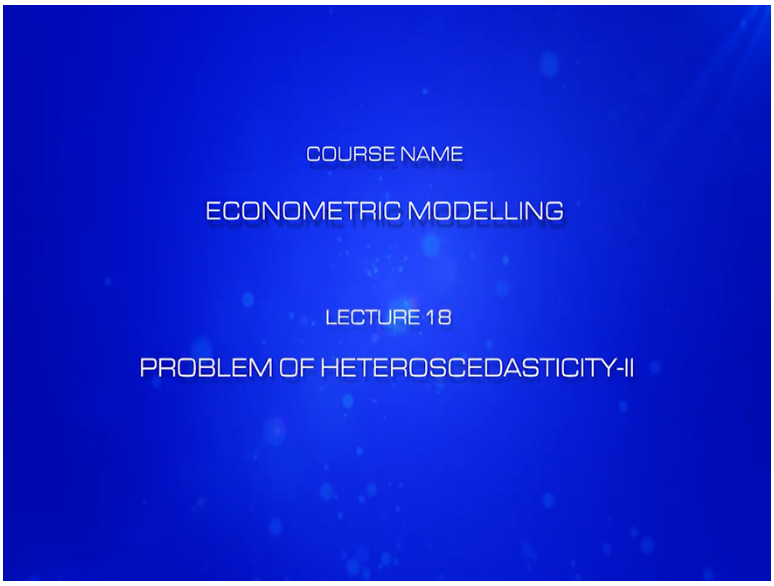
Econometric Modelling
Professor Sujata Kar
Department of Management Studies
Indian Institute of Technology, Roorkee
Lecture 18
Problem of Heteroscedasticity - II

(Refer Slide Time: 00:11)



Hello, everyone. So, I continue with the problem of Heteroscedasticity in Module 18 of the course on Econometric Modelling.

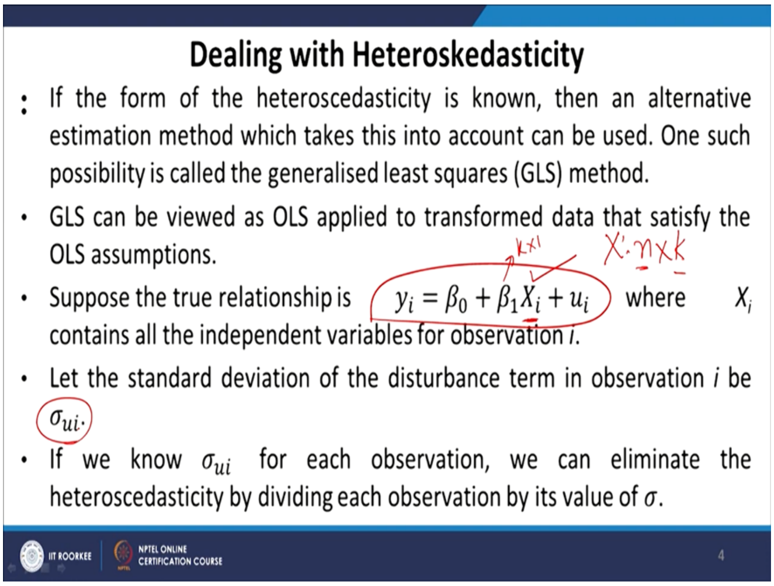
(Refer Slide Time: 00:35)

In Module 17, we discussed the definition of Heteroscedasticity, what kind of problem it leads to, and then how we can test for Heteroscedasticity. There I had discussed four alternative tests for the presence of Heteroscedasticity and they were the Spearman rank correlation coefficient test, we discussed Goldfeld- Quandt test, then we discussed the Breusch-Pagan test, and finally, the White's test.

Now, I'm going to discuss how to deal with Heteroscedasticity. Now, dealing with Heteroscedasticity has actually two different aspects. So, when we study as I have mentioned in the previous module, there is something called autoregressive conditional heteroscedastic errors.

So in those situations that heteroscedasticity is actually modelled, or the heteroscedastic errors are incorporated into the models instead of trying to remove the problem of heteroscedasticity. But here currently, what we try to do is, primarily, while dealing with heteroscedastic errors, we try to remove the problem of heteroscedasticity by removing the problem of heteroscedasticity, what we obtain is, of course, the efficient estimators, but we are not actually sure about whether they are unbiased or consistent or not. So, most often they will not be unbiased estimators, but they are certainly the best estimators that are the most efficient estimators.

