**Econometric Modelling**

**Prof. Rudra P. Pradhan**

**Department of Management**

**Indian Institute of Technology, Kharagpur**

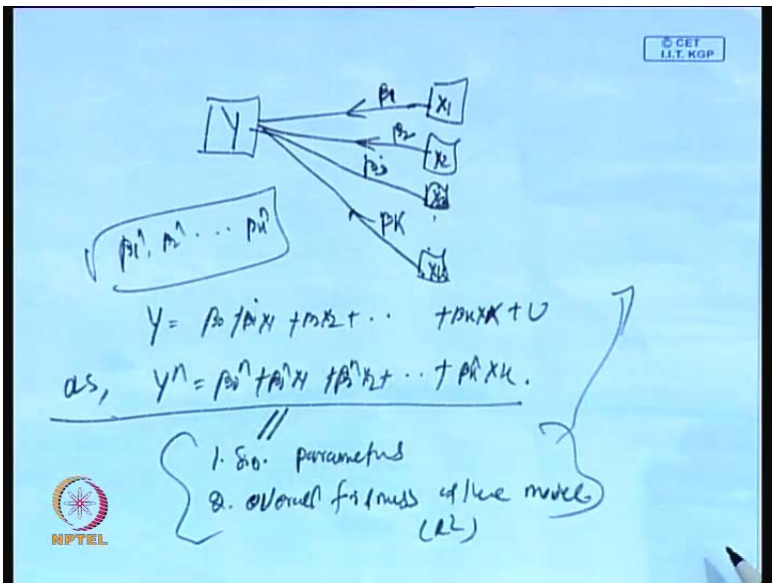
**Lecture No. # 21**

**Multicollinearity Problem (Contd.)**

Good evening, this is doctor Pradhan here, welcome to NPTEL project on econometric modelling. So, today we will continue the same problem multicollinearity issue. So, in the last lectures, I have discussed or you can say, I have highlighted briefly the structural multicollinearity, how it comes to the econometric modelling issue? And what are the steps? And you know problem altogether, and how you have to go for its solutions.

So, let me start with a highlighting once again this particular problem. So, in the last class, I mentioned very clearly that multicollinearity is a multivariate issue. The problem will be there, when the system about you can three variables at a times or more than three variables. If the system consists of only two variables, then multicollinearity may not be a means, it will not come at all. So that means, it will start only either in the trivariate case or you can say multivariate case.

(Refer Slide Time: 01:32)



So now we like to see, what is all about the structure of multicollinearity? So since, it is a multivariate problem so, we will first write what is multivariate problem all about? So, multivariate is a system where there is one dependent structures and series of independent structures. We have mentioned in the last class, the series of independent is recognized as X 1, X 2 upto X K. So, the model setup is like this.

So that means, the modelling behavior will be like this. So, this is supported by a beta 1 beta 1 coefficient, this is supported by beta 2 coefficient, this is supported by beta K coefficient. For multivariate problems so, we may not have a serious issue or we should not think much about this beta zero coefficient that is you know intercept terms that is supporting component. But our major concern is on all beta coefficients, that is beta 1, beta 2 up to beta K.

So, that means the coefficient of independent variables, direct independent variables not the intercept terms. So, that is why I have not mentioned here anything about beta 0. So, we have we have to target the beta 1 hats, beta 2 hats up to beta K hat is our issue. Now what is all about this multicollinearity? There are many ways we can define the term multicollinearity, but the simplest way to define multicollinearity is that it is having the linear relationship among the regressor or independent variables.

So, that means it is a problem having the linear relationship among the regressors. So, the linear relationship among the regressor that is the core ideology or fundamental issue of multicollinearity; the linear you know relationship among the regressor not non linear. So, we are discussing the linearity issue of regressors. So, if there is a any such relationship then obviously this is a issue of multicollinearity, that means the game all together is like this.

