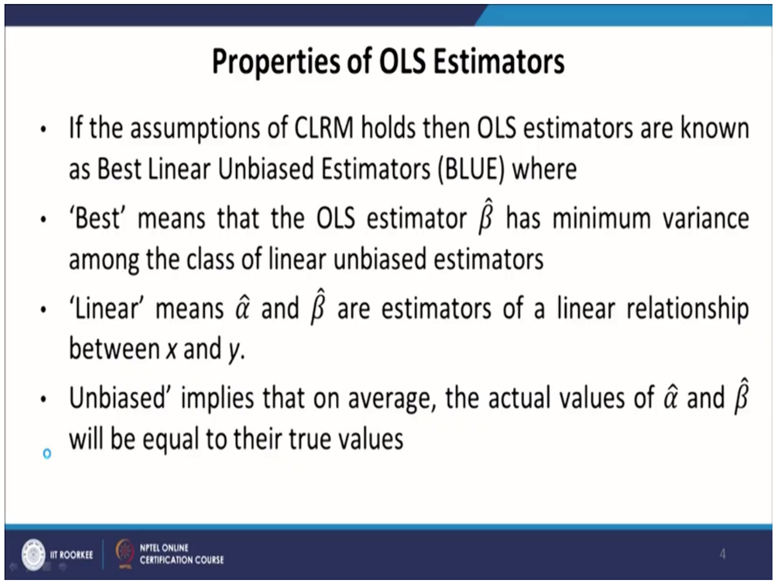
Econometric Modelling
Professor Sujata Kar
Department of Management Studies
Indian Institute of Technology Roorkee
Lecture 09
Properties of OLS Estimators

Hello. This is the ninth module of the course on Econometric Modeling. So, in the previous module, we have introduced the methodology under simple regression. And I have also discussed the assumptions behind OLS.

(Refer Slide Time: 00:44)

Now, this module basically talks about the properties of the OLS estimators. OLS estimators need to have certain properties which are primarily dependent on the assumptions that we make. These properties are essential for the OLS estimators because unless and until the assumptions are not fulfilled the properties will also not be met and accordingly the application of OLS on the certain problems might not be relevant or the right one. So, from those prospective, the properties of OLS estimators are important things to discuss.

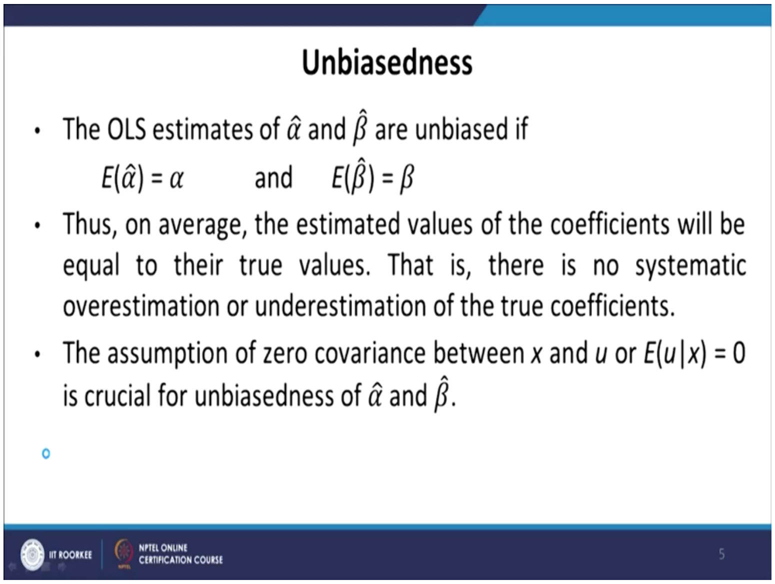
(Refer Slide Time: 01:23)



First of all, I just mention that if the assumptions of CLRM holds then OLS estimators are known as the best linear unbiased estimators (BLUE). So, this actually summarizes the properties or the most important characteristics of OLS estimators. Best here means that the OLS estimator $\hat{\alpha}$ *as well as* $\hat{\beta}$ has minimum variance among the class of linear unbiased estimators. Linear means $\hat{\alpha}$ *and* $\hat{\beta}$ are estimators of a linear relationship between x and y.

