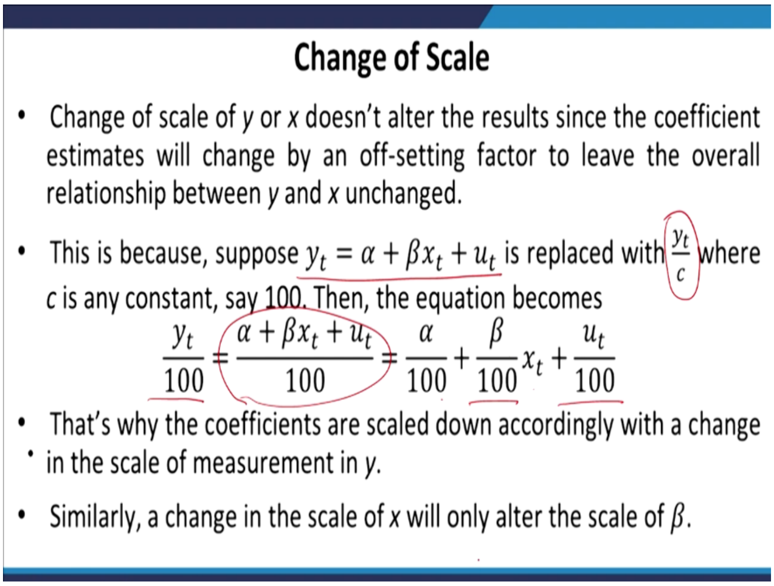
Econometric Modelling
Professor Sujata Kar
Department of Management Studies
Indian Institute of Technology Roorkee
Lecture 07
Simple Regression – II

(Refer Slide Time: 00:38)

Hello everyone and this is module 7 of the course econometric modelling, till module 6 or rather in module 6 we discussed simple regression so that was the first module under part-2, part-2 is an overview of classical linear regression model and I discussed module 6 in the last session where I dealt with simple regression and there, we discussed the method of ordinary least square regression, its mathematical interpretation as well as geometric interpretation. So, specifically, what ordinary least square regression is all about is how it actually fits a line through a set of data and gives us the estimates of the line, that was the purpose.

(Refer Slide Time: 1:37)

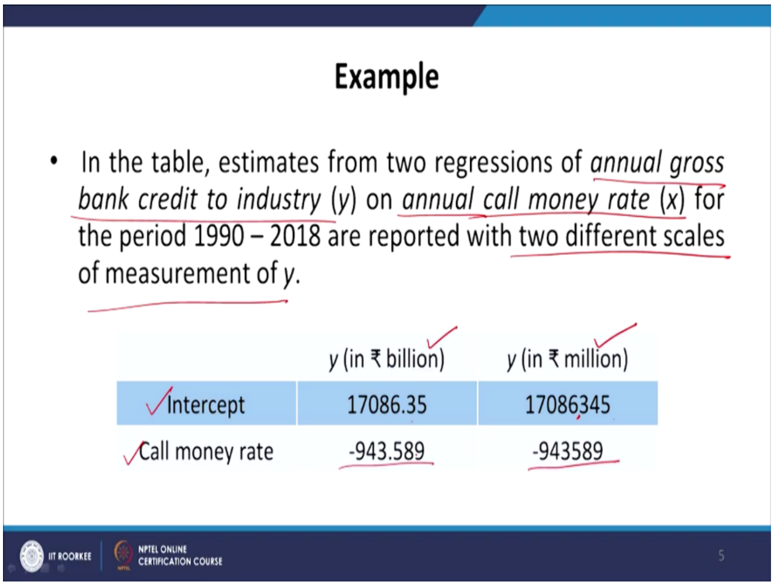


Now in this module again, on simple regression, we are actually going to discuss some of the features or characteristics of simple regression. The first thing that we are going to discuss is a change of scale, change of scale here refers to the fact that when we are changing the scale of measurement of y or x then how it is going to impact the parameter estimates whether it is going to impact the parameter estimates or not.

So, a change of scale of y or x does not actually alter the results since the coefficient estimates will change by an offsetting factor to leave the overall relationship between y and x unchanged. So, this is actually explained below that suppose our original model is a $y_t = \alpha + \beta x_t + u_t$ and we replace y_t with y_t/c where c is any constant so it can be, say 100, it can be 1000 anything or any other number 25, 50 then the equation becomes $\frac{y_t}{100} = \frac{\alpha+\beta x_t+u_t}{100} = \frac{\alpha}{100} + \frac{\beta}{100} x_t + \frac{u_t}{100}$

So, you can see that when y_t is scaled down by an amount say 100, then the parameter estimates are also scaled down accordingly, that is why the coefficients are scaled down accordingly with a change in the scale measurement in y. Similarly, a change in the scale of x will only alter the scale of β, if I replace x_t with $x_t/100$ or $x_t/1000$ then the parameter estimate β will become β by 100 or 1000. So, other than that no change is going to take place.