(Refer Slide Time: 02:11)



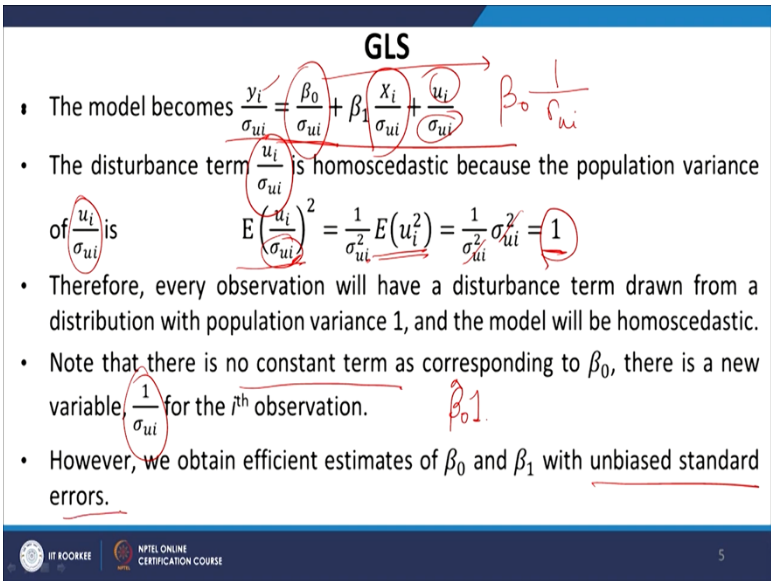
So if the form of heteroscedasticity is known, which is actually our trouble many times, then an alternative estimation method that takes this into account can be used. One such possibility is called the generalized least square or GLS method.

GLS can be viewed as an ordinary least square technique or method applied to transform data that satisfy the OLS assumptions. So, whenever we transform or modify the data, so that the data fulfills the basic assumptions of OLS or CLRM assumptions or Gauss Markov theory assumptions, then that is called GLS or generalized least square.

Suppose, the true relationship is given by this where this x contains all the independent variables for observation i. So, this is for a single observation. So, if you remember our x was an n by k matrix where there were n observations. So going by that this x actually will have only one observation and of course, k independent variables.

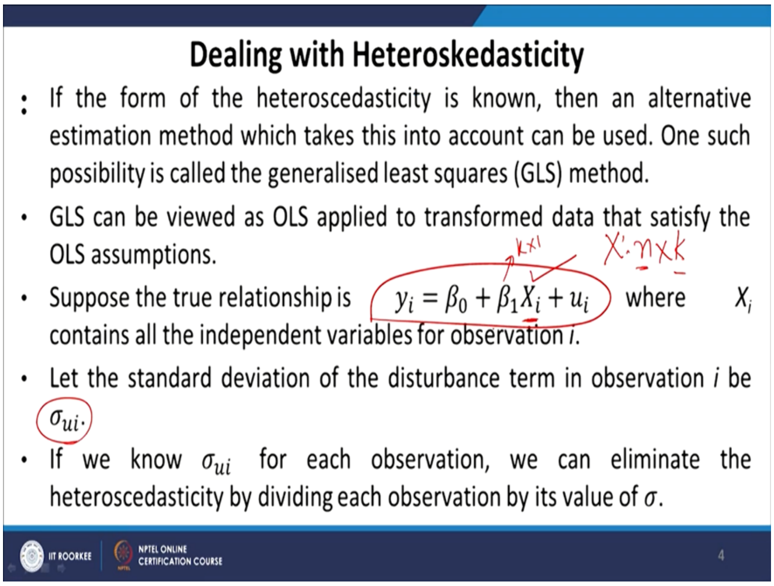
So, (refer slide time 3:30- 4:06).

(Refer Slide Time: 04:16)



If we know (refer slide time: 4:07- 4:38).

