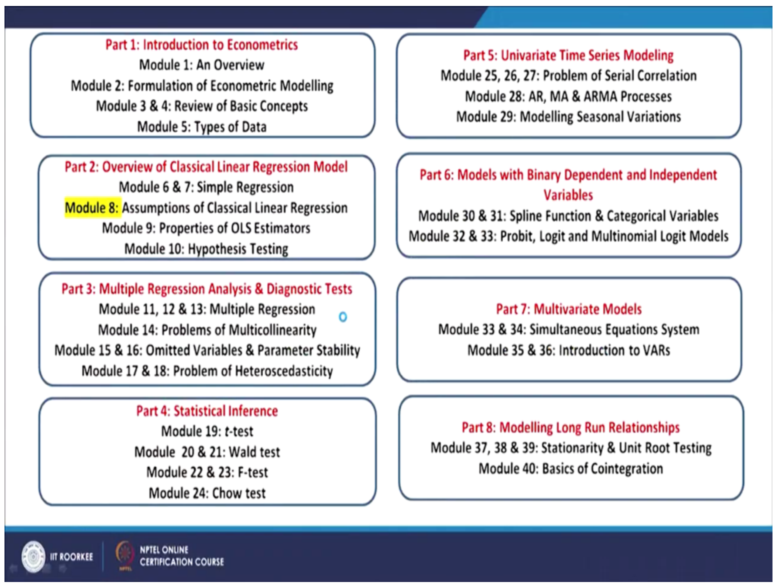
Econometric Modelling
Professor Sujata Kar
Department of Management Studies
Indian Institute of Technology Roorkee
Lecture 08
Assumptions of Classical Linear Regression

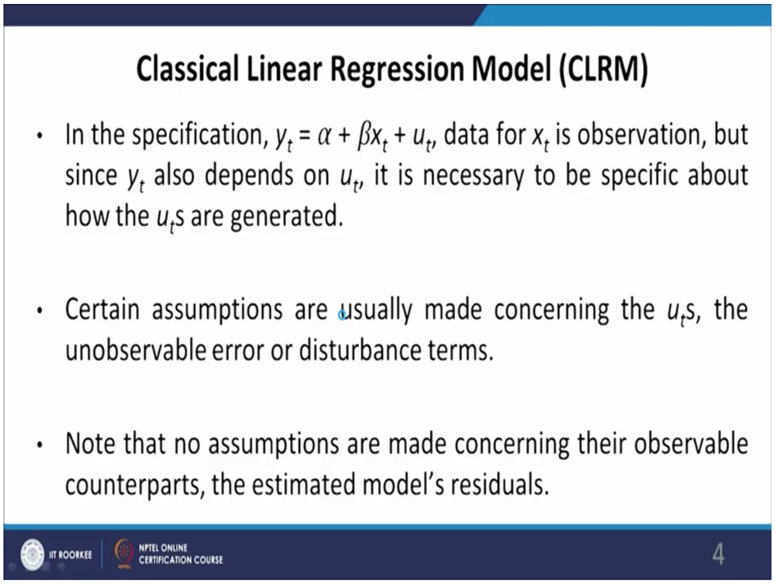
Hello, and this is module 8 of the course on Econometric Modelling.

(Refer Slide Time: 00:32)



Module 8 presents the assumptions of classical linear regression. So once we have talked about the basic structure of simple regression, the next thing is the assumptions of classical linear regression. They are a very crucial part of any regression analysis. And basically, a lot of things, or whether how good a model is, that to a large extent depend on these assumptions of classical linear regression.

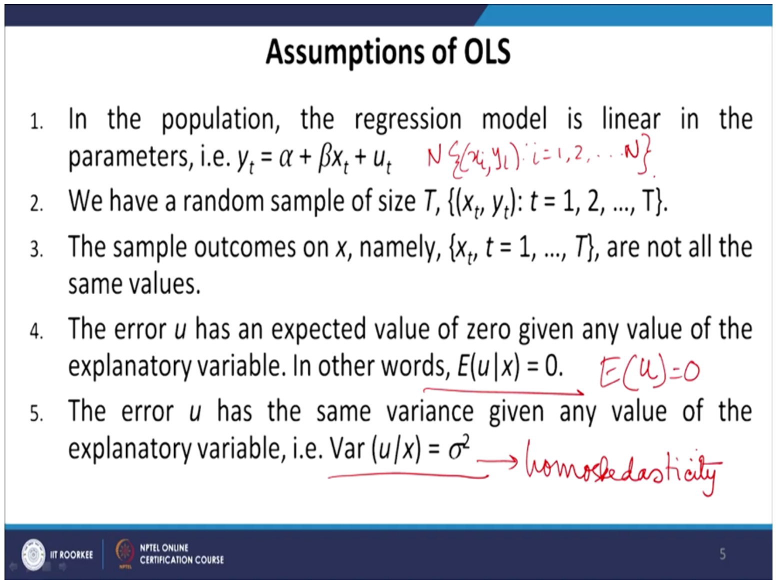
(Refer Slide Time: 01:02)



