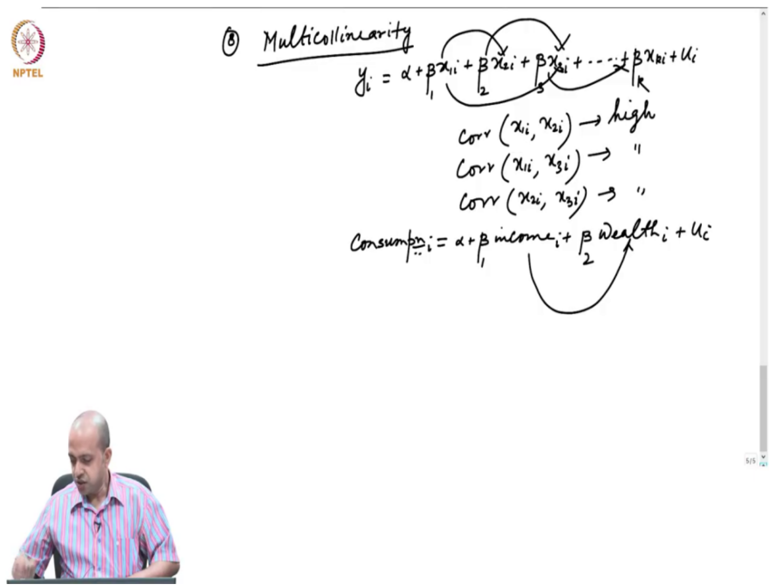
Introduction to Econometrics
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras
Lecture 10
Classical Linear Regression Model Part-4

(Refer Slide Time: 00:22)



Our next assumption after autocorrelation is the assumption of multicollinearity. Let us say our model is $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_k x_{ki} + u_i$. So, basically multicollinearity says that there should not be a high degree of correlation between any two variables. That means the correlation between x_{1i} and x_{2i} or correlation between x_{1i} and x_{31} or correlation between x_{2i} and x_{3i} should be insignificant or very small. If it is high then we say that there is a presence of multicollinearity. Collinearity means perfect linear relationship and multi means multiple like perfect linear relationship between x_1 and x_2, x_2 and x_3 or x_3 and x_k.

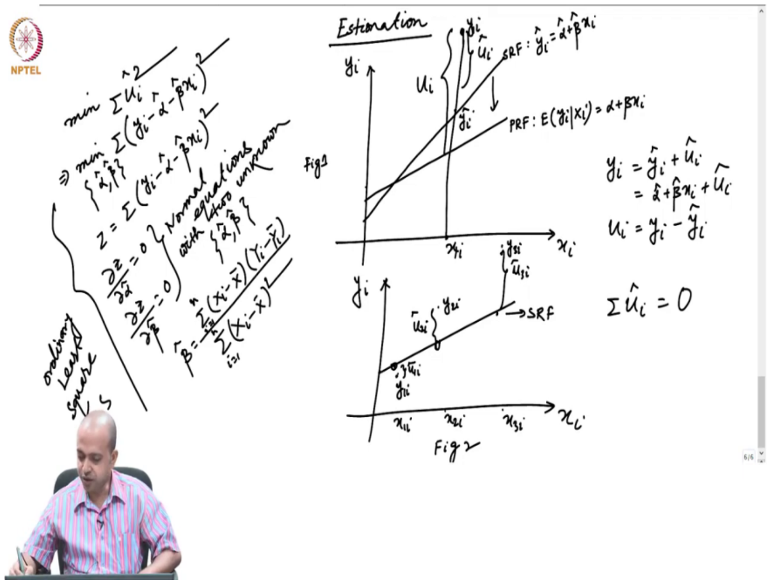
So, that means we have perfect linear relationship and when number of such relationship is multiple, it is linear multiple multicollinearity. For example, let us say consumption=f(α +β_1 income + β_2 wealth + u_i). So, consumption function theory says that the consumption depends not only on income but also on wealth and it is possible that income and wealth is highly correlated among each other. If that is the case then we say that this particular data suffers from multicollinearity problem and again in the presence of multicollinearity problem, the three desirable properties of unbiasedness, efficiency and consistency may get disturbed.

So, these are the assumptions that we specify before we estimate the model using CLRM. After maintaining this assumption we need to proceed for estimation and now we will try to understand how we estimate a model and the technique that we apply to estimate the model.