(Refer Slide Time: 04:38)

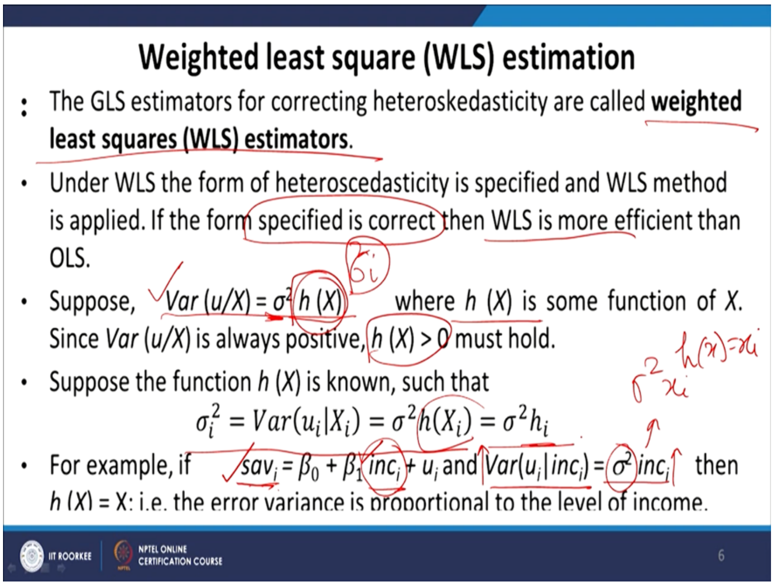


So (refer slide time 4:38- 5:53).

Therefore, every observation will have a disturbance term drawn from a distribution with population variance 1 and the model will be homoscedastic. Note that, there is no constant term as corresponding to (refer slide time: 6:10).

So, so far for all the observations, we had one constant term that was 1, and corresponding to that constant term our parameter was β_0 and the estimated parameter was $\hat{\beta}_0$, but now what I am having is (refer slide time: 6:30).

(Refer Slide Time: 07:15)



Now, σ_{ui} is an observation-specific variable and as a result of which, we do not have any constant term. However, we obtain efficient estimates of β_0 and β_1 with unbiased standard errors, because you can see that since we are now having a homoscedastic variance, our estimates are now efficient, and the standard errors are actually unbiased estimators of the population standard deviation.

The GLS estimators for correcting heteroscedasticity are called weighted least square estimators. So GLS can be applied in various situations or in various circumstances. Now, when we use them for correcting heteroscedasticity then they are specifically called weighted least square estimators. So, GLS or WLS is actually one kind of GLS. Under WLS, the form of heteroscedasticity is specified and the WLS method is applied if the form specified is correct, then WLS is more efficient than OLS.

Of course, specifying the form correctly is actually not an easy thing. But if we can specify the form, then as I have just explained, that WLS becomes more efficient. So, the previous example also is an example of WLS only, but in a more generalized context. Now, we are becoming more specific by bringing in the definition of WLS. Suppose, (refer slide time: 8:19).

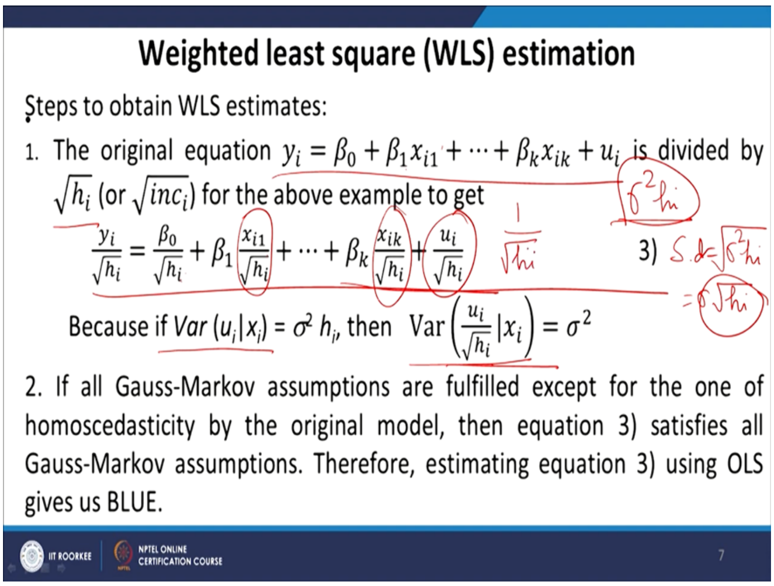
So instead of writing σ_i or σ_{ui}, I am now specifying the functional form. And in that functional form, we have a constant component which is σ^2, and a component that is dependent on the independent variables, which is given by x. Now, again, I am still not specifying the functional form of the independent variables. It is just that the σ_i^2 or σ_{ui}^2, has been segregated by taking out a constant component denoted by σ^2 and leaving the rest as dependent on the independent variable.

Now, since the variance of u/X is always positive, $h(X) > 0$ must hold because if $h(X)$ is not greater than 0, it is a negative number, then the variance of u/X will also be a negative number. But that is not possible because variances are always calculated on the basis of the squared sum. So, squared sum of the deviations of the individual values from its mean and as a result of which they can never be negative numbers.