So first of all we define the classical linear regression model. In the specification $y_t = \alpha + \beta x_t + u_t$, data for x_t is observed. That is they are observations. But since y_t also depends on u_t it is necessary to be specific about how $u_t s$ are generated. So most of the assumptions are actually made concerning the $u_t s$, the unobserved error, or the disturbance terms. Since x_ts' are observed that is why we do not need many assumptions about them. Or for the time being, no assumptions are made concerning their observable counterparts, the estimated model's residuals.

Once we obtain the sample from the population then we obtain the residuals. So the assumptions are all about the error terms primarily and they pertain to the error terms. Whether the assumptions are fulfilled that is certainly checked by considering the residuals or with the residuals but assumptions are made about the error terms only and not about the residuals. On the other hand, very few assumptions are made about the observed part that is x_t. Most assumptions revolve around the unobserved part.

(Refer Slide Time: 02:37)



So the assumptions of OLS are primarily five. First of all, in the population, the regression model is linear in the parameters. So linearity, what it implies, why do we need it, has already been explained. We must have a broad formula or framework of the model as $y_t = \alpha + \beta x_t + u_t$, which basically states that the relationship between the variables in terms of the parameters is linear.

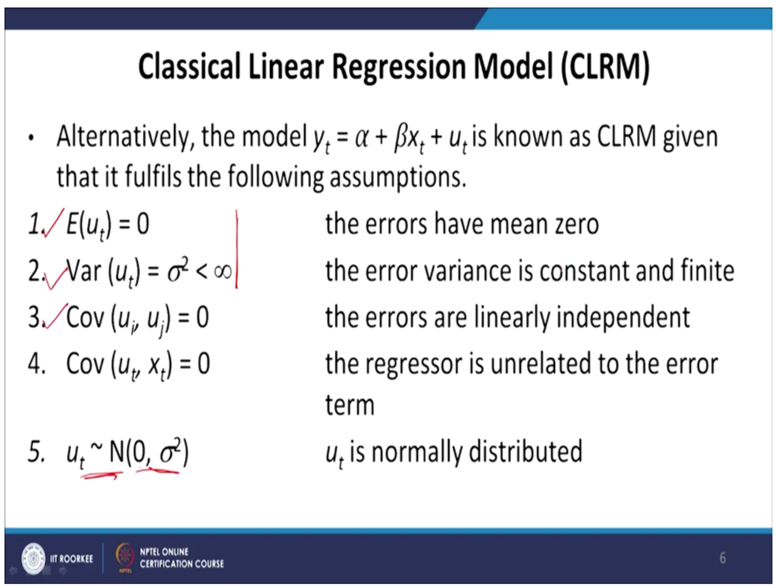
We have a random sample of size T. So the sample should be a random sample, where the sample consists of the observations on two variables $\{(x_t, y_t): t = 1, 2, \ldots, T\}$. So here we are considering time series. We can similarly have alternatively t replaced with i. So we can have a random sample of size N where the observations are (refer slide time 3:50). Then we are dealing with cross-section data.

The sample outcomes on x namely, $\{x_t, t = 1, \ldots, T\}$, are not all the same values.

The error u has an expected value of 0 given any value of the explanatory variable. In other words expected value of u given x is equals to 0 $E(u|x) = 0$. So, so far we have talked about in parts that the expected value of u is equal to 0. Now we are saying that the expected value of u conditional upon the values of x is 0. We are going to deal with this assumption which is a very crucial one, at length, in the upcoming slides.

The error u has the same variance given any value of the explanatory variable. So Var $(u|x) = \sigma^2$. This is also another very important assumption. The error u has the same variance conditional upon the values of x. And this is called the assumption of homoskedasticity. So in the later slide when we deal with this assumption then that will be directly referred to as the assumption of homoskedasticity.

(Refer Slide Time: 05:16)