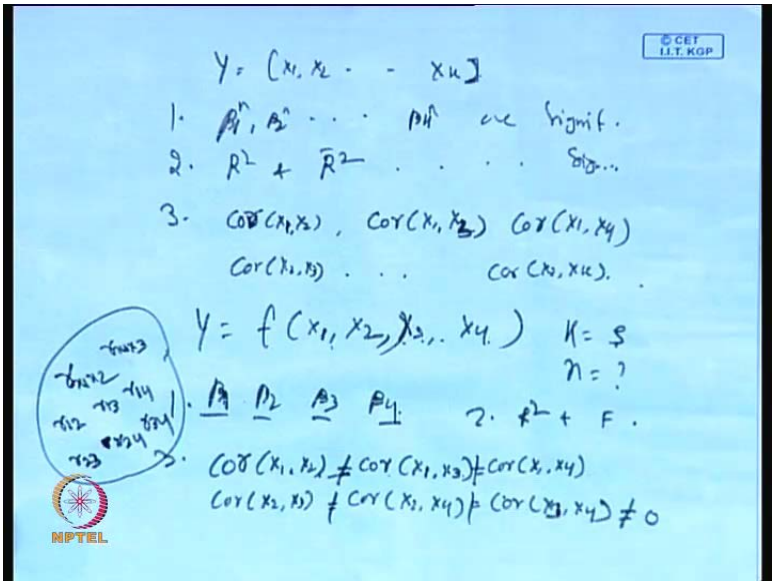
So, that means obviously by default X 1 we are expecting that X 1 has a influence on Y X 2 has a influence on Y X K has a influence on I .So, this is little bit interesting so, I will put two another variables in the system. Then we will call it beta 3 hats or beta 3 coefficient. So, now you know what is our usual procedure? Usual procedure is to set the regression models .So, beta 0 plus beta 1 X 1 plus beta 2 X 2 continue up to beta K X K plus U. So, obviously by various techniques, we have the estimated models beta Y hat equal to beta 0 hat plus beta 1 hat X 1, plus beta 2 hat X 2, plus beta K hat X K. So, this is how the estimated model we have received. Now, you see the moment you will have

this estimated model then obvious standard issue is that earlier we are checking the significance of the parameters and the significance of the overall fitness of the models.

So, that means significance of parameter that too beta 1, beta 2 up to beta K and significance of the overall fitness of the model that is R square, adjusted R square. Then you can say followed by f statistic etcetera etcetera. So, that means it includes the s s r s a hat t s s. So, these are the components which will take care the overall fitness issue and here the beta 0 hat, then variance of beta 0 variance of beta 0 hat, beta 1 hat, variance of beta 1 hat, beta 2 hat, variance of beta 2 hat and standard error of beta 1 hat, standard error of beta 2 hat. Then, obviously t of beta 1 hat, t of beta 2 hat. These are the components, which will take care the significance of the parameters.

So, right now what we have discussed is the significance of the parameters. And, second is the overall fitness of the models. It is the overall overall fitness of the fitness of the overall fitness of the model. So, that is the overall fitness of the model that too you know R square so adjusted R square f statistic etcetera. But suppose multicollinearity is concerned, so far as multicollinearity is concerned, we have to check the linear relationship among the regressors.

(Refer Slide Time: 06:24)




So, now that means Y is a function of X 1, X 2 up to X K. So, that means we have to check whether all these beta 1 hat, beta 2 hat, and you know beta K hat are significant. So, then R squares and followed by adjusted R square has to be statistically high and

significant. This is the one issue, this is second issue. Till now we are much concerned about these two issues. But now so far as a so far as a multicollinearity issue is concerned so, I have mentioned what is all about this multicollinearity; multicollinearity is having the linear relationship among the regressors.

So, now we have K number of regressor, so that means X 1 X 2 X 3, up to X K. So, linear relationship means then X 1 upon X 2, X 1 upon X 3, X 1 upon X4, X 1 upon X K. Then again X 2 upon X 3, X 2 upon X 4, X 2 upon X 5, or X 2 upon X K. Similarly, there may be again X 3 upon X 2, X 3 upon X 1 then, X 3 upon X 4, X 3 upon X 5, then X 3 upon X K like these there are various structures you will find.

So, we like to know whether this association is there or not? If it is not there then you are in the right track, if it is there then you are in the wrong track, you have to come to the right track. So, that is how we need to have solution provided your objective must be very, very specific. So, that means third interesting thing we like to observe for this particular multicollinearity problem is that. So, covariance of X 1 upon X 2 or correlation or covariance between X 1 upon X 2. Then covariance correlation between X 1, X 2, X 1, X 3 then correlation between X 1, X 4 like this then, correlation between X 2, X 3 so, it will continue correlation of X 2 X K. So this is how we have to first catch all these reports. That means we have to apply the permutation combination you get to know how many such relationship we can have in this particular system, if your system is all about K number of variables.