Once again we can recap the assumptions. We say that firstly the model is linear and there should not be any non-linearity in parameters. It might be linear in variable. The second assumption says that when you specify the model then expectation of u_i given x_i is equal to 0 and u_i follows a normal distribution. The third assumption say that in the model we specify x_i is fixed repeated sampling. Assumption number 4 says that the covariance between x_i and u_i is 0. So, x_i is strictly exogenous in the model. The fifth assumption says there is no model misspecification either due to improper functional form or due to inclusion of irrelevant variable or exclusion of any relevant variable. The sixth assumption says that the total number of observations should be much-much higher than the number of parameters to be estimated. The seventh assumption say that in your model there should be enough variation in x_i and y_i. Assumption eight says that there should not be any autocorrelation. Assumption nine says that there should not be any heteroscedasticity and number 10 says that there should not be multicollinearity. But basically these are the assumption in your textbook also you will see there are 10 such assumptions. I might have made some mistake that is why my counting is coming 8. But there are actually 10 such assumptions. So, once again we specify these assumptions.

Then the next step is how to estimate the model.

(Refer Slide Time: 06:56)



So, I will try to make you understand about the estimation procedure with a simple diagram. Once again this is x_i and this is y_i and this is your line. Let us, say this is my PRF which is given by expectation of y_i given x_i and is equal to $\alpha + \beta x_i$ and let us say this is your SRF so that means this is the population regression function which we are trying to infer about because the true population parameters alpha and beta are never known to us.

So, that is why this is the population regression function but we have estimated this SRF and this is basically $\hat{y}_i$ that we have estimated and is equal to $\hat{\alpha} + \hat{\beta} x_i$. Now let us say for any given value of x_i this is my x_i. What will happen if you draw a line? This is your y_i and this y_i is actually your actual or observed y_i. So, this y_i indicates the actual y_i. Let us say this is y_{1i} and you can write this is x_{1i}. For a specific value of x_{1i} your observed y is y_{1i}. Let us say this is $\hat{y}_i$ and this is your y_i -population regression function. This is your srf. So from this relationship I can write that $y_i = \hat{y}_i + u_i$.

So, that means we can write $\hat{\alpha} + \hat{\beta} x_i + u_i$. This is what you can write and u_i is basically $y_i - \hat{y}_i$. That means in this technique we are trying to fit a line which is basically srf by drawing a particular sample from the population. So that we can infer something about this prf and if you draw multiple samples then you will get multiple such srf and our objective is to estimate the srf in such a way that srf goes as close as possible towards this prf. This is what we are trying to estimate. We specify a line once again- this is x_i and this is y_i. This is figure one. Let us say this is figure 2. This is x_i and this is y_i. So, for any given value of x_i you will get your predicted y_i like this. Your actual y_i would be let us say this is y_{1i} but this is your predicted value so that means you will commit some mistake.

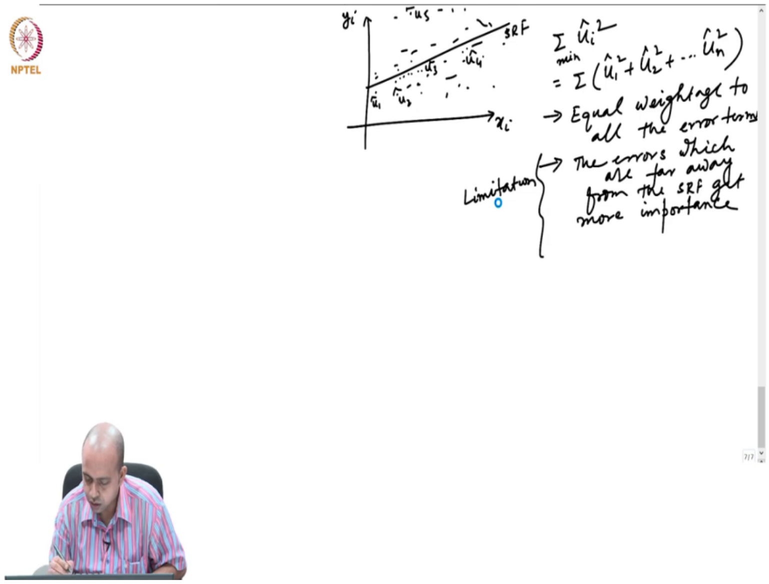
Let us say, this is u_{1i} when x is x_{1i} and then for the second one let us say this is x_{2i} which is your predicted but your actual may go here this is let us say y_{2i}. So, that means this is u_{2i}. Similarly for x_{3i} let us say, this is your actual y_{3i} so that means this is your $\hat{y}_i$. So that means this becomes your $\hat{u}_{3i}$. This is $\hat{u}_{1i}$, this is $\hat{u}_{2i}$ and this is $\hat{u}_{3i}$. We will try to minimize this error. If you minimize the sum of these errors then your srf will go as close as possible to the prf. You have estimated a line and then you are trying to predict somebody's consumption given his or her income now. When you specify a given level of income, your model says that individual's consumption should be here on the line.