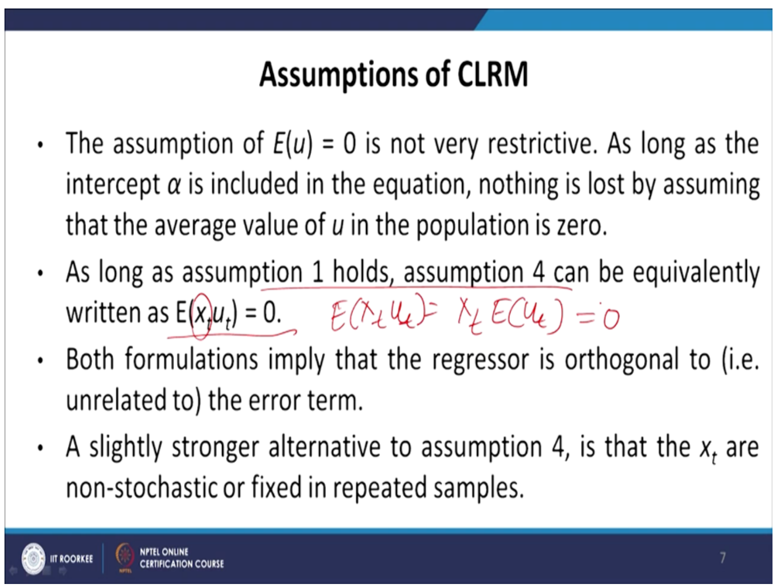


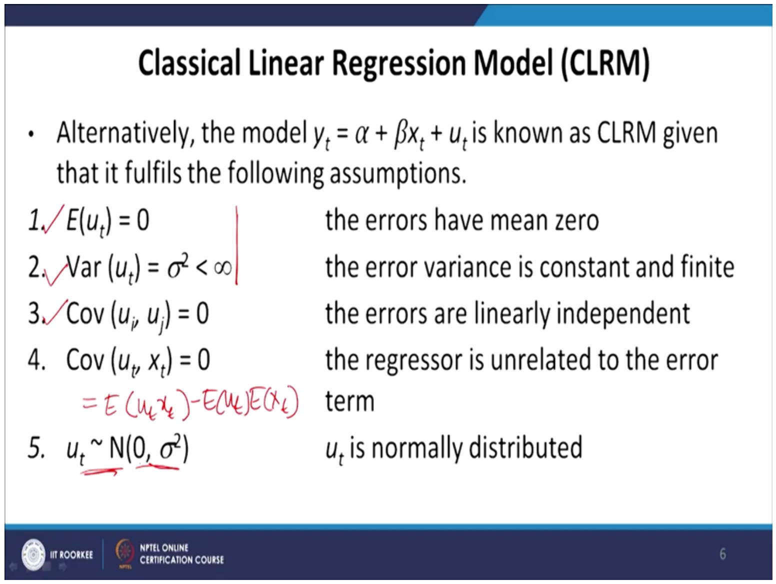
So classical linear regression model, alternatively $y_t = \alpha + \beta x_t + u_t$ is known as the classical linear regression model or CLRM given that it fulfills the following assumptions. The first assumption is expected value of ut given is equal to 0. The variance of ut is equal to sigma square. Covariance between ui uj equals 0, so that is the different observations of the error term are actually independent of each other. If you remember we had defined in the very beginning that independence between two variables implies that their covariances are 0. (refer slide time 5:30)

Now if covariance between $u_t u_j$ is equal to 0 here implies that the observations that the random u or error term has, are independent of each other. Covariance between ut and x_t is equal to 0 which is the regressor is unrelated to the error term. Again the error term and the independent variables, are basically independent of each other. And finally, ut follows a normal distribution with 0 mean and sigma square as the variance. (refer slide time 6:07)

So this, of course, is partly referred to by the first two assumptions, that it has mean 0 and variance sigma square. The only thing that is being added here is that u_t is normally distributed. So, this is the same set of observations or assumptions that are now written in a different way and this entire thing together is known as the classical linear regression model.

(Refer Slide Time: 07:09)



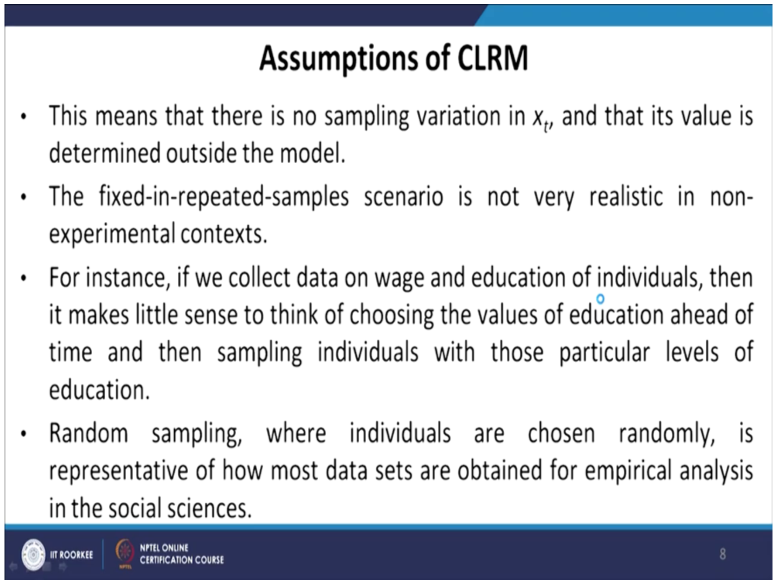


Now the assumption of $E(u) = 0$ is actually not very restrictive as long as the intercept α is included in the equation. Nothing is lost by assuming that the average value of u in the population is 0 because even if it is not 0 it can be rendered 0 by modifying the model in such a way that α will take care of the non-zero mean of the error term.

