is a scalar. So by taking its transpose, we can write it as beta r hat prime R prime lambda. So that is how we are having here beta hat R prime, R prime lambda. And by differentiating it with respect to beta r hat prime, we have R prime lambda left with us.

(Refer Slide Time: 28:47)





So, multiplying both sides by X prime X inverse, minus one, so we are multiplying by X prime X inverse. So, we are having X prime X inverse X prime y, 2 2 cancels out. So we are having minus X prime X inverse X prime y, then plus X prime X inverse, X beta X prime X beta r hat. So X prime X inverse and X prime X cancels out. I am left with only a beta r hat.

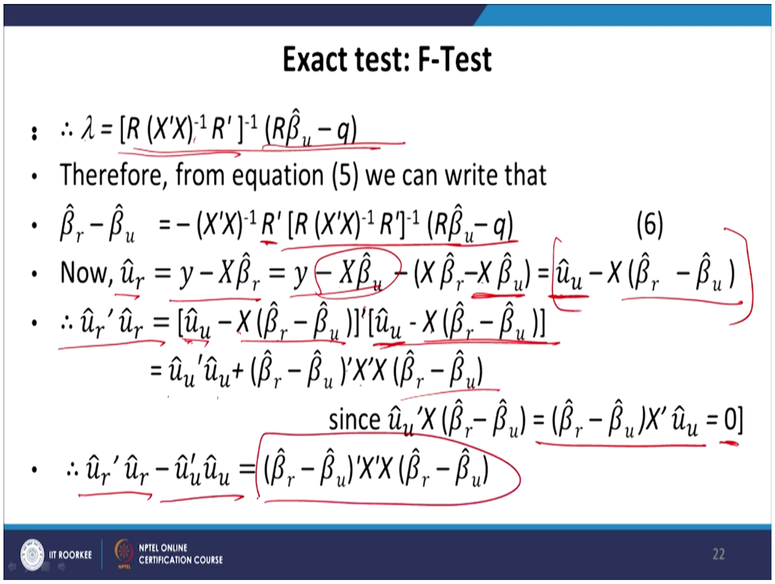
And then, we have X prime X inverse R prime lambda. So X prime X inverse R prime lambda. If beta hat equals X prime X inverse X prime y, so this is my usual expression. Whenever we go for multiple regression, then my, our beta is X prime X inverse X prime y. So, we are calling this X prime X inverse X prime y is equal to the beta hat. And if the beta hat is denoted by beta u hat in order to differentiate it from the restricted parameter estimates *(refer to slide time 28:47).*

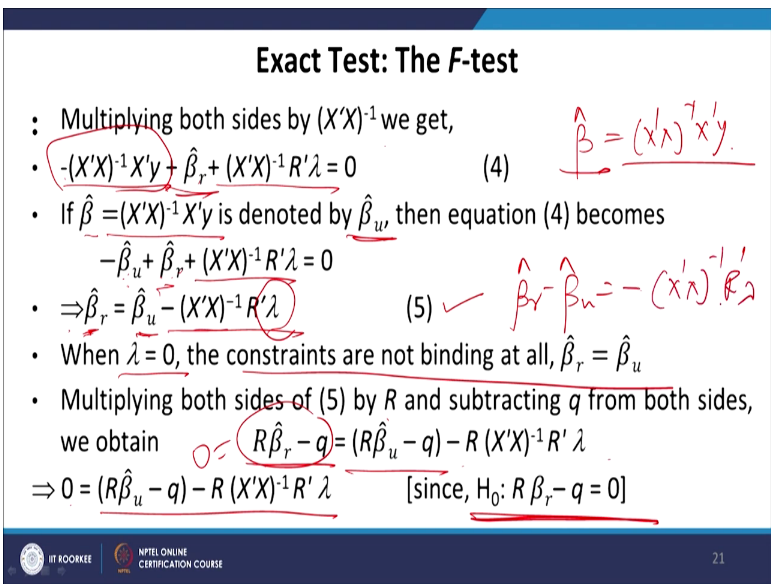
So, we write it, minus beta u hat plus beta r hat, and the same thing, which implies that beta r hat minus, beta r hat is equal to beta u hat minus X prime X inverse R prime lambda. So, this is equation 5, which I am going to use later.