(Refer Slide Time: 03:39)

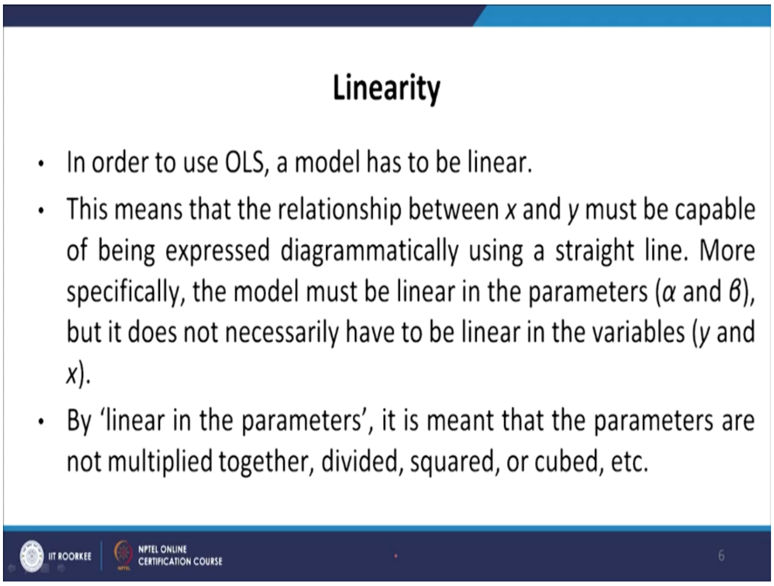


Now we take an example, in the table estimates from the 2 regressions of annual gross bank credit to industry on annual call money rate for the period 1990 to 2018 taken from India, so these are Indian data we have collected data on annual gross bank credit to the industry that is our dependent variable and call money rate is our independent variable, annual call money rate. So, we are trying to find out whether the annual call money rate impacts the annual gross banking, bank credit to industry or not.

So, that is done with respect to 2 different scales of measurement of y. So, y once is measured in terms of billions of rupees and the second one is millions of rupees. So, you can see that both these parameter estimates, intercept as well as call money rate, their values are the same just the decimal points or places have changed.

So, here the decimal place has come 3 points before when we moved rather from million to billion and similarly another one has also changed by, the decimals have moved 3 points to the left, so here we had minus 943589 and now we have minus 943.589. That is the only difference we can observe, so this is an example that shows that with change in scale, nothing changes, or the values remain, the estimates remain the same. Again, only their scale of measurement changes.

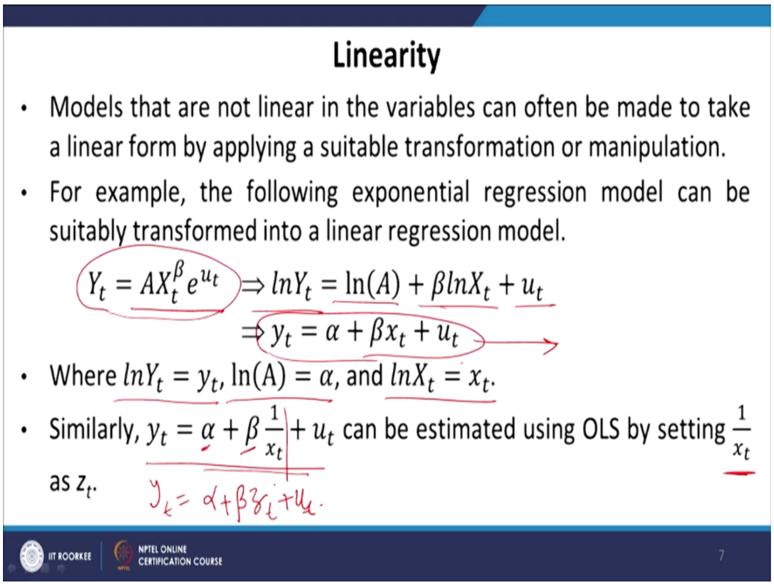
(Refer Slide Time: 5:31)



The next characteristic or feature that we will be talking about is linearity. In order to use OLS, a model has to be linear, this means that the relationship between x and y must be capable of being expressed diagrammatically using a straight line, more specifically the model must be linear in the parameters that are α and β but it does not necessarily have to be linear in the variables that are y and x. So, now again I will show with examples how linearity in x and y or in variables can be resolved in most cases while we certainly do not want non-linearity in the parameters.