But the actual consumption is lower than this line. That means you are committing a mistake. Let us say, that is $\hat{u}_{1i}$ for x_{1i} and your model predicts the consumption here. But your actual consumption is over the line. That means there also you are committing mistake. That means some errors are positive and some errors are negative and our objective here is to minimize this errors. The sum of $\hat{u}_i$ is actually 0 because this srf line what we have discussed earlier is basically an average line. So all these errors $\hat{u}_1$, $\hat{u}_2$ are basically deviations from the average line and the sum of deviation from the average or mean is actually 0. That is why summation of $\hat{u}_i$ is 0. We cannot actually minimize this $\hat{u}_i$ rather we have to minimize summation of $\hat{u}^2_i$ because otherwise the positive errors will get cancelled out by the negative errors.

If you look here $y_i = \hat{\alpha} + \hat{\beta}x_i + u_i$ .So that means basically you are minimizing $(y_i - \hat{\alpha} - \hat{\beta}x_i)^2$ and the control is that you are trying to minimize this with respect to α and β. To minimize you have to differentiate this function. If you differentiate with respect to α, you will get one equation and you have to set it equal to 0. By the rule of minimization or maximization, you have to differentiate this again with respect to β. Let us assume that z = **Σ** $(y_i - \hat{\alpha} - \hat{\beta}x_i)^2$. So, basically then minimization requires δz/δ which is equal to 0; and del z by del beta hat which is also equal to 0. If you write these two then you will get two normal equations with two unknowns- $\hat{\alpha}$ and $\hat{\beta}$. If you solve these two equations then you will get your $\hat{\beta}$ which is equal to $\frac{\sum(x_i - \bar{x})(y_i - \bar{y})^2}{\sum(x_i - \bar{x})^2}$. This is i running from 1 to n. This is your $\hat{\beta}$.

This particular technique what we have applied here is known as Ordinary Least Square method or OLS in short. So, this is actually the meaning of estimation and this is what we actually do. We minimize the sum of $\hat{u}^2{}_i$. Now, couple of things that we need to remember here.

(Refer Slide Time: 19:28)



This is x_i (along the x axis) and y_i (along the y axis) and this is your line. Let us say that these are all my error terms- $\hat{u}_1$, $\hat{u}_2$, $\hat{u}_3$ $\hat{u}_4$ $\hat{u}_5$ and so on. When you are minimizing the summation $\hat{u}_i$ it is basically minimizing summation of $\hat{u}^2{}_1 + \hat{u}^2{}_2$ etc upto $\hat{u}_n^2$. This means I am giving equal weightage to all the error terms. If you square it up, $\hat{u}^2{}_1$, $\hat{u}^2{}_2$ and all these get equal weightage. So, if you look at from the diagram some of the points are very close to your estimate line which is why some predictions are very close towards the point towards the srf and some predictions or some points or some errors are much away from the line.

So, obviously the lines which are closer to the line should get more weightage because these are the points that actually contribute more towards constructing this lines and these are the points which contributed less in constructing this point. But when I am minimizing it by taking summation even though this point is much away from the line and contributed very less, when I am squaring, it is getting more significance and more importance. So, while u_5 actually should get less weightage, because of our mechanism of OLS u5 hat square has actually more weightage in this mechanism. So, that is some kind of limitation of this ordinary least square method that we have to keep in mind. In ordinary least square method all these predicted error squares get equal weightage while we should put more importance to the point which are closer to the line and less to the error term which are far away to the line because those points contributed very less in the construction of the line because this line is nothing but an average of your scatter plot.

If, you plot your x and y you will get a scatter plot like this and this line which is srf is actually the average representation of the scatter plot and in that scatter plot the point which are closer towards the line they should receive more importance and that is why lesser the error more should be the weightage and higher the error less should be weightage but here it is happening just opposite when I am squaring, more the error larger the weightage is the importance they are giving. The limitation of this OLS method will be discussed in detail in later part of our discussion when we discuss about weighted least square method. Limitation of OLS should be overcome by the weighted least square method where we see that larger the deviation from the line, those errors will receive lower weightage. This implies that the error terms errors which are far away from the srf get more importance. This is the limitation of the OLS method. We will overcome this using our weighted least square method.

So far we have learned what is basically the sample regression function, population regression function, we are estimating $\hat{\alpha}$ and $\hat{\beta}$ to predict about or to infer something about the true population parameter α and β and that we are doing this using ordinary least square method where we are just trying to minimize the sum of $\hat{u}_i^2$ in that process.

Since we are putting equal weightage, we are committing some mistakes because larger the deviation from the srf, we are giving higher importance because if we are squaring the errors which are already larger in their magnitude getting magnified which ideally should not be the case. So, with this we are closing our discussion today and tomorrow we will discuss something about some important properties of your estimates $\hat{\beta}$ and we will also take one data set and we will see how to use the software statistical software to estimate the model that would become very interesting.

Thank you.