The concept of linearity has already been discussed at length and again here it implies that only linear relationships can be estimated using OLS. So $\hat{\alpha}$ *and* $\hat{\beta}$ are linear estimators of the relationship between x and y. They are unbiased. This implies that on average the actual values of $\hat{\alpha}$ *and* $\hat{\beta}$ will be equal to their true values.

(Refer Slide Time: 02:50)

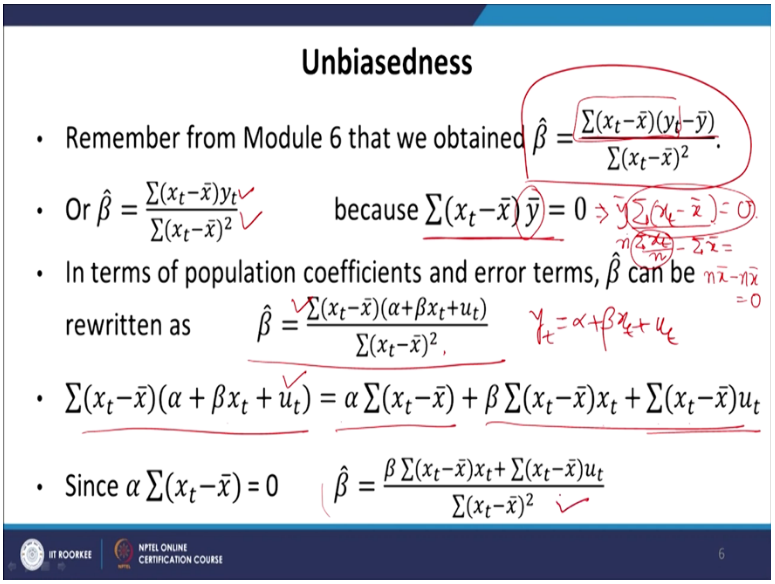


So first we talk about unbiasedness. I said it is represented by BLUE. So best linear unbiased estimators $\hat{\alpha}$ *and* $\hat{\beta}$ are estimators. We begin with first unbiasedness. Linearity has already been discussed. So, at the end or next, we will be discussing the best. And there is another property which is the property of consistency that will be discussed after that.

So, the OLS estimates of $\hat{\alpha}$ *and* $\hat{\beta}$ are unbiased if $E\left(\hat{\alpha}\right) = \alpha$ and $E(\hat{\beta}) = \beta$. This essentially implies that in repeated samples from the population if we obtain $\hat{\alpha}$ *and* $\hat{\beta}$ then their expected values will be equal to the population parameters. Thus, on average the estimated values of the coefficients will be equal to their true values; that is there is no systematic overestimation or underestimation of the true coefficients.

So, the true coefficients may vary slightly from the actual parameters that are alpha and beta. But the thing is that there is no systematic overestimation or under estimation. So, whatever little over-estimation or under estimation could there be that is completely random. The assumption of zero variance between x and u or expected value u, given x, equals 0 is crucial for unbiasedness of $\hat{\alpha}$ *and* $\hat{\beta}$. We will be proving the unbiasedness of $\hat{\alpha}$ *and* $\hat{\beta}$

(Refer Slide Time: 04:18)