So, linearity in parameters actually means that the parameters are not multiplied together divided, squared, or cubed etcetera. So, in case we observe any non-linearity in the parameters that actually cannot be handled or a model having non-linearities in the parameters cannot be estimated using OLS, whereas non-linearities when present in the variables can at time be estimated using OLS.

(Refer Slide Time: 6:58)



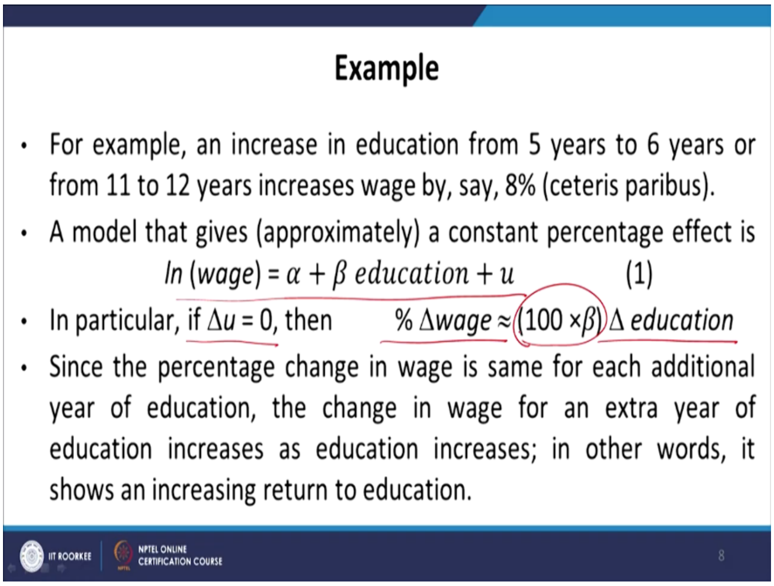
I will now show you examples of that, so models that are not linear in the variables can often be made to take a linear form by applying a suitable transformation or manipulation. For example, in this model, the following exponential regression model can be suitably transformed into a linear regression model.

To begin with, we have $Y_t = AX_t^{\beta} e^{u_t}$ now if I take natural logarithm, then this becomes $lnY_t = \ln ln\,(A) + \beta lnX_t + u_t$ and then if I call $lnY_t = y_t$, $ln(A) = \alpha$, and $lnX_t = x_t$, then we actually from this expression can get our original regression model or generally the way we write it. So, simply $y_t = \alpha + \beta x_t + u_t$

Similarly, if we have a regression model like $y_t = \alpha + \beta\frac{1}{x_t} + u_t$ then this can be estimated using OLS by setting $\frac{1}{x_t}$ as z_t So, in that case, this model will simply look like $y_t = \alpha + \beta z_t + u_t$. So, z_t is a variable that is originally $\frac{1}{x_t}$. Now converting $\frac{1}{x_t}$ or renaming $\frac{1}{x_t}$ as z_t, actually does not impact α or β.

So, my parameter estimates are untouched and that is why this is an estimable model and there is actually no harm in this kind of conversion when we can convert it into linear formats or just change the name of the variables from $\frac{1}{x_t}$ to z_t. Our parameter estimates remain the same and these models are estimable using OLS.

(Refer Slide Time: 9:13)

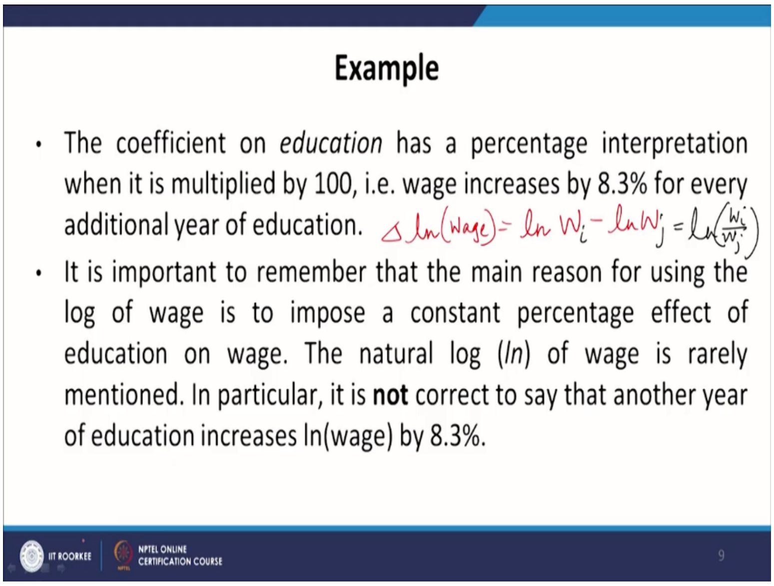


Now we take another example in order to explain the interpretation of β in the context of non-linearity in the variables. So, for example, an increase in education from 5 years to 6 years or from 11 to 12 years increases wage by saying, 8 percent, ceteris paribus which implies that other things holding constant if we have an estimate that with an increase in education by 1 year, wages increase by 8 percent, then how I am going to write it? A model that gives approximately a constant percentage effect is $ln\,(wage) = \alpha + \beta\, education + u$.