As long as, assumption 1 holds, assumption 4 can be equivalently written as $E(x_t u_t) = 0$. So assumption 1 is $E(u) = 0$. Now you can see that assumption 4 (refer slide time 8:00) So if you remember an alternative formula for covariance was (refer slide time 8:06). Now that is why we are saying that the longer the first assumption holds that $E(u) = 0$, then this assumption can be written as $E(x_t u_t) = 0$. Both formulations imply that the regressor is orthogonal, which is unrelated to the error term.

A slightly stronger alternative to assumption 4 is that x_t are non-stochastic or fixed in repeated samples. So this actually may also imply that expected value, if x_t is fixed in the repeated sample then we can consider x_t as non-random. And the moment it becomes non-random we can always write (refer slide time 9:24), and as a result of which this becomes 0. So, but this is a slightly stronger alternative to assumption 4. This is a stronger alternative because we are assuming away the randomness that could be there in x_t.

(Refer Slide Time: 9:49)



This means that there is no sampling variation in x_t, and that its value is determined outside the model. The fixed-in-repeated-samples scenario is not very realistic in a non-experimental context. For example, if we collect data on the wage and education of individuals then it makes little sense to think of choosing the values of education ahead of time and then sampling individuals with those particular levels of education. Alternatively, suppose I collect data from 100 individuals and on y_t and x_t that is wage and education.

Then I go to another set of 100 individuals for collecting data on wage and education. But if I hold x_t constant then this means that I must choose those individuals having the same amount or the same number of years of education. So that education remains constant, only my y_t keeps on changing with the change in the individuals.

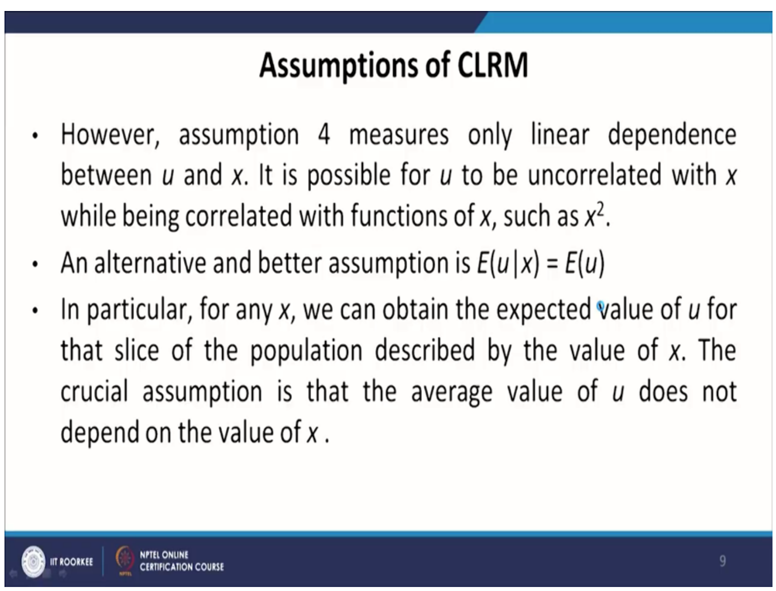
So, we are holding x_t or independent variable constant in repeated samples which is unrealistic, and as a result of which, this is, we call a slightly stronger assumption. Random sampling where individuals are chosen randomly is representative of how most datasets are obtained for empirical analysis in the Social Sciences. And if we assume repeated, x fixed in repeated samples then this assumption of random sampling is lost in the case of or in the context of non-experimental data.

So that was a special case. But in general, what we have obtained is that it the expected value of u_t multiplied by x_t is actually 0. Alternatively, the covariance between u_t and x_t is 0. And one of the assumptions of CLRM is that the expected value of u_t is equal to 0.

Now if you remember in the previous module while discussing adjusted R-squared value I had assumed that (refer slide time 12:21). And then I have proved that why that is 0 while I made an assumption that (refer slide time 12:31). So, this is actually very similar to this assumption but this is in the context of population.

And if we work with the sample residual then this assumption, if applicable on the sample residuals as well, then would mean the same thing, (refer slide time 12:53), which is alternatively the covariance between the residual and the independent variable is actually 0.