So, now for the simplicity we will restrict our model to only 3 variables, so that we can have the accurate discussion. So, otherwise it will have K number of variables, then the problem is very, very complex very, very you know pathetic issue all about. So, now what you have to do? So for the simplicity, we will take Y equal to function of X 1 and X 2. So, let us assume that this is the models we have right now in the system.

Now, the problem is very, very simple and very, very easy to detect so far as a multicollinearity concerned, because here only such relationship is with respect to X 1 and X 2. So, in fact this is also not accurate to detect the multicollinearity. Means, most of the components we cannot observe from this particular issue. Let us take help with another 2 variables. So, let us say X 3 and X 4.

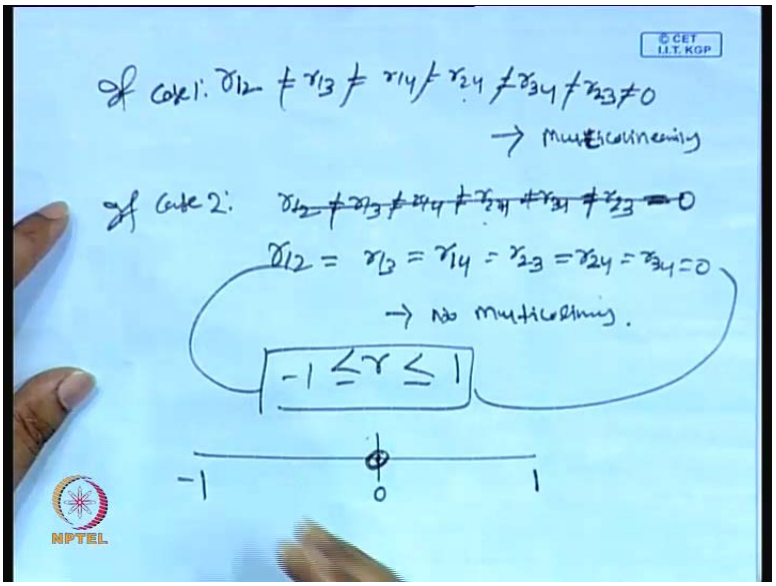
Now the model is all about where K equal to 5 sorry K equal to number of variables in the system so, independent variables K equal to 5 dependent variables and n equal to forget something something. So Y is one dependent variable and X 1, X 2, X 3, X 4 is 4 independent variables all together there are 5 variables in the systems.

So now, as usual the coefficients are beta 1, beta 2, beta 3 then beta 4 forget about beta 0 coefficient, beta 0 is a intercept concept. So we are not bothering much about this one right now, for this multicollinearity problem .So now what we have to do here is so, we like to highlight the multicollinearity issue. That means, this is 1 and second is the R square issue and F statistic issue. Third problem is here is so, we like to know means in this particular jargon. So we have correlation between X 1 and X 2, then correlation between X 1, X 3, then correlation between X 1, X 4, then correlation between X 2, X 3, then correlation between X 2, X 4, then finally correlation between you know X 1, X 4, X 1, X 4 is already there so X 3, X 4.

So, these are the six possibilities that are there for this particular setup. So that means if your game boundary is with respect to Y and for independent variable, then so far as multicollinearity is concerned, then we have an additional problem. We have to observe or we have to investigate is that having correlation among the regressor. So, that means in this particular setup we like to know, what is the correlation between X 1 and X 2 X 1 and X 3 X 1 and X 4 X 2, X 3, X 2, X 4 and X 3, X 4(s); by any chance if all these values are equal and is equal to 0 then there is no such multicollinearity issue.

So, this problem is absolutely free from multicollinearity. However, if all are may equal, it may not equal obviously. So, if it is not equal and not equal to 0 then obviously it is a serious problem. But you remember one thing so we know correlation coefficient is denoted as r for correlation X 1, X 2 you will call it X 1, X 2. Then this is correlation X 1, X 3 we will call it r X 1, X 3. So, otherwise you can call it r 1 2, r 1 3, r 1 4 then r 2 1, r 2 3, r 21 or r 1 2 same because it is symmetric in nature; so, 2 3 r, r 2 4 then r 3 4. So, these are the possible correlation coefficients.