In particular, if $\Delta u = 0$ so all other factors contained in u and all of them are held constant then a change in education is going to tell us how what will be the percentage change in wages, and that is given by 100 multiplied by beta i.e $\%\ \Delta wage \approx (100 \times \beta)\ \Delta\ education$. Since the percentage change in wage is the same for each additional year of education, the change in wage for an extra year of education increases as education increases. In other words, it shows an increasing return to education.

So, ideally increasing returns to scale is a concept in economics where it says that when the inputs employed in a production process increase in a certain proportion and output increases more than proportionately then we call it increasing returns to scale. So, here we are observing increasing returns to education that is with an increase in education, the wage is increasing. And this increase is by an amount which is nearly 8 percent.

(Refer Slide Time: 11:36)



Now how are we able to comment on this? So, the coefficient of education has a percentage interpretation when it is multiplied by 100, that is wage increases by 8.3 percent for every additional year of education, this is because when we are measuring wage as the logarithm of wage then $\Delta \ln ln\,(wage) \;= lnw_t - lnw_{t-1}$ or here it is probably better to talk in terms of individuals.

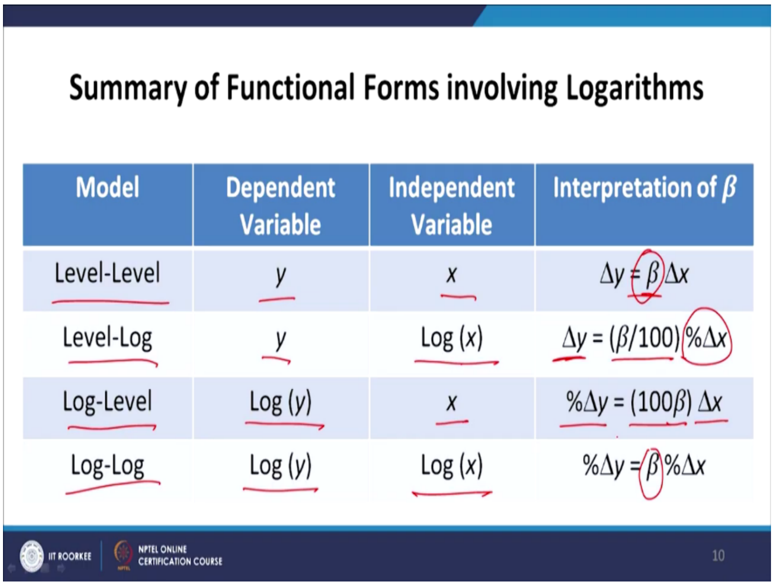
So we are considering individuals having different levels of education, I can write it as an lnw_i and lnw_j, so j has a different wage rate, i has a different wage rate, their education levels are different and accordingly, I am having different wages associated with them.

Now you know, that this $\Delta \ln ln\,(wage) \;= lnw_i - lnw_j$ can be written as $ln(w_i/w_j)$ and this is actually another way of measuring percentage changes or growths like growths are calculated either as $w_i - w_j/w_j$ or it can be $ln(w_y/w_j)$ so the left-hand side that is $\Delta ln(wage)$ is actually measuring the percentage change in the wages earned and this is changing because of a change in the education of individuals.

So, the coefficient of education has a percentage interpretation because of this reason, it is important to remember that the main reason for using the log of wage is to impose a constant percentage effect of education on wage. The natural log is denoted by ln most often of which is rarely mentioned.

So, when we interpret or explain or we talk about the impact of change in education on wages, we actually do not say that with change in education by 1 year, log wage changed by 8 percent. We say that wages changed by 8 percent because that is inherently implied by the way we have framed the model. So, that is why it is not correct to say that another year of education increases log wage by 8.3 percent, rather we say that another year of education increases wage by 8.3 percent.

(Refer Slide Time: 14:22)



Summary of Functional Forms involving Logarithms

Model	Dependent Variable	Independent Variable	Interpretation of β
Level-Level	y	x	$\Delta y = \beta \Delta x$
Level-Log	y	Log (x)	$\Delta y = (\beta/100) \%\Delta x$
Log-Level	Log (y)	x	$\%\Delta y = (100\beta) \Delta x$
Log-Log	Log (y)	Log (x)	$\%\Delta y = \beta \%\Delta x$

Then we come up with a summary of functional forms involving logarithm so since logarithm is in some way, converting or treating one of the variables or both the variables as non-linear or converting non-linearities into linearities, that is why logarithm as a special case is mentioned here, this is also applicable when we measure elasticities, when we measure growths then a lot of time we use logarithm in economics and as a result of which a summary of functional forms involving logarithms are specifically mentioned here.