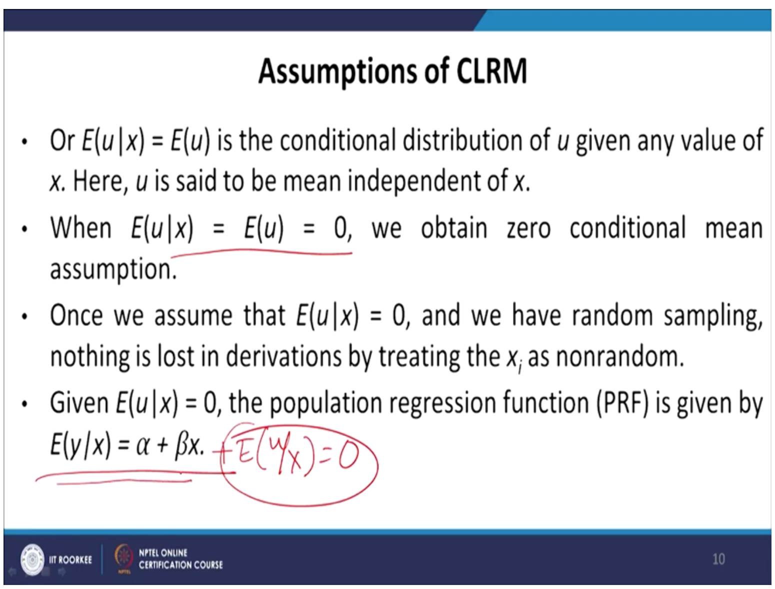
(Refer Slide Time: 13:02)



- Anyway, however assumption 4 measures only linear dependence between u and x. It is possible for u to be uncorrelated with x while being correlated with functions of x such as x^2. So an alternative and better assumption is $E(u|x) = E(u)$. So, you can see that when we assume that covariance between u and x actually equals 0 that is specifically having an expression would imply that the expected value of u_t multiplied by x_t equals to 0. But the problem is that if u_t can be correlated with any other functional form of x, for example, u_t can be related to x^2, x^3, 1/x, and so on.

So, if we can assume the errors are actually unrelated to any functional form of x then we would be writing it as the expected value of u given x, that is any value of x, any functional form of x, we are having that equal to the expected value of u. So in particular for any x we can obtain the expected value of u for that slice of the population described by the value of x. The crucial assumption is that the average value of u does not depend on the value of x. So u is actually mean independent of x.

(Refer Slide Time: 14:44)

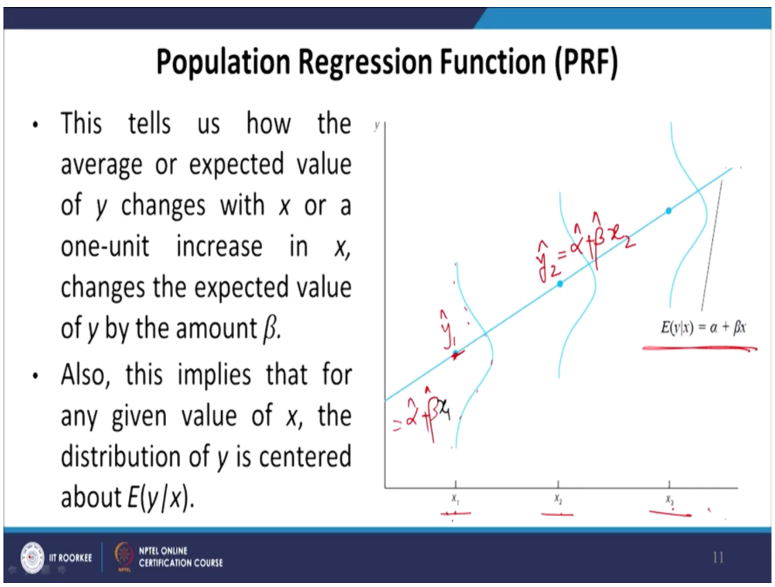


So alternatively $E(u|x) = E(u)$ is the conditional distribution of u given any value of x. Here u is said to be mean independent of x. When $E(u|x) = E(u) = 0$ we obtain 0 conditional mean assumption. Once we assume that $E(u|x) = 0$ and we have random sampling nothing is lost in derivation by treating the x_is as non-random.

So even if I go for a fixed regressor in the repeated sample we do not lose much because we have already made a sufficiently strong assumption that u_ts are actually independent of any value of x. So you keep x fixed in the repeated samples, you keep on changing x in repeated samples, which is in no way going to impact the values of u_t or rather the expected values of u_t.

Given $E(u|x) = 0$ the population regression function (PRF) is given by $E(y|x) = \alpha + \beta x$, because we have this additional term $E(u|x) = 0$. This portion is 0 and that is why this is what is our population regression function.