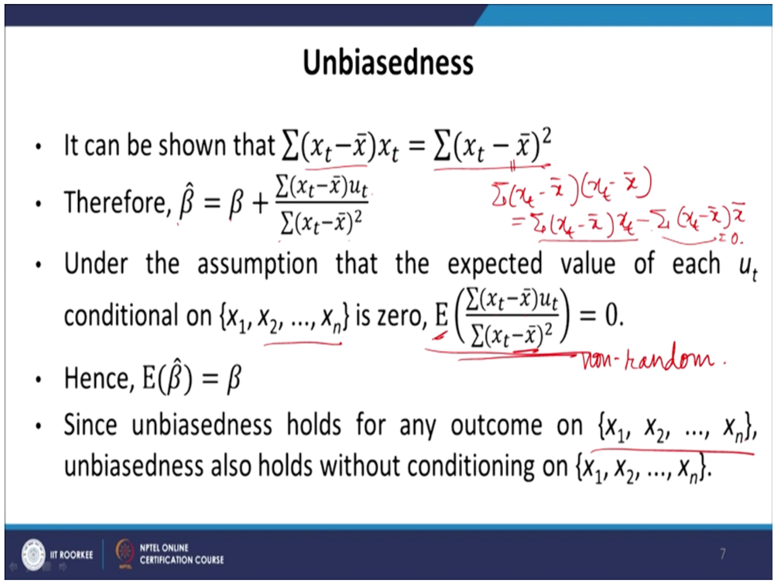


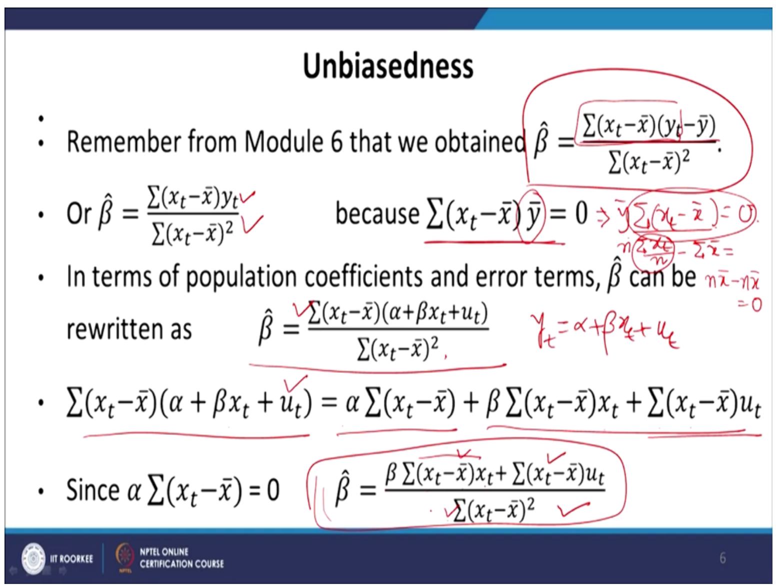
So first, of all, we consider $\hat{\beta}$. So, remember from module 6 that we obtained $\hat{\beta}$. That was the formula for $\hat{\beta}$ that we had obtained. Or this can also be written as (refer slide time 4:37) is because xt minus x bar multiplied by y bar is 0.

So, if I expand this expression, (refer slide time 4:53).

Now this can be written as further, now focusing on only this part, you can see that if I write (refer slide time 5:30 – 7:30)

(Refer Slide Time: 7:31)





Now it can be shown that (refer slide time 7:33). That is actually, can follow from here itself that this expression (refer slide time 7:45). The square is just broken. And if I further write it is the summation (refer slide time 7:50-8:17).

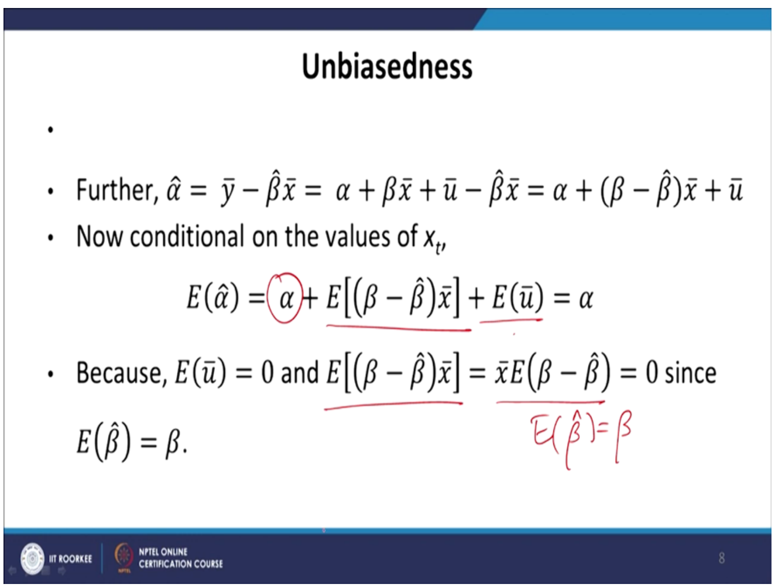
Now if you remember or rather let me show you that we had this expression for β where we have (refer slide time 8:25). I just showed that this expression is equivalent to this expression. So, if I separate these two entities then what happens is that (refer slide time 8:38 - 8:51).