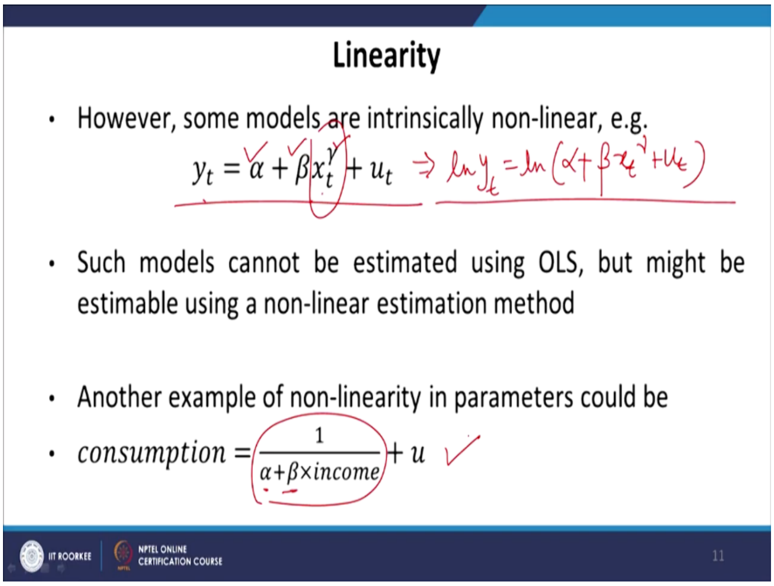
So, the models could be like both of them are in levels so y is in level, x is in level and we have a simple interpretation of β, x changes by certain unit, y changes by certain unit, or the change in y is given by the parameter estimate β.

The model could be a level log that is when y is in level and independent variable x is measured in terms of the logarithm, then the interpretation of β would be, you can see that, since x is in logarithm, so $\%\Delta x$ measures percentage change in x straight away while Δy is simply a change in y and that is given by $\beta/100$.

Log level when y is measured in terms of logarithm x is measured in its level, the example that we had just taken, the example of the impact of education on wages, then we have $\%\Delta y$ as measured by 100β and y and that is given by an Δx.

So, Δx that is a change in x changes y by β multiplied by 100 times, and finally, we can have log-log, so y is measured in terms of the logarithm, x is also measured in terms of logarithm and then again β has the usual interpretation, that $\%\Delta x$ leads to β percentage change in y. So, this is the summary of functional forms involving logarithms.

(Refer Slide Time: 16:45)



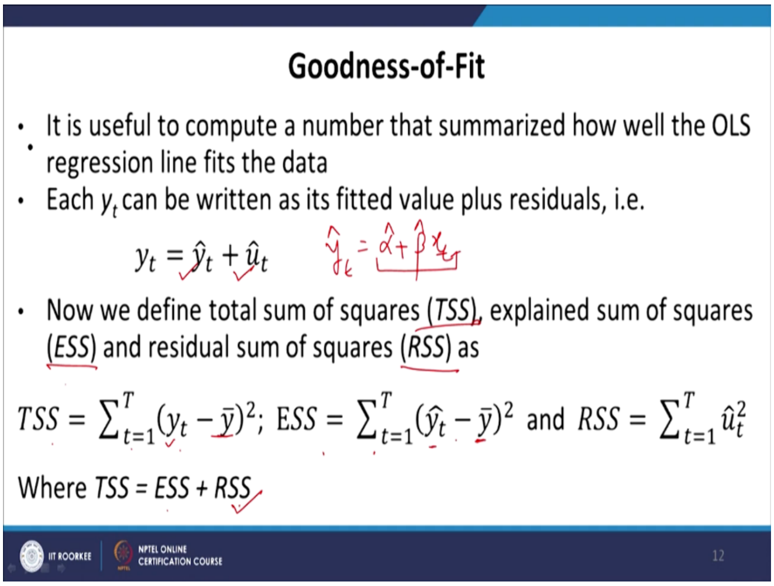
However, some models are intrinsically non-linear, for example, this is a model which is intrinsically non-linear so here, first of all, we note that there are 3 parameters that need to be estimated, parameter α, parameter β , and parameter γ.

We have only 2 variables y_t and x_t, so first of all unless and until I know the value of γ, I cannot come up with the variable x_t, raise to the power γ, we can linearize it partially by writing it as the logarithm of y_t Of course, it should be also supported by a theoretical argument under economics or whatever field from where we are taking the examples or fitting the model in.