Suppose the function (refer slide time: 9:40). For example, this is my equation where we are considering savings as the dependent variable and it is regressed on income. So, we are trying to find out whether savings are impacted by the incomes of individuals or not.

So this is an individual specific observation where we are saying that we are trying to find out whether the income of an individual i impacts his or her savings or not and suppose (refer slide time: 10:25). That is the error variance is proportional to the level of income. As income increases, the error variance also increases proportionally because σ^2 here is a constant term.

(Refer Slide Time: 11:02)



Now, we discussed the steps to obtain WLS estimate. The original equation is again given by this, now it is divided by $\sqrt{h_i}$. So, I stick to my previous example, if the form of variance is given as $\sigma^2 h_i$, then this is variance as a result of which my standard deviation or standard deviation of the population error will be $\sigma\sqrt{h_i}$. Now, since σ is constant, we are ideally dividing this equation by only r$\sqrt{h_i}$. So, if I divide it by $\sqrt{h_i}$ then what happens?

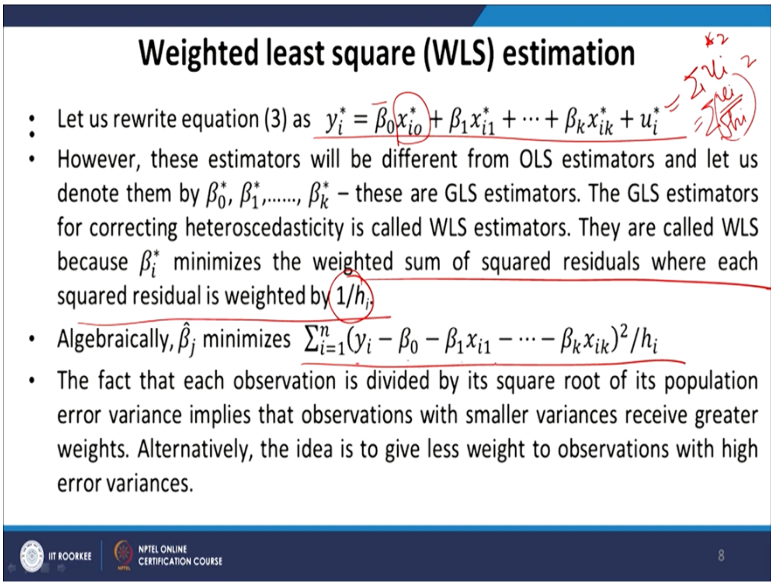
We have this expression. So, (refer slide time: 11:57).

So, if all Gauss Markov assumptions are fulfilled, except for the one of homoscedasticity by the original model, then equation three satisfies all Gauss Markov assumptions because our only problem was that our errors were not homoscedastic. Errors were heteroscedastic. But if I write equation three, if I write it in this form, knowing the form of heteroscedasticity then you can see that the variance of this term, (refer slide time: 12:55).

So, these are actually homoscedastic errors. So if homoscedastic errors are there and other assumptions of Gauss Markov theorems have already been fulfilled, then we can estimate equation 3 using OLS and then OLS would give us the best linear unbiased estimators. So we have the most efficient estimators, the best estimators along with the fact that the estimators will also be unbiased.

But one thing we need to notice here, and that is the variables are not now, x_1, x_2, and so on. Instead, what are the variables? (Refer slide time: 13:48).

(Refer Slide Time: 14:03)