And then we have (refer slide time 8:53- 9:10). This part will be 0 when we consider an expected value ahead of it. So, since this is conditional upon the x values, so all these summation x_t minus x bar whole square and summation x_t minus x bar, all of them become non-random.

And consequently, we have (refer slide time 9:34) and that is equal to summation, as well as along with the fact that we have covariance between u_t and x_t, these two assumptions taken together renders this expression 0. Hence, we have $E(\hat{\beta}) = \beta$.

Since unbiasedness holds for any outcome on x_1, x_2 to x_n , unbiasedness also holds without conditioning on x_1, x_2 to x_n. So on, for any outcome on x_1, x_2 to x_n implies that for any functional forms of xs' then biasedness holds. As a result of which we can say that it also holds without conditioning on any values of x.

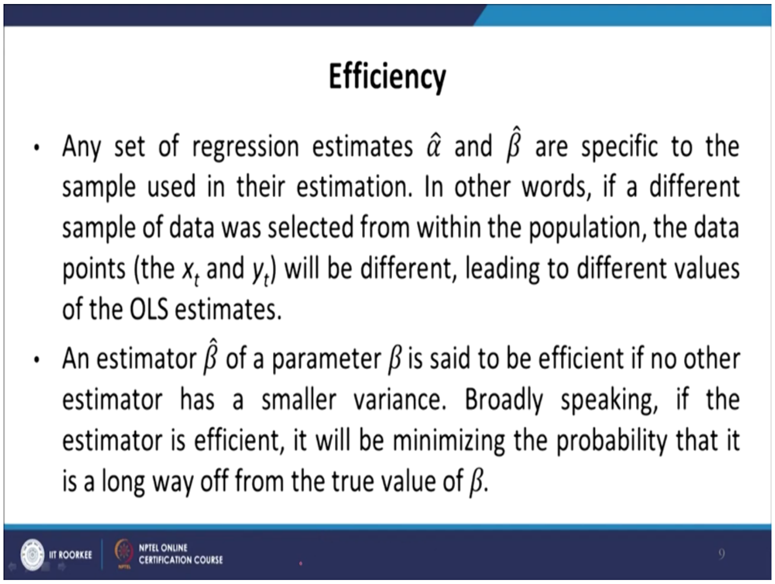
(Refer Slide Time: 10:21)



Now talking about the unbiasedness of alpha hat; (refer slide time 10:25 -11:31)

The expected value of the alpha hat is alpha i.e. $E(\hat{\alpha}) = \alpha$. That is how we prove the unbiasedness of $\hat{\alpha}$ *and* $\hat{\beta}$.

(Refer Slide Time: 11:37)



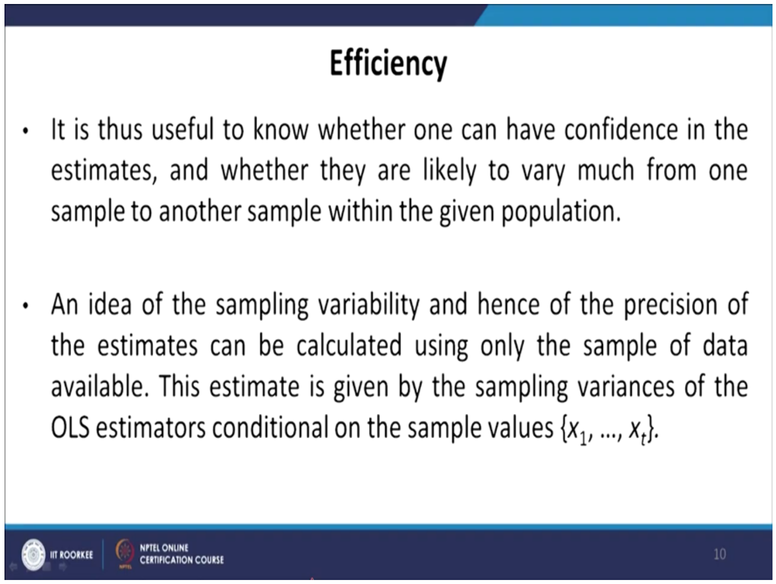
Now we talk about efficiency. So best implies minimum variance estimators. And this is an alternative term for best estimator is the most efficient estimators or alternatively, we call an assumption of efficiency. Any set of regression estimates $\hat{\alpha}$ *and* $\hat{\beta}$ are specific to the sample used in their estimation.