So this can be (refer slide time 17:35) and this actually does not solve any problem and as a result of which, this remains inestimable or you know an intrinsically non-linear model. That is why we say that such models cannot be estimated using OLS but might be estimable using a non-linear estimation method.

- Another example of non-linearity in parameters could be this example where consumption is expressed as a function of 1 upon alpha plus beta multiplied by income plus u, i.e. $consumption = \frac{1}{\alpha+\beta\times income} + u$ so one cannot come up with a variable involving income which is the independent variable unless and until I know the value of α and β. So, this is another intrinsically non-linear model which is not estimable using methods like OLS where it is required to have a linear relationship between the parameters. Certain non-linearities in the variables can be converted into linearity.

(Refer Slide Time: 18:54)



The next thing that we talk about is, the goodness of fit, it is useful to compute a number that summarizes how well the OLS regression line fits the data. So far we have talked about obtaining values of α and β fitting a line. There is a lot more to discuss how these parameters are, but the basic study starting point is that how the model fits is, how good the line fit is?

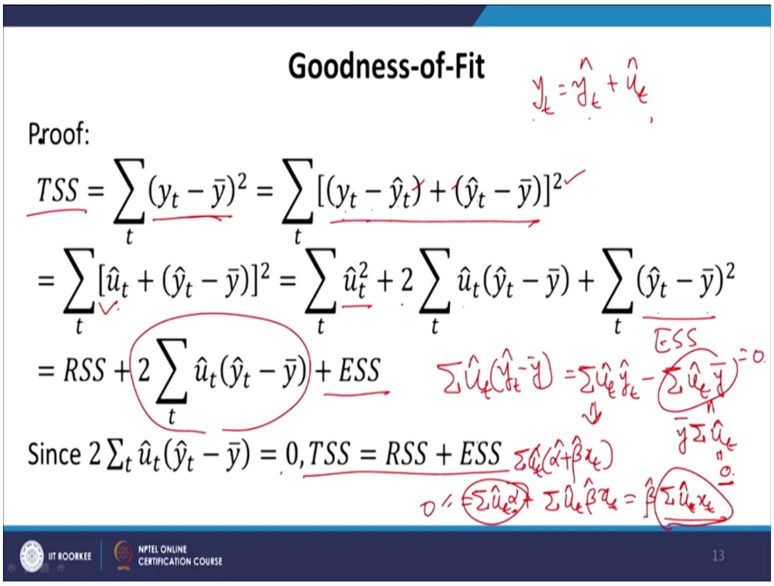
So, each yt can be written as a fitted value plus the residuals. So, I had probably explained it in the previous module that (refer slide time 19:35) and what remains is the residuals which is the ut hat, residuals are separate from error terms, I repeat.

Now we define total sum up squares denoted by TSS, explained sum of squares denoted by ESS, and residual sum of squares denoted by RSS as $TSS = \sum_{t=1}^{T}\left(y_t - \bar{y}\right)^2$; $ESS = \sum_{t=1}^{T}\left(\hat{y}_t - \bar{y}\right)^2$ and $RSS = \sum_{t=1}^{T}\hat{u}_t^2$

Where $TSS = ESS + RSS$

This is something we are going to prove in the next slide but before that, I need to tell you that, you note TSS is the total sum of squares, now what is the sum? The sum is the deviation of the actual observations from the mean value, the mean of the series y. Explained sum of squares again, what is the sum? The sum is actually the deviation of the estimated values from the mean values of the series again, the yt series. And RSS or the residual sum of squares simply considers the squares of the residuals and sums them up.

(Refer Slide Time: 21:25)



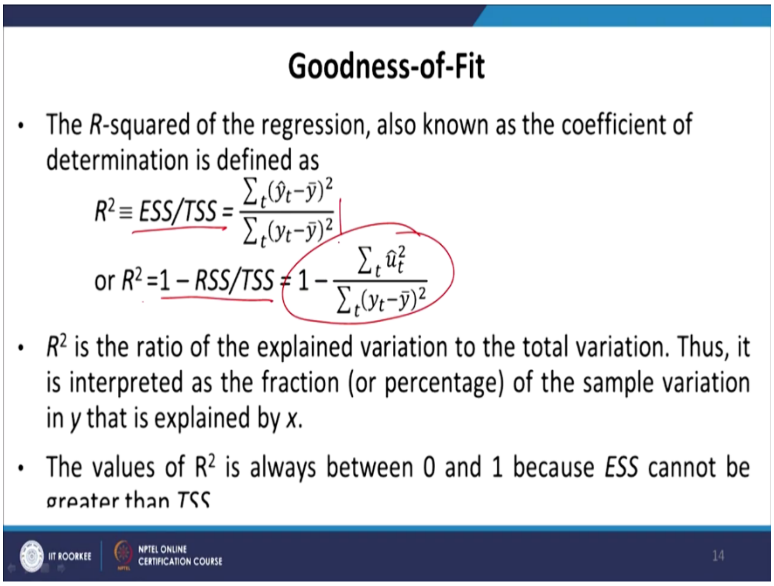
(Refer slide time 21:25) Now I prove that TSS is equal to ESS plus RSS is actually pretty simple, what we do is, first of all, we define TSS, we have already defined TSS, so TSS (refer slide time 21:25)