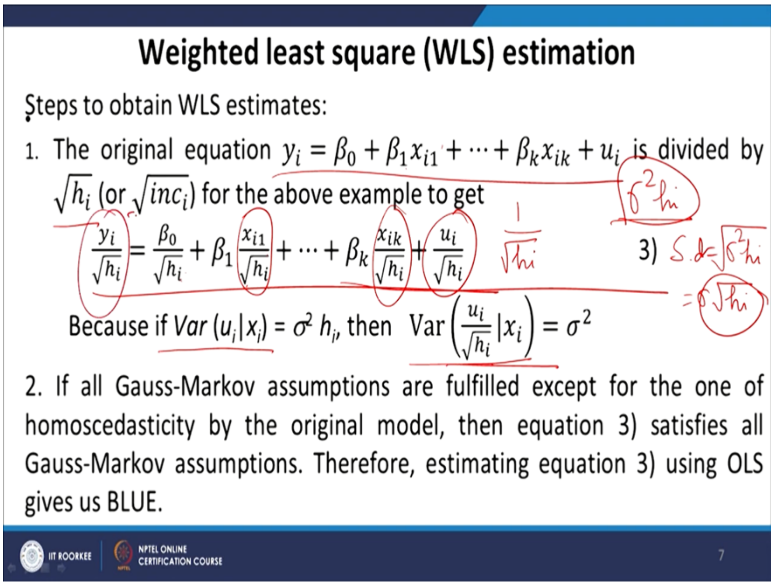


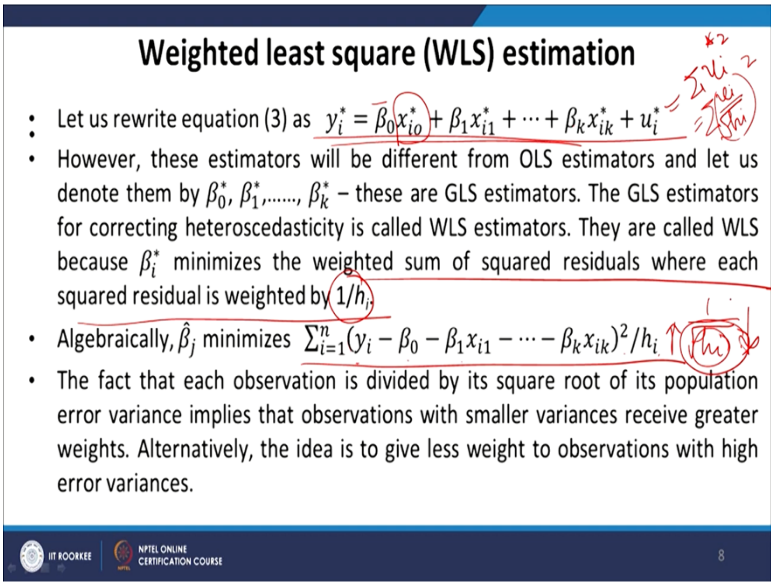
So, we rewrite equation-3 as (refer slide time: 14:08). However, these estimators will be different from OLS estimators, because as you can see that our variables have changed. So, if our variables have changed, then of course, the estimated parameters would also change and let us denote these estimated parameters or other parameters as (refer slide time: 14:43) and they are the GLS estimators. They could be different from OLS estimators.

The GLS estimators for correcting heteroscedasticity is called WLS estimators. They are called WLS because β_i^* minimizes the weighted sum of squared residuals, where each squared residual is weighted by $\frac{1}{h_i}$. So if you remember earlier when we talked about OLS $\hat{\beta}_j$ was designed to minimize the sum of square residuals and we did not use any weights there.

But here in case, in this context, what we are minimizing? We are minimizing (refer slide time: 15:34- 16:08).

(Refer Slide Time: 16:28)





So algebraically (refer slide time: 16:10). So I write the original model and then the entire thing divided by h_i because for each and every variable it is divided by $\sqrt{h_i}$, and when I go for squaring the entire expression, then basically this is h_i in the denominator. So that's why we have h_i.

The fact that each observation is divided by the square root of its population error variance implies that observations with smaller variances receive greater weights. Alternatively, the idea is to give less weight to observations with high error variance. So, observations that are the combination of y and x, suppose there are only two variables one y and one x, one dependent variable and other independent variables.

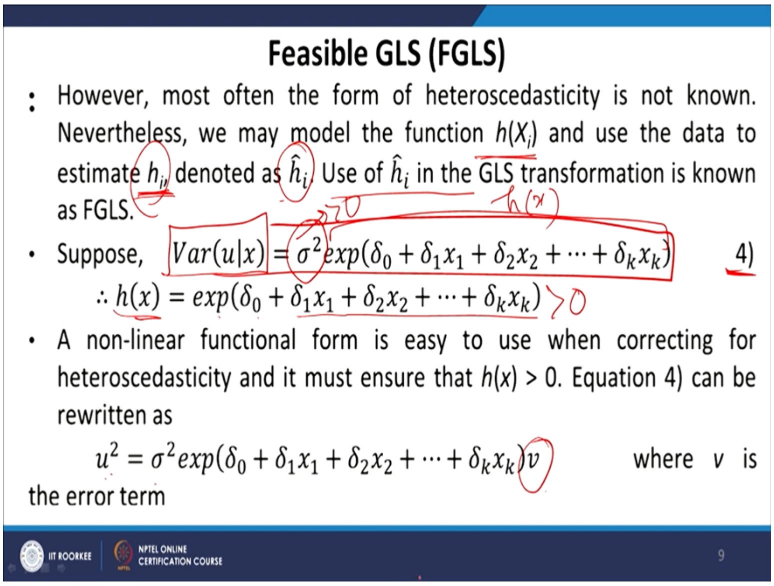
Now, if we observe very high variance between y and x specific to a particular observation, then we would consider the quality of that observation to be poor compared to observations having lower variances because lower variances implies that they are more close to the mean of the series and the mean of the series is expected to represent the population expected value or population mean as a result of which we can say that the quality of those observations having lower variances is actually higher or better.