(Refer Slide Time: 16:23)

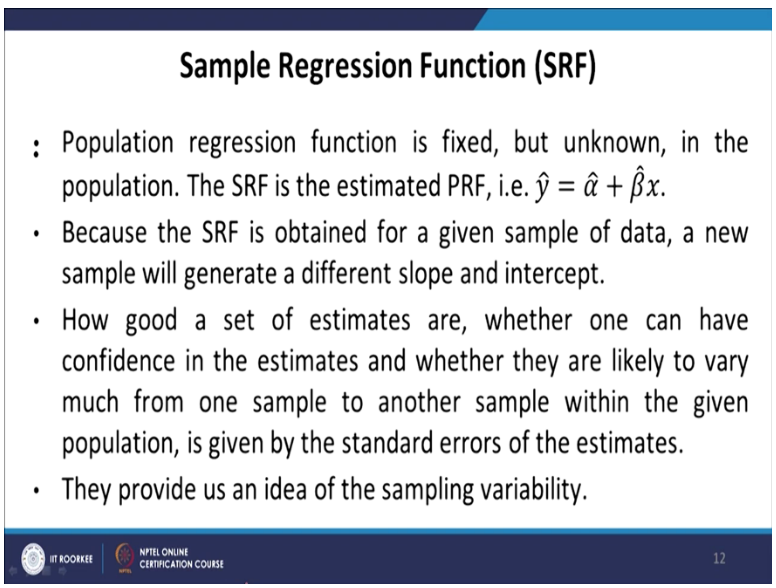


Now we draw this population regression function. So on the right-hand side, you can see a diagram where for various values of x1, x2, x3 we have a certain value of y. Now you can see that if in repeated samples, if I collect samples repeatedly and x values are fixed then I can have for a fixed value of x1, so x, we have taken 3 alternative values of x; x1, x2, x3. And for the individual value of x, we have different observations of y.

And this is my fitted line which is the population; this is the population regression function. In the population, this line would fit the data given as (refer slide time 17:12). So, this is the line that is passing through and these are the points of $\hat{y}_t$. (Refer slide time 17:22)

Now this tells us how the average or expected value of y changes with x, or a one-unit increase in x changes the expected value of y by the amount β. Also, this implies that for any given value of x the distribution of y is centered about $E(y|x)$. So this is the distribution of y. For any given value of x, they are centered around the distribution of y. The population regression function is at the center and the distribution of y is centered around or about the population regression function.

(Refer Slide Time: 18:34)



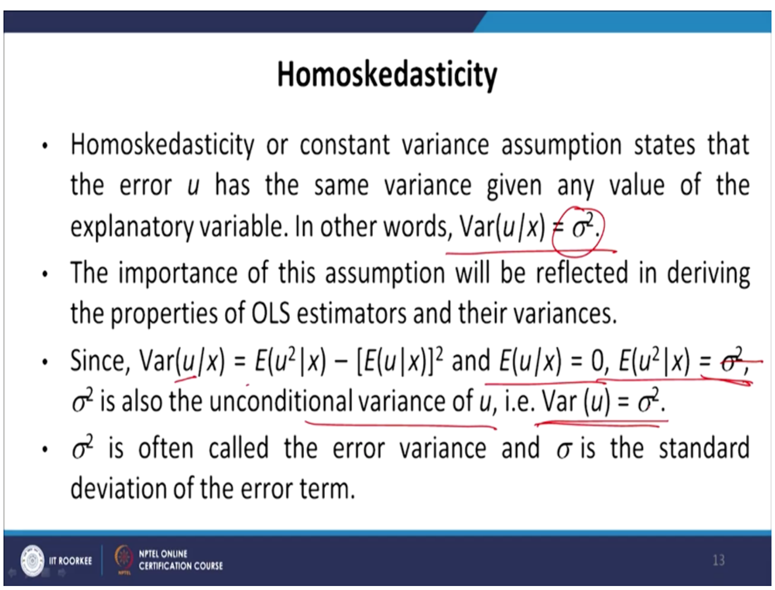
So, its counterpart for the sample is the sample regression function or SRF. Population regression function is fixed but unknown in the population, and that is why the SRF is estimated as PRF. So, with samples, we try to estimate the population regression function and which is denoted by the sample regression function which is $\hat{y}_t = \hat{\alpha} + \hat{\beta}x$.

Now because the SRF is obtained for a given sample of data, our new sample will generate a different slope and intercept. Every time we change the sample our $\hat{\alpha}$ *and* $\hat{\beta}$ values are expected to change. How good a set of estimates are, whether one can have confidence in the estimates and whether they are likely to vary much from one sample to another sample within the given population is given by the standard errors of the estimates.

So, we have a set of estimates given by different samples. So as samples changed my $\hat{\alpha}$ *and* $\hat{\beta}$ values changed. That is how if I draw 100 samples, I will be getting 100 estimates of $\hat{\alpha}$ *and* $\hat{\beta}$. And these 100 estimates are never going to be the same because my sample has changed or whatever be the variations between them that are measured by the sample variance, or variance of the sample estimates or their standard errors.