(Refer Slide Time: 13:32)




So, by any chance by any chance if you have like r 1 2 equal to r 1 3 equal to r 1 4 equal to r 2 4 equal to r 3 4 (( )) 1 2, 1 3, 1 4, r 2 4, r 3 4. So, these are the possible solution r 1 2, 1 3 then 1 4, 2 4, 3 4 and in fact 2 3 is there. So, there is another item r 2 3. If these are these are these are all not equal to 0; if it is 0, then no problem. If it is not equal to 0, then there is multicollinearity. If case 1, if case 2 if r 1 2 equal to r 1 3 equal to r 1 4 equal to r 2 4 equal to r 3 4 is equal to r not equal to not equal to not equal to not equal to not equal to 2 3 is equal to say 0 then, obviously obviously it should be equal to 0.

Then we can put like this r 1 2 equal to r 1 3, r equal to 1 4, r equal to 2 3, r equal to 2 4, r equal to 3 4 is equal to 0. Instead of writing this one, it is better to write this one if that is the case then it is no multicollinearity. So, that means the moment you have a system where Y is a dependent variable X 1, X 2 ,X 3, X 4 are independent variables. So, as usual you have to go for specification test by judging the significance of the parameters beta 1, beta 2, beta 3, beta 4 and in the other sides, we have to judging the overall fitness of the model that is R square, adjusted R square F statistic followed by s s t s s and r s s etcetera.

So, now in the mean time so, for as a multicollinearity is concerns if your objective is to look into multicollinearity issue, then obviously we have to see what is the correlation among this regression? Then these are the possible possible correlations, and if it is equal to 0, then obviously there is no multicollinearity .So, you have to assume that whatever

model we have received then, that model is a absolutely best fitted and it can be used for forecasting and policy use.

But remember when these are all equal to 0, then by default then most of the parameters will be significant and your r square will be very high and f will be also highly significant. But if this particular items are not equal to 0, if we will go by case 1 if all are not equal to 0 then there is a multicollinearity it is a serious problem. Even if you know r 2 r r 2 r 1 3 r 1 4 r 2 4 r 2 3 and r 3 4 all are equal but not equal to 0 then obviously it is also problem of multicollinearity.

That means there may be possibility that r 1 2 equal to r 1 3 or equal to r 1 4 so, it may be equal but the value cannot be equal to 0. If cannot equal to 0 then it is called as a multicollinearity issue. If it is equal to 0, then there is no such multicollinearity issue. So that means what is all about multicollinearity? Multicollinearity is having the existence of linear relationship among the regressor. It is the degree of association between two variables, it does not matter whether they are negatively correlated or positively correlated. So that means, so we like to know what is the degree only.

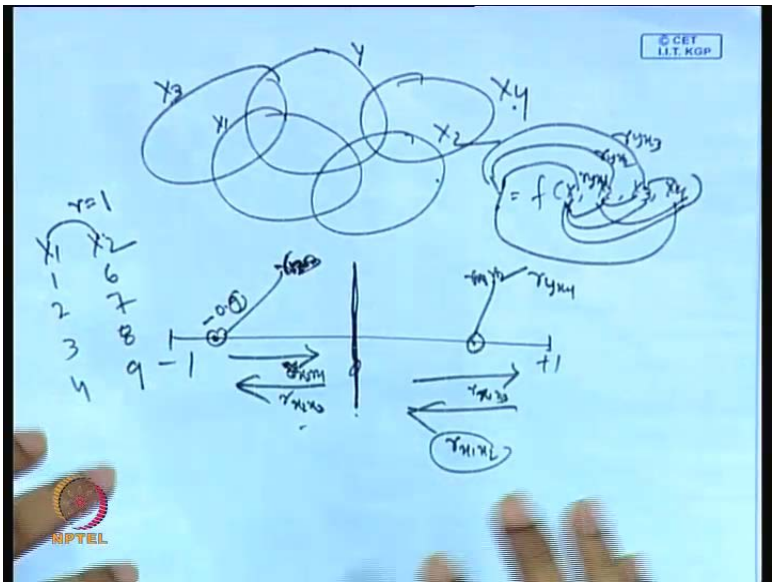
So, that means r always lies between minus 1 to the plus 1 so, this is the standard tricks how you have to integrate with the multicollinearity issue. So, this is the standard tricks we have to issue in the multicollinearity angles. So, that means when r is simply as represented as a correlation coefficient.