Now, when lambda is equal to 0, this implies that beta r hat is equal to beta u hat. The constraints are not binding at all. Now, multiplying both sides of 5, by R and subtracting q from both sides, what do we obtain? We obtain R beta hat minus q on the left-hand side. Then R beta hat u minus q on the right-hand side.

And then again, I multiply this expression by R. So R X prime X inverse R prime lambda. Since this equals 0 under the null hypothesis, so we are having 0 equals to this expression.

(Refer Slide Time: 30:50)





This tells us that lambda is equal to R X prime X inverse R prime whole inverse R beta hat u minus q. Therefore, from the equation, we can write that, this is my equation 5 which is actually equal to beta r hat minus beta u hat equals to minus X prime X inverse R prime lambda *(refer to slide time 30:50)*.

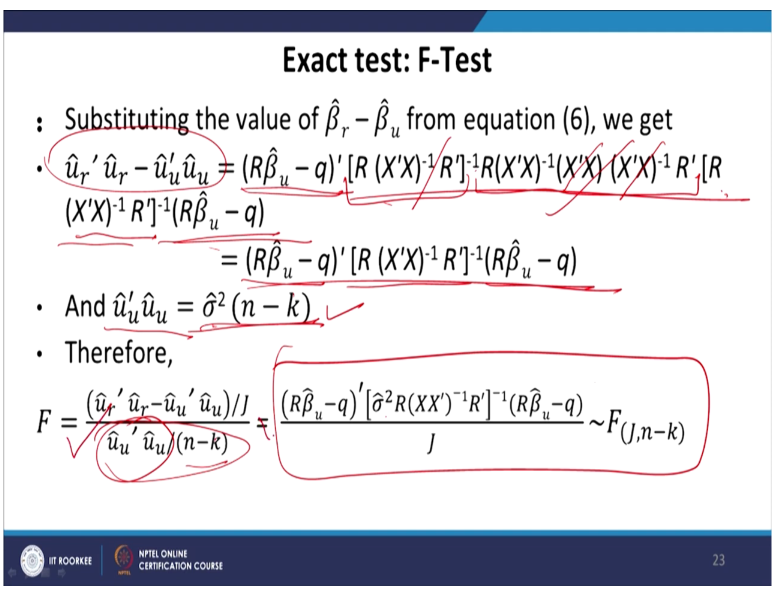
So, what we are doing is that, in place of lambda, we are substituting the value. Now, ur hat is equal to y minus x beta r hat. This is further written as y minus X beta u hat. I incorporate 1 minus X beta u hat. As a result of which, I also incorporate 1 X beta u hat minus plus, so that they cancel out.

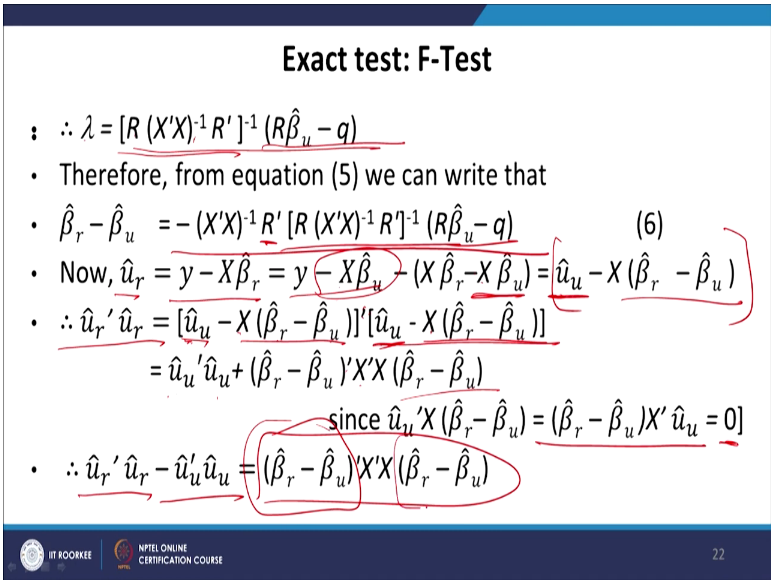
This equals to uu hat y minus X beta u hat. And then, by taking X common, we have beta r hat minus beta u hat. So, ur hat prime ur hat is equaled to this entire thing prime multiplied by again this entire thing.