So, variance, square root of the variance is generally standard deviation. But when it comes to the estimators the square root of the variance is known as standard error. So they provide us an idea of the sampling variability, how much my parameter estimates are varying due to variations in the sample.

(Refer Slide Time: 20:52)

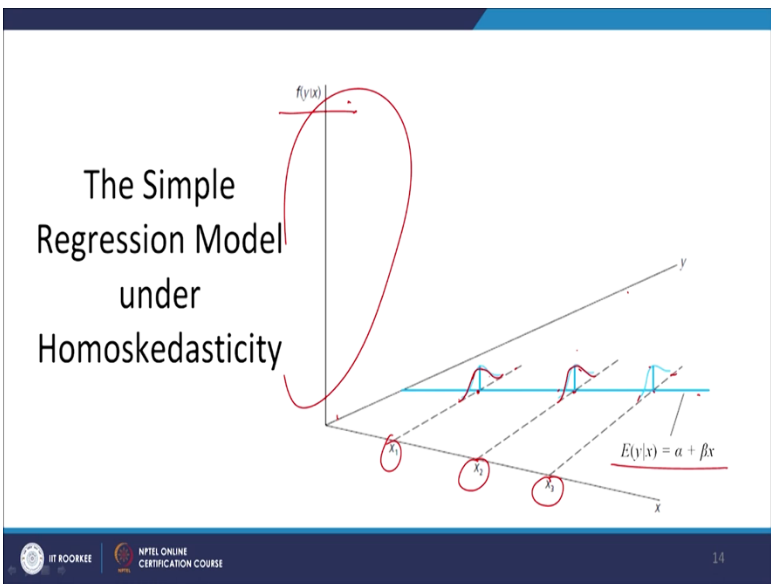


So next we talk about the last assumption of homoskedasticity. I mentioned that homoskedasticity implies constant variance of the error term. Homoskedasticity or constant variance assumption states that error u has the same variance given any value of the explanatory variable. In other words, $Var(u|x) = \sigma^2$. You can see that the sigma square is a constant number. It does not have any subscripts like t, i, or no variable is associated with it. So, it is a constant number, and as a result of which this refers to the assumption of homoskedasticity which implies that the error has a constant variance.

The importance of this assumption will be reflected in deriving the properties of OLS estimators and their variances. $Var(u|x) = E(u^2|x) - [E(u|x)]^2$ and $E(u|x) = 0$, $E(u^2|x) = \sigma^2$, σ^2 is also the unconditional variance of u.

$Var(u) = \sigma^2$ which implies that here we are deriving the variance of u conditional upon the values of x, and we are getting σ^2. So, this is the conditional variance of u given x. And this is the unconditional variance of u given x. Both are σ^2. σ^2 is often called error variance and sigma is the standard deviation of the error term.

(Refer Slide Time: 22:31)

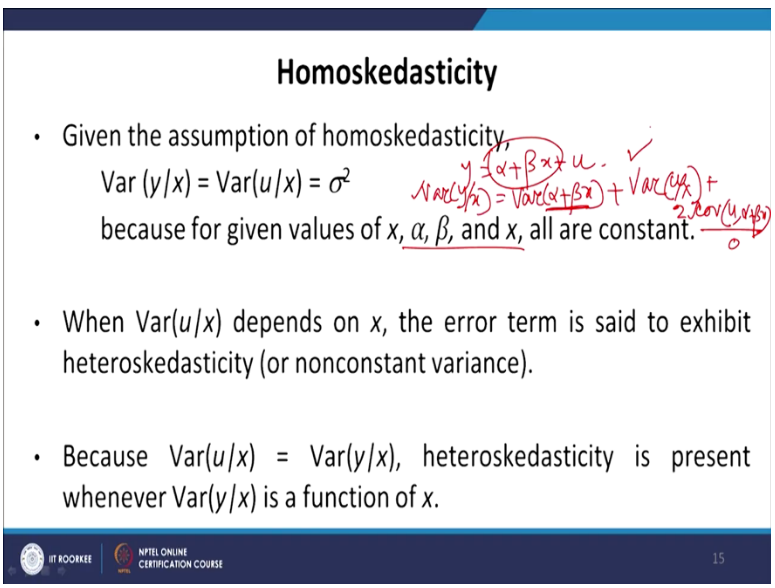


Now the simple regression model under homoskedasticity, how does it look like? So in the previous graph, I had shown you how the population regression function looks like. That was the 2D graph where we are measuring y on the vertical axis and x on the horizontal axis. Now if you also turn this diagram, then it would look very similar to what was shown in the 2D diagram.

Now by considering a three-dimensional where on the third dimension we are measuring the function of, functional form of y given x. So, this talks about the importance of homoskedasticity. You can see that for given values of x the values of y are centered on the population regression function. But their variations around the fitted lines or the population regression function are constant.