But I have not specifically highlighted any subscript here, because it is generalized. If it is with respect to first and second variable then put 1 2, if it is first and third variable you put 1 3, if it is second and fourth put 2 4, like this you have to abbreviate and you have to interpret accordingly.

Now, in any case r is considered as a correlation coefficient. It should be in between minus 1 to 1. So, that means if I will take a range here, then this is 0, this is 1, this is minus 1; so now, if in a particular regression, whatever correlation coefficients can possible, that too among the regressors only. So every relationship and every correlations should be tends to this particular circle 0, then obviously there is no such multicollinearity. But if any other point if you find there in tracking. So then there is a multicollinearity, whether it is a left side or whether it is a right side. So, the nature of relationship is not important for multicollinearity issue. It is the degree of association is

very important, if the degree is high then it is a high multicollinearity issue, if the degree is low it is a low multicollinearity issue. So, how extent, and what context we have to look these methods or you have to consider this problem, so that we have to take care here.

(Refer Slide Time: 19:46)




Now, let us see here so, our game plan is like this. So, this is Y and this is say X 1, this is X 2 and this is X 3 and this you can call it as X 4. That means, obviously the variables which you have taken X 1, this is X 2, this is X 4 the variable which have taken then obviously by default there is some degree of association with Y. So it should be otherwise, you cannot go for econometric modelling.

That means, so you has a function of X 1, X 2, X 3, X 4. So that means there should be some connection with some degree of association between Y X 1, Y X 2, Y X 3 then Y X 4. So, this can be as usual there should be so, that means in that case you have to put r y x 4 so this is r y x 3, then this is r y x 2, this is r y x 1. So this is how you have to abbreviate but in this context this should not be equal to 0. If this should be equal to 0 then that variable cannot be in the system. So, this is how the system is all about the econometric modelling.

So, that means there is a some relationship; correlation must be dependent to independent variable. But correlation should not be among the independent variables. So,

that means X 1 X 2 and X 1 X 3, X 1 X 4 or X 2 X 3 or X 2 X 4 or X 3 X 4 these are not our concern it should not be there.

So that means the variables which included in the system should be completely independent. That is the fundamental issue or fundamental agenda of all econometric modeling. So, the variable which we have chosen or which we have included in the systems that too independent variables, should be totally independent to each other. Then, the clustering or the influence of independent variable to dependent variable will be very accurate, very perfect, very systematic and very feasible.

So, that means we like to know this particular setup how ease is the entire system? This is little bit you know very complicated pictures, because why I am saying multicollinearity has a 2 dimensional issues, one dimensional issue is the nature of relationship, another dimensional issue is the degree of association. So far as nature is concerned either it is negative or positive so, that is not our concerned. So we should not bother about to having the negative association, whether r is negative or r is positive.

So, we like to know, what is the value of r? So, is it close to minus 1 or is it close to plus 1 or is it close to 0 or is it close to again 0 that means, the trend is like this. So if it is 0 here this plus 1 and this is minus one. So, that means whether the trend is this side or whether this trend is this side, whether the trend is this side or whether this trend is this side so these are all correlation coefficients among the regressor.

So x 1 x 2 then you say x 1 x 3, this is x 2 x 3, this is say r x 3 x 4 like this these are all. So you first find out the correlation coefficient. Let us say correlation coefficient between x 1 and x 2 is coming 0.7 so, you will track here. So, this particular component we will call it r x 1 x 2.

So, let us say we will get x 2 x 3 is minus 0.9 so these particular (( )) This is called as r 2 3 so, this is correlation coefficient between 2 and 3 so, negative related at the higher level. So, we are not bothering whether you are here or whether you are here. This is the this is the you know initial setup, whether you are in this direction or whether you are in this direction that is not our issue. Our issue is what is this position? Where is that position? Whether it is close to minus 1 or whether it is close to plus 1 or whether it is close to 0 or whether it is close to 0?