As I am saying that as the sample changes $\hat{\alpha}$, $\hat{\beta}$ or estimates of $\hat{\alpha}$ *and* $\hat{\beta}$ changes. In other words, if a different sample of data was selected from within the population the data points, that is xt and yt, will be different leading to different values of the OLS estimates that is $\hat{\alpha}$ *and* $\hat{\beta}$.

An estimator $\hat{\beta}$ of a parameter β is said to be efficient if no other estimator has a smaller variance. Broadly speaking, if the estimator is efficient, it will be minimizing the probability that it is a long way off from the true value of β. So, lower the variance lower is basically on average the deviations, $\hat{\beta}$ as an estimator we take from a population parameter β.

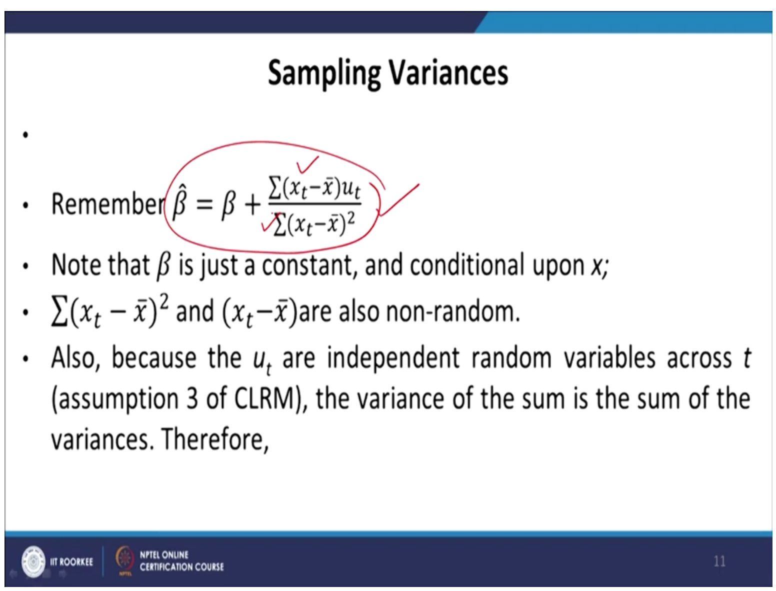
So lesser the variations the better the estimate is, and that is why this is a desirable property, the property of efficiency, that it is $\hat{\alpha}$ *and* $\hat{\beta}$ both are supposed to be the minimum variance estimators.

(Refer Slide Time: 13:11)



It is thus useful to know 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. So, lower the variations, sample to sample estimates will be also less different. An idea of the sampling variability and hence of the precision of the estimates can be calculated using only the sample of data available. This estimate is given by the sampling variances of the OLS estimators conditional on the sample values x_1, x_2 to x_t.