So the variations here, the variations here, the variations here all are the same. All do look very same, and these are the assumptions or implications of homoskedasticity, as opposed to the assumption of heteroskedasticity. I will also show you a graph quickly which shows how heteroskedastic errors would look like.

(Refer Slide Time: 24:09)

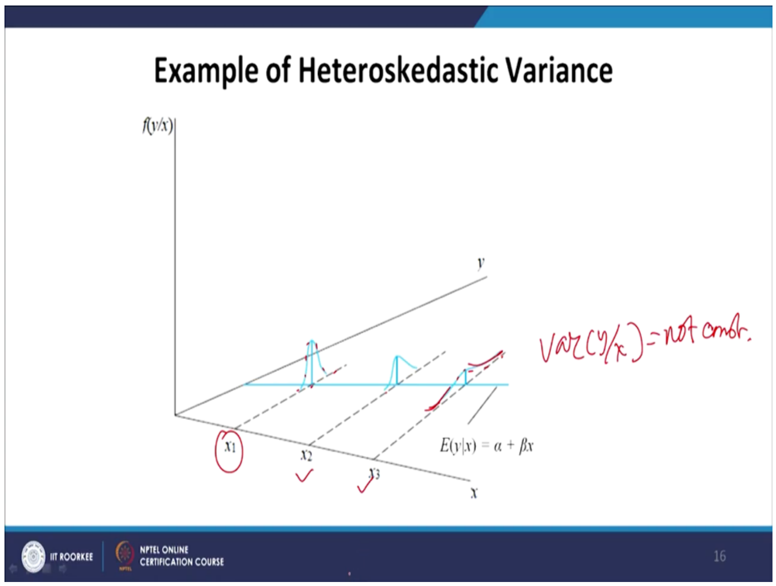


Now given the assumption of homoskedasticity, Var $(y|x)$ = Var$(u|x)$ = σ^2 because for given values of x, α, β, and x, all are constant. (refer slide time 24:27)

In that case, my (refer slide time 24:45) 0, because they are now no more random variables, they are not varying altogether, their variations would be 0. So their variance is 0. So what I am left with is actually variance of u given x and then there will be covariance terms between u and this part that is α plus beta x. So by now, we know that the covariance between u and x is 0.

So anyway, this part also becomes 0 (refer slide 25:24). When Var($u|x$) depends on x the error term is said to exhibit heteroskedasticity or non-constant variance, because Var($u|x$) = Var($y|x$). Heteroskedasticity is present whenever Var($y|x$) is a function of x.

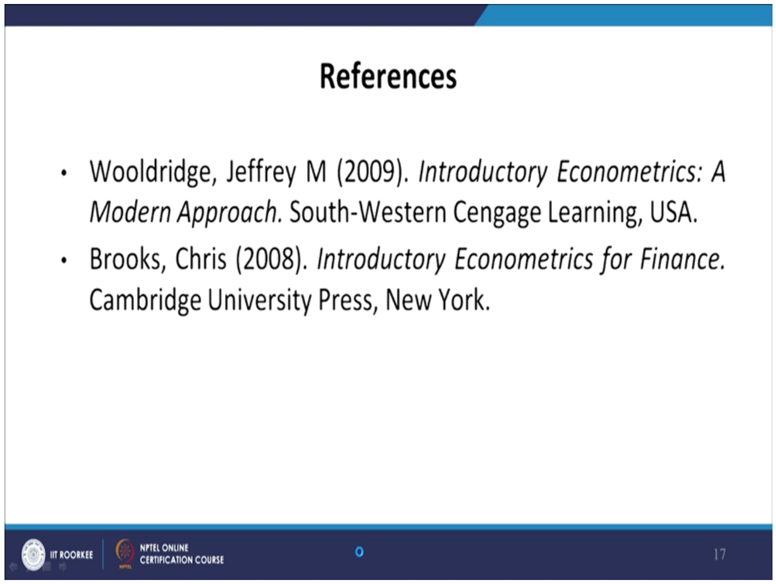
(Refer Slide Time: 25:46)



So, this is the presence of heteroskedasticity graphically. You can see that for individual values of x these are the values of y in repeated samples, and you can see that the variations are not the same for different values of xs'. So when we are changing x then the variations in y are actually not constant. So, the variance of y given x is not constant, and this is the problem of heteroskedasticity.

Somewhere the variations are different, that is reflected through these graphs that is how they are distributed. Their distributions are changing. Some are nearer to the mean, some are much broad-tailed, or fat-tailed as a result of which the variations are changing continuously, and we do not call it homoskedastic, error terms are homoscedastic, y series, they are heteroskedastic.

(Refer Slide Time: 26:54)



So that brings me to the end of this module. We will continue with the properties of OLS in the next module. Thank you.