That means so, first part of the problem that is nature of the relationship is not so important but the degree of association is very important. So far as degree of association is concerned, obviously you know the degree will starts from 0 to plus 1 and 0 to minus 1. So you will go 0 by minus 1 and you will go by 0 to plus 1. Now, the moment it will increase this side or increase this side, then the multicollinearity problem will be more and more and more dangerous. So, that means the virus will be aggravate accordingly.

So, the degree of influence for instance, boundary is like this so it will effect first file, second file, third file. So, it will automatically highlight so, if it is very small level then the problem is not such issue. That means, let us say in a folder there are 100 files, if it is affecting only 1 file then obviously it is not serious problem. But if it is affecting 99 files then obviously it will be serious problem. That is how you have to look in to this multicollinearity issue.

So, now it is the degree which brings the issue of multicollinearity. Now, so far as degree is concerned so the degree may be 0, if it is 0 then no multicollinearity, if degree is very very low then it is very very minor multicollinearity. If the degree is low, then obviously multicollinearity is also low, if the degree is very high then obviously multicollinearity is very high. If the degree is very very high, then obviously multicollinearity is very very high. If the degree is extremely extremely high what is mean by extremely high? That means it is exclusively equal to plus 1 or exclusively equal to minus 1, then it is very very serious or extremely serious problem for multicollinearity issue.

That means, this type of situation model you should not consider that model at all for forecasting or policy use. So, this means close to one(s) means two variables are same that means, we have to drop one variable. For instance, I will take a series here. So, let us say X 1 and X 2, so I will put 1, 2, 3, 4 only. Then by default I will put here 6, then 7, then 8, then I will put here 9. So, this is one series and this is another series. So I will like to correlate it. Then by the instance, I will get I think positive correlations it is a positive correlations then that to 1 only.

Because this X 1 and X 2 are linearly dependent. X 2 is the item plus 5 for every case item plus 5, 1 plus 5, 2 plus 5, 3 plus 5, 5 plus 5. There are many ways such relationship can be. The relationship can be additive, may be multiplicative, may be divisions, may be something something. So in that case it will give you signal the linear dependency.

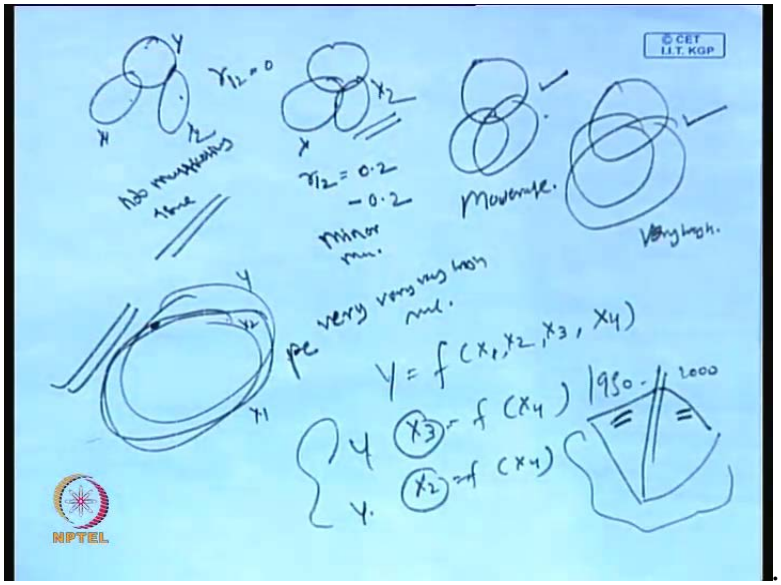
So that means the linearly independent variables should be considered in to the system and that system will be more accurate, more appropriate, more practical; but if the variables are linearly dependent, l d then obviously the system is very inconsistent, practically not valid and it is totally infeasible in (( )). So, you have to bring in to the feasible regions.

So, how do you go for that particular structure altogether. So we like to know what this shape of this multicollinearity issue is. Since the degree starts from 0 to extremely high, then obviously multicollinearity will be extremely high. So how does it look like? So, that is our main concern let us see here. So let us say, for knowing the complexity of that particular problem with respect to low degree to high degree. Then, we consider this system where there is a Y variable and X 1 X 2 are two independent variables, that is the basic statistic point of multicollinearity. So if it is less than and that then there is no such multicollinearity you see very interesting.