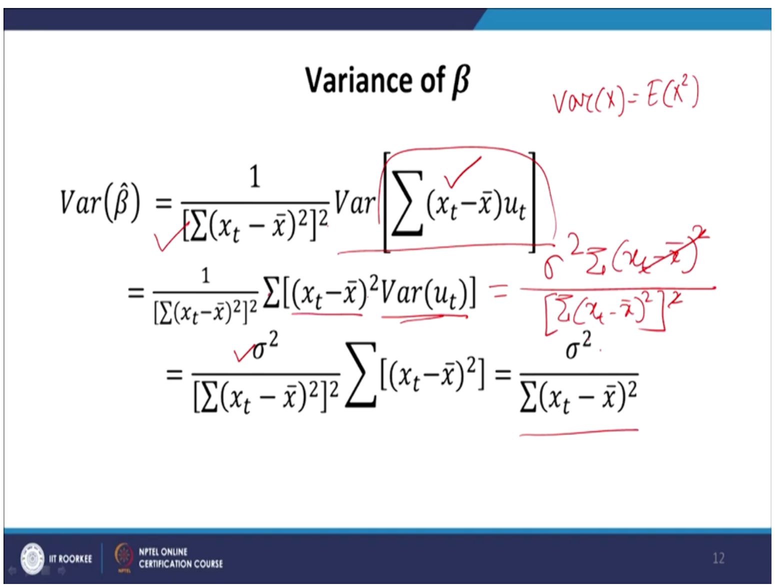
(Refer Slide Time: 13:47)



So now we talk about sampling variances, and we also basically derive the sampling variances of the estimators. Now remember this was the expression of $\hat{\beta}$. $\hat{\beta}$ we arrived at this expression, after that taking expected value of $\hat{\beta}$ we proved that $E(\hat{\beta}) = \beta$ which imply unbiasedness. Now we begin from this expression. Note that β is just a constant and conditional upon x.

So, which renders (refers slide time 14:24). So, this component and this entire component both also are non-random. Also, because the ut are independent random variables across t which is the assumption 3 or under CLRM that is classical linear regression model, the assumptions that we made in the previous module, the variance of the sum is the sum of the variances, and which implies that if all of them are independent then there will be no covariance between the terms of which variances are considered. And because of which what we obtain is, this is the variance of β.

(Refer Slide Time: 15:05)



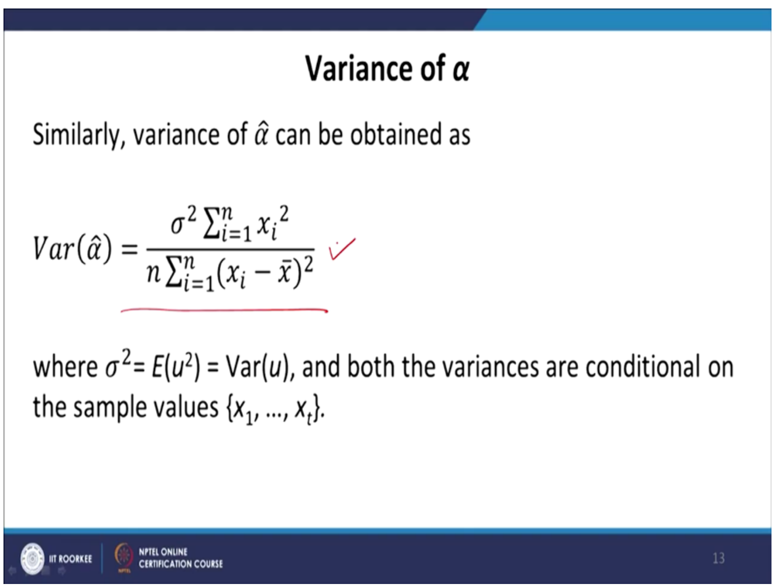
So how do we arrive at the variance of β? If you remember I discussed that (refer slide time 15:22), now which is the sum of individual observations, minus the mean of the variable. Now this implies that the moment I take a variable out of the variance expression I need to square it. So, this being non-random is taken out of the variance operator and it has been squared.

So, in the denominator or first we have (refer slide time 15:59). And we are left with the variance of the numerator expression (refer slide time 16:07). This is also non-random. So, this also comes out with a square term. And I am left with a variance of u_t only.

We already have discussed the variance of the population error term and that was denoted by sigma square. It can also be alternatively denoted by sigma square u. So currently going with the expression of sigma square, we have sigma square here. Now you note that sigma square being constant actually comes out of this summation operator.