So, this is simple and this expression is actually expanded here by incorporating (refer slide time 21:54 - 22:35) So, this is simply RSS, the way we have defined RSS, this is simply ESS, the way we have defined ESS in the previous slide. So, I am left with this expression and this expression is equal to 0, why? Because I take this summation, (refer slide time 22:53 - 24:07)

And this term is 0, that is because of one of the assumptions of classical linear regression that is something which we are going to take up in the next module.

For the time being, I tell you that this multiplication is 0 because we assume independence between the residual and the independent variables or the error terms and the independent variables. So, as a result of which, this entire expression becomes 0 and consequently, I have TSS equals RSS plus ESS, it has been proved.

(Refer Slide Time: 24:39)



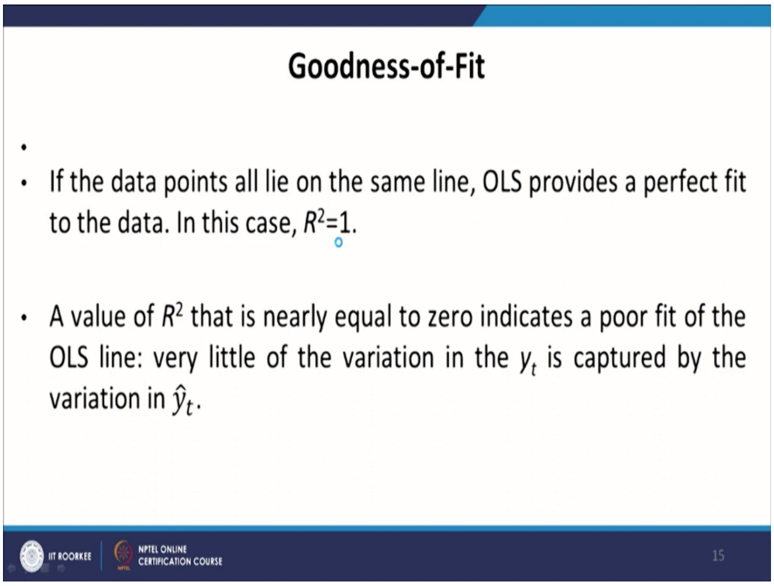
Now we are going to define the goodness of fit measure, so R-squared of the regression also known as the coefficient of determination is actually a goodness of fit measure that is, it measures how good the model fit has been. Now R-square is defined as ESS upon TSS explained sum of squares divided by the total sum of squares and the expressions are mentioned here.

So, R-square actually measures that what percentage of the total sum of square is actually explained by the fitted line, alternatively R-square can be written as $R^2 = 1 - RSS/TSS = 1 - \frac{\sum_t \hat{u}_t^2}{\sum_t (y_t - \bar{y})^2}$, so this is the same thing, residuals sum of square divided by TSS gives us what percentage of the model is not explained by the line and one minus that gives us what explained by the line and this is the expression for R square.

So these are two alternative expressions, R square is the ratio of the explained variation to the total variation. Thus, it is interpreted as the fraction or percentage of the sample variation in y that is explained by x, the values of R square are always between 0 and 1 because ESS cannot be greater than TSS. So, the sum of the square cannot be greater than the total sum of square because just now we have proved that explained sum of square plus RSS, residual sum of square, they add up together to form TSS, and all of them are positive numbers because all of them are squared sums.

So, once I take the square, then anything negative even becomes positive and when we sum them up, they cannot be negative numbers. So, individually all of them are positive numbers, 2 positive numbers sum up to make TSS, as a result of which TSS, RSS, ESS are all positive and that is how ESS upon TSS can never be less than 0. That is how the R-square value always lies between 0 and 1, it is a fraction.