Now, in this scheme of things that is in weighted least square technique, what we are doing is that, giving lower weights to the observations having higher dispersion or higher variances because what we are doing is that, taking aside the constant term, we are dividing each and every observation by $\frac{1}{\sqrt{h_i}}$ or $\frac{1}{h_i}$.

Now, the thing is that since h_i is a part of the error variance, the larger the error variance, the larger the denominator, and the smaller is the weight. Alternatively, the smaller is the error variance corresponding to a particular observation, the smaller is h_i and larger are the weights associated with that observation.

So that is how we are actually going for an estimation where observations having higher variances are actually penalised by assigning lower weights to them and better quality observations having lower dispersions or deviations from their mean value are given higher weight in terms of $\frac{1}{\sqrt{h_i}}$.

(Refer Slide Time: 19:08)



Then, we talk about something called feasible GLS. So this actually is applicable when you do not know the form of heteroscedasticity. So most often the form of heteroscedasticity is not known. Nevertheless, we may model the function hX_i and use the data to estimate h_i denoted as $\hat{h}_i$. Use of $\hat{h}_i$. hat in the GLS transformation is known as FGLS or Feasible GLS.

So, what exactly we are trying to do here is that we are trying to hypothesize the functional form of the relationship between the independent variable and the error variance. So, the relationship between independent variables and the error variance is given by hX_i or h_i. And on that basis of that hypothesized relationship, we try to find out whether such relationship actually exists or not, and if we observe that relationship actually exists then we incorporate that relationship. So, feasible GLS actually goes one step further and first tries to estimate the relationship between the error variance and the independent variables.

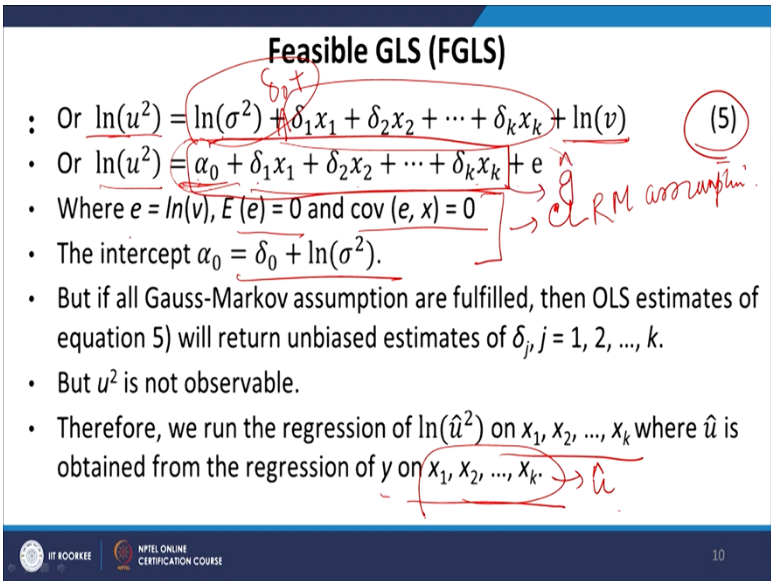
So, suppose the (refer slide time: 20:30- 21:00).

This is the same condition which we required earlier also because error variances have to be 0 and as a result of which $h(X)$ has to be greater than 0 provided σ^2 is always greater than 0. Equation 4 can be written as (refer slide time: 21:16).

What is done is that instead of having this variance of u given x, we are considering u^2 because that variance of u given x is actually the expected value of u^2. And then by taking out of the expectation operation, we are introducing a random error term here, which is v.

Now, once it is done, it takes an estimable form and it can also be converted into a linear functional form which could be estimated using the OLS method.