Now, I expand them. I have uu hat prime multiplied by uu hat. Then uu hat, and then uu hat prime multiplied by X beta r hat minus beta u hat is equaled to 0 because we, this is under one of the CLRM assumptions that, there is independence between the residual terms and the independent variables.

Similarly, this multiplied by this will also be equals to 0, for the same reason. So what I am left with is beta r hat minus beta u hat prime multiplied by X prime, then X and beta r u, beta r hat minus beta u hat, which implies that Ur hat prime Ur hat minus uu hat prime uu hat is equal to this expression.

(Refer Slide Time: 32:56)





Substituting the value of beta r hat minus beta u hat from equation 6. What I get is R beta hat u minus q prime, then the entire thing, again multiplied by R beta hat u minus q.

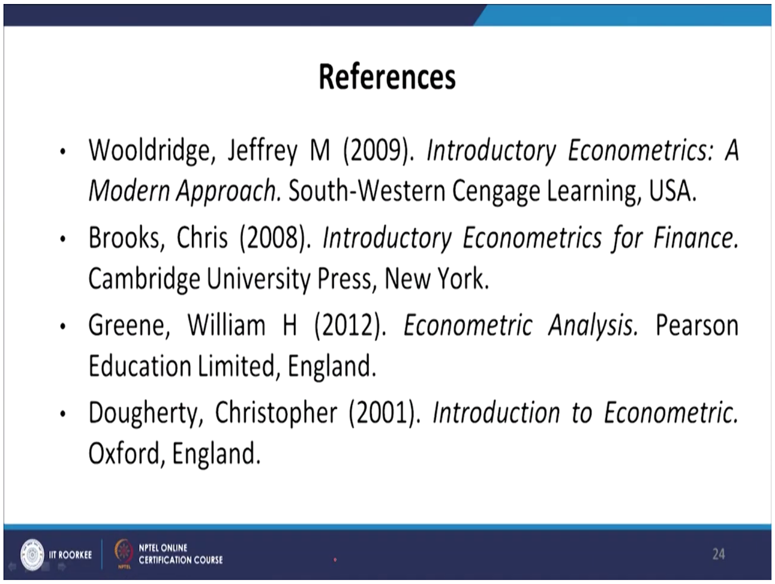
You can see that this expression gets canceled out. First of all, two cancels out. Then RX prime X inverse R prime also gets cancels out with the inverse. So, I am left with R beta hat u minus q prime multiplied by RX prime X inverse R prime whole inverse multiplied by R beta hat u minus q.

And uu hat prime uu hat is equal to sigma hat square multiplied by n minus k. This is something we have already discussed and derived earlier. Therefore, the expression is equal to the expression in the numerator, except for the sigma hat square term. And divided by J.

And, also this expression. Rather, this expression is divided by n minus k. So as earlier I have shown you that n minus k, n minus k cancels out and this sigma hat square basically goes up. So we are left with this expression which follows an F distribution with J and n minus k degrees of freedom.

So this proves that this expression is exactly equivalent to another expression. So, whatever form of F distribution you use, we arrive at basically the same result.

(Refer Slide Time: 34:48)



So this is all about the F distribution, its basic characteristics. I have followed these books in order to come up with the discussion *(refer to slide time 34:48).* In the next module, I will discuss some of the examples of F distribution along with some of its applications. Thank you.