So, when it comes out of the summation operator I have (refer slide time 16:53). So, this shows that this thing actually cancels out. And that is how we have (refer slide time 17:15)

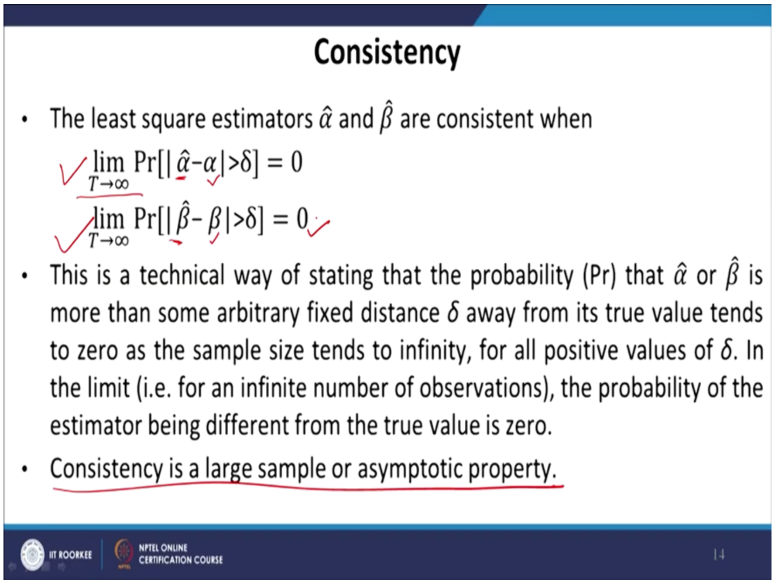
(Refer Slide Time: 17:24)



Similarly, the variance of $\hat{\alpha}$ can be obtained like this. Now variance of $\hat{\alpha}$, the derivation is slightly longer, so I am not getting into the derivation. All these are available in the standard textbooks I have referred to. Besides that, the derivation of this will be easier when we get into multiple regression and then in the matrix form, we can reduce the variance of $\hat{\alpha}$ *and* $\hat{\beta}$ or rather the variance of the constant term more easily.

So, this is the (refer slide time 18:00). And both variances are conditional on the sample values of x_1 to x_t. So that was about efficiency or minimum variance estimators. I had just derived the mean variances of the estimators. There are proofs that show that in the class of estimators OLS estimators are basically the minimum variance. I did not get into the proof that $\hat{\alpha}$ *and* $\hat{\beta}$ are the minimum variance estimators.

(Refer Slide Time: 18:47)

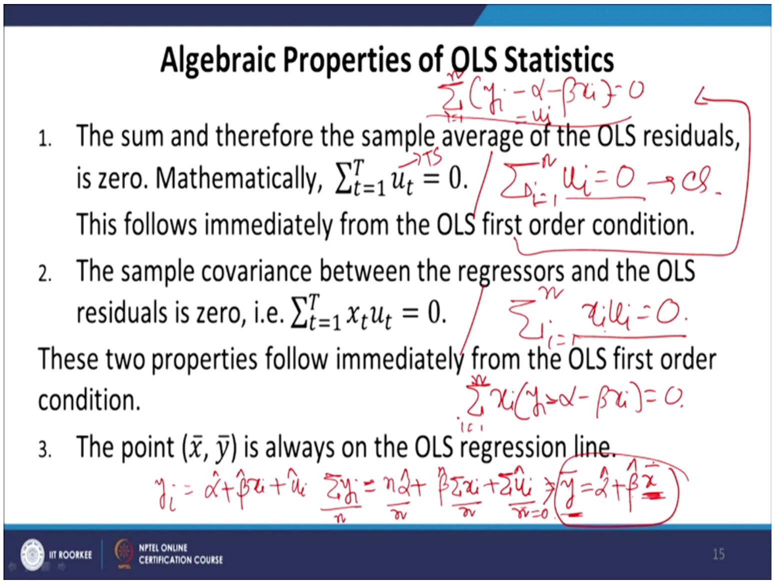


Next, I talk about the property which is consistency. This is basically a large sample or asymptotic property. So as the sample size increases then probably, we observe this property among the OLS estimators. So, the least square estimators $\hat{\alpha}$ *and* $\hat{\beta}$ are consistent when (refer slide time 19:02). And similarly, we have an expression for β. So, this is how we define consistency.