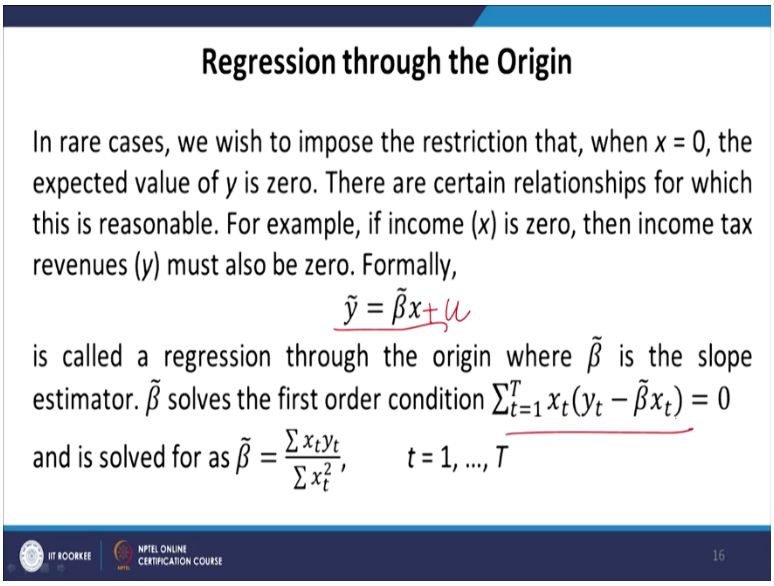
(Refer Slide Time: 26:53)



If the data points all lie on the same line, OLS provides a perfect fit to the data, in this case R^2=1, if the line fit is perfect that is any deviation from y to its mean value is completely explained by the fitted line, so $\left(y_t - \bar{y}\right)^2$ and summed up is equal to $(\hat{y}_t - \bar{y})^2$ and summed up, they are the same, then r-square becomes 1.

A value of r-square that is nearly equal to 0 that is the other opposite extreme case, then it indicates a poor fit of the OLS line, very little of the variation in the yt is captured by the variation in $\hat{y}_t$. So, basically, the difference between y_t and $\bar{y}$ is hardly explained by the very difference between $\hat{y}_t$ and $\bar{y}$. Then we have a case where r-square is actually close to 0.

(Refer Slide Time: 28:07)

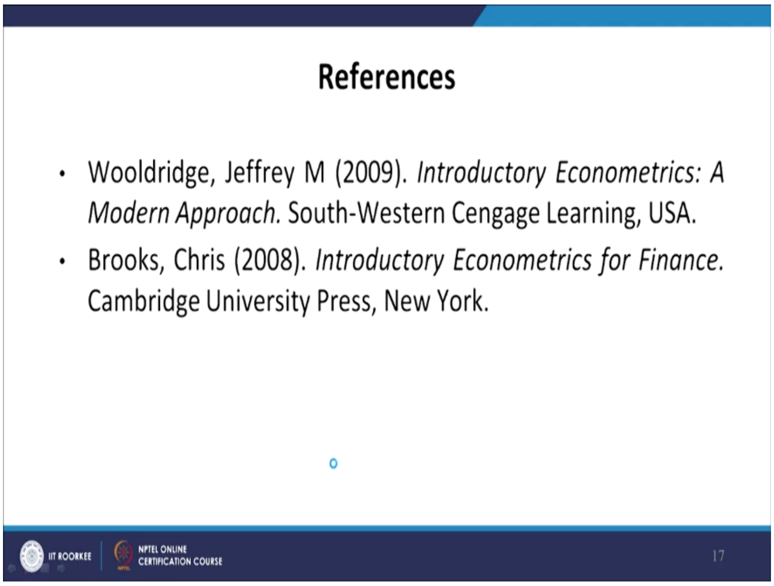


So, now we are actually talking about a special case where the regression is through the origin. In this module we are discussing certain characteristics of simple regression, we started with a change in the scale of measurement, then we talked about linearity, then we talked about the goodness of fit measures and the final thing that we are discussing is a regression through the origin.

So, in rare cases, we wish to impose the restriction that when x equals to 0, the expected value of y is 0. There are certain relationships for which this is reasonable, for example, if income (x) is 0 then income tax revenues (y) must also be 0. So, in that case, we cannot have or we need not have a positive or negative or inclusion of an interceptor constant term. So, we simply write it as $\tilde{y} = \tilde{\beta}x + u$ and this is the deterministic component. For a random component, we can add certainly u here.

Now this is called a regression through the origin where $\tilde{\beta}$ is the slope estimator, $\tilde{\beta}$ solves the first-order condition $\sum_{t=1}^{T} x_t(y_t - \tilde{\beta}x_t) = 0$, so when we had two parameters to be estimated then we had 2 first-order conditions and when we have only 1 parameter to be estimated, we have only 1 first-order condition. And from this first-order condition $\tilde{\beta} = \frac{\Sigma x_t y_t}{\Sigma x_t^2}$, $t = 1, \ldots, T$So, this is a regression through the origin.

(Refer Slide Time: 30:09)



And these are the books that I have followed in order to present the content of module 7. Thank you!