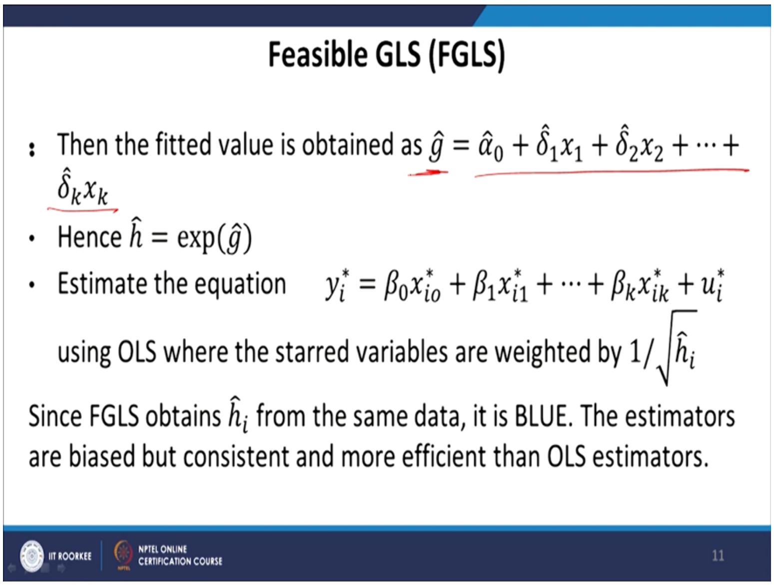
(Refer Slide Time: 21:59)

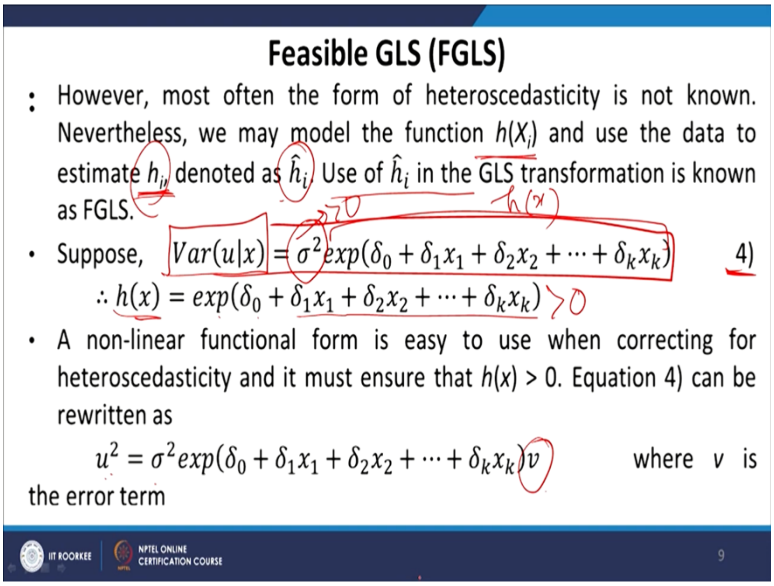


So we write by taking natural logarithm. On both sides, we write it (refer slide time: 22:06-23:05).

These are required as part of our classical linear regression model assumptions, CLRM assumptions. But, if all Gauss Markov assumptions are fulfilled, then OLS estimates of Equation 5 will return unbiased estimates of (refer slide time: 23:26).

(Refer Slide Time: 24:03)

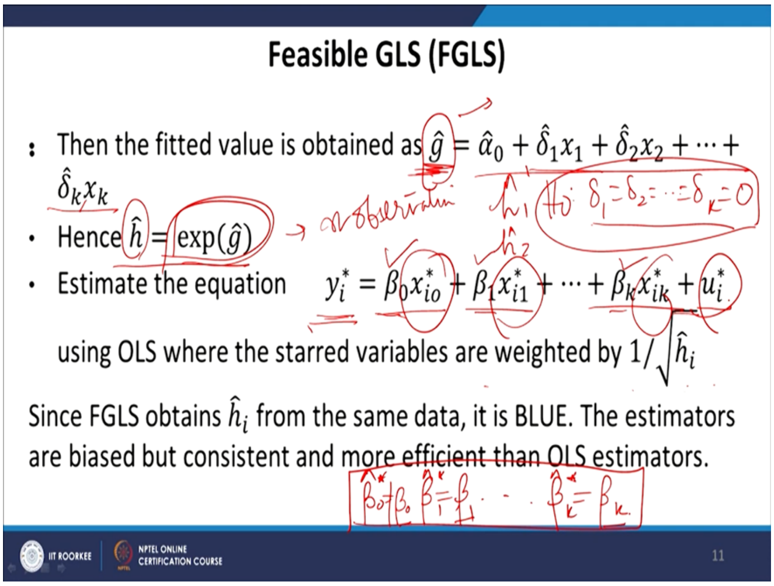




So first of all, we go back to the original model, run a regression of y on x_1, x_2 to x_k collect the residuals that are $\hat{u}$ then run a regression of $\hat{u}$ on x_1, x_2 to x_k. (Refer slide time: 24:05).