Now, what does it imply? This is a technical way of stating that the probability that $\hat{\alpha}$ *or* $\hat{\beta}$ is more than some arbitrary fixed distance δ away from its true value tends to 0 as the sample size tends to infinity for all positive values of δ. Alternatively, in the limit, that is for an infinite number of observations that is as my sample size grows, the probability of the estimator being different from the true value is 0.

So simply put, as my sample size grows to infinity, that is the sample approaches the population, the probability that the estimated $\hat{\alpha}$ or estimated $\hat{\beta}$ will differ from the population α and population β by any positive number is 0. So, this is the property of consistency, which is actually, as you can see a large sample or asymptotic property.

(Refer Slide Time: 20:25)



Now we briefly just talk about the algebraic properties of OLS statistics that is first the sum and therefore, the sample average of the OLS residuals is 0. So, we have assumed that the expected value of population error is 0. Its counterpart sample residuals tend to have an average which is equal to 0, (refer slide time 20:51).

Alternatively, this is, this can be written for (refer slide time 21:02). This is for cross-section, and this is for time series. Otherwise, there is as such no difference. This follows immediately from the OLS first-order condition. If you remember the OLS first-order condition was (refer slide time 21:24).

This is equivalent to u_i. So that is why it directly follows from OLS first-order condition, the sample covariance between the regressor and the OLS residuals is 0. So, we assume it for the population counterpart which is one important assumption of the CLRM. And for the sample, we are considering it to be one of the algebraic properties of the OLS statistics. And this is actually not an assumption.

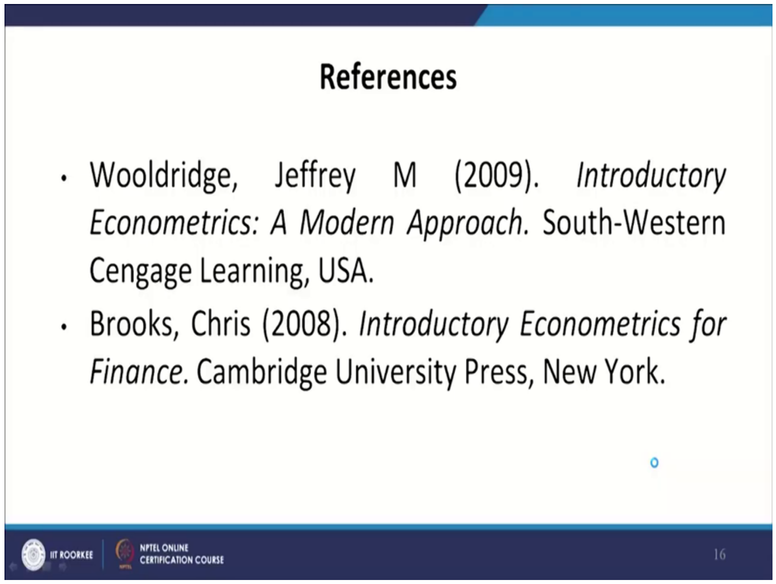
This is, ideally the sample counterpart of the population assumption that is the covariance between the regressor and OLS residuals, is 0. So, (refer slide time 22:18). These two properties follow immediately from the OLS first-order condition. So, this also follows from the OLS first-order condition.

This is obtained when I divide or when I take the first derivative of the residual sum of square with respect to $\hat{\alpha}$. This follows when we take the first derivative of the residual sum of square with respect to $\hat{\beta}$. So, if you remember this is equal (refer to slide time 22:56). So, these two properties follow from the OLS first-order condition.

And the final algebraic property is that the point x bar y bar is always on the OLS regression line. This has also very simple proof. (Refer slide time 23:23)

Now when I divide both the sides by n then you can see I can write it as (refer slide time 24:13). So, this shows that when x is at its average for given values of $\hat{\alpha}$ *and* $\hat{\beta}$, y will also be at its average. So $\hat{\alpha}$ *and* $\hat{\beta}$ basically for any values of the slope and the constant term; when x is at its average, $\hat{y}$ will also be at its average. And since this expression is true that is why we can say that the regression line always passes through the averages of both variables.

(Refer Slide Time: 25:04)



So, this brings me to the end of the discussion on the properties of OLS estimators. These are the books that you can follow for the discussion on the properties of OLS estimators. Thank you.