Now, you can see that my original equation (refer slide time: 24:22- 25:07)

(Refer Slide Time: 25:21)



So, there are slight differences, but broadly I can consider it to be (refer slide time: 25:15-26:02).

In the case of feasible GLS, what has been done, which is different from WLS is that we do not know the functional form or rather we have hypothesized a functional form of the error variance with respect to the independent variables, and then we have tried to come up with an estimate of the hypothesized relationship. The estimate of the hypothesized relationship is given by (refer slide time: 26:27).

Now, you can see that, if the error variance is not dependent on the independent variable, which alternatively means that there is no heteroscedasticity then $\delta_1 \ldots \delta_k$ will be equals to 0. So, in case we have a null hypothesis, which says (refer slide time: 27:05), this null hypothesis will not be rejected.

And if this is not rejected, then this means this functional specification that actually has two alternative implications. First of all, this kind of a linear functional specification or this kind of nonlinear functional specification is actually not valid. So, this relationship does not exist.

Alternatively, it can also mean that the error variance is actually non-heteroscedastic. So, it is homoscedastic and in that case of course, we do not need to apply or go for a WLS. But if all or some of the $\delta_1 \ldots \delta_k$ turns out to be nonzero statistically then we can go for this kind of an

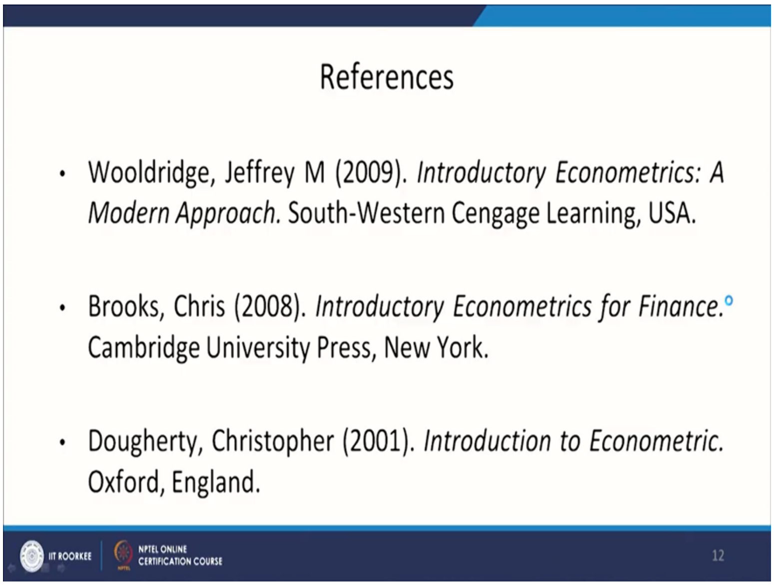
FGLS or feasible GLS where we divide the original variables by $\sqrt{\hat{h}_i}$, the estimated a hypothesized functional form and then obtain the values of β_0, β_1 ... δ_k.

Again, this β_0, β_1 ... δ_k are not equivalent to the OLS estimators or OLS estimates because you would see that again our variables are not what we have observed in case of OLS or what we would have used in case of OLS. The variables have changed, which has already been discussed.

So, since FGLS obtains $\hat{h}_i$ from the same data, it is actually BLUE. The estimators are biased but consistent and more efficient than the OLS estimators. So, we call it biased. The reason is whatever we estimate, suppose the estimated values are denoted by or estimated parameters are denoted by (refer slide time: 28:58).

And that is why the estimators are biased, but we can easily prove that they still are consistent which is a large sample property, and they are more efficient. So, consistency is there the long we assume that the covariance between x and u or the covariance between these variables and u_i^* is equal to 0. The estimators are consistent and they are more efficient than the OLS estimators.

(Refer Slide Time: 30:09)



So that brings me to the end of the discussion on heteroscedasticity. We have discussed different perspectives, that is its definitions, detections as well as how to deal with or how to correct for heteroscedasticity. These are the references. Thank you.