So, when will we go for econometric modeling, then you start with a bivariate system. The econometric modelling is useless or cannot be possible. When the system is only one variable and having no classification and the root of the system or starting point of the system is that it must have one dependent variable and one independent variable.

But in that particular context multicollinearity is not at all pictures, it is not at all coming there. Now, you add another variable Y X 1 and X 2 then multicollinearity will start adding in the process. So, if you will add another variable then it will start another complexity, if you add another variable then it will start another complexity of multicollinearity. So this is how the problem is extended extended and extended.

(Refer Slide Time: 29:37)




So, what you have to do? so you have to solve the degree of association let us say Y equal to here dependent variables. So this is Y and this is X 1 and this is X 2. So, in this case r 1 2 is exactly equal to 0 so, that means there is no connection between connection between X 1 and X 2. So in that context in that context this model is perfectly accurate, perfectly fit for forecasting and policy use.

Now, I will take another models say Y u Y and this is X 1, and this is this is X 2. So that means r 1 2 there is a relationship and if I do not know whether positive or negative because we are not bothering about degree of association, nature of the relationship not the degree of (( )).

So, if that is the case let us say it is 0.2 only, 0.2 or minus 0.2. So in that case it is called as; this is no multicollinearity issue. I can call it minor multicollinearity issue. So similarly, there is another issue here. So, this is called as a moderate multicollinearity issue. So, I will take another case here. So, this is very high multicollinearity very high multicollinearity issue. So I will take another case. So this is called as a a perfectly multicollinearity issue.

In fact we cannot say perfectly because there is something missing here, perfectly means it is the 2 circles which merge each other equally. So in that case this is X 1 and this is X 2 this is Y. So in that case it is a very very very very high multicollinearity. So this is very serious issue for econometric problem, this is also serious issue problem, this is in fact also small issue of multicollinearity, this is very minor issue, this is no issue at all.

So this is how the shape of the multicollinearity is all about. So, now we get to know what is the; how do we define the term multicollinearity? What is the nature of multicollinearity? And now we are going to discuss why multicollinearity occurs in the econometric modelling or in statistical process. So, the reason is that there are many reasons.

So, one standard reason is that most of the variables in the environment are very interdependent. It is very difficult to find 2 variables which are completely independent. If it is completely independent, yes you will find there are variables which are completely independent. But you know when you are targeting the causality issue then obviously we are looking for some interdependence, that too dependent to independent only. But if you know within x s if some X is function of another x.

So, that means that itself is regression analysis all together. Means I am putting here is Y equal to function of X 1, X 2, X 3 up to you know say X 4. Then I will find X 3 is a function of X 4 or you will find X 2 is a function of X 4. So, that means X 3 can be considered as a another Y, X 2 is considered as a another Y because, it is all together dependent side. So that means it is the structure or understanding only.

So the cause is that since most of the variables are interdependent in natures. So by default or by natural process multicollinearity is always be there in the econometric model. But the thing is how whether you will go ahead with multicollinearity or you have to get its solutions. You see that is depends upon your objectives specification, if your objective is to go for forecasting or predictability then, obviously in that case high R square having low significance errors. Low significance of the parameters can be considered as the base models.

So that means, if you go for forecasting or prediction issue that times high R square will give you better prediction and better forecasting. But when you will go for reliability of the reliability of the modelling that times you need to have both significance of the parameters, exclusively highly significant and the high value of R square and exclusively highly statistically significant for F statistics.

So, this is how the objective of reliability is concerned. But if your objective is prediction and forecasting, then obviously multicollinearity is not such serious problem. But reliability case it is a very serious problem. But if will we go together then obviously

minor multicollinearity or slightly moderate multicollinearity is not a serious issue you have to go ahead with, because it is very difficult to find a system, where all the independent variables are totally independents. There may be some kind of interconnections. So it is very difficult so that is why, by default there will be some relationship but that relation should be very at the minor level. Otherwise, if it is a high level what is the point to take or to consider two variables differently so, it has to be one variable only.

For instance, in economic environment or you can say a management environment we have g d p and per capita g d p. So g d p is considered sometimes as a variable and per capita g d p sometime considered as a variable. Both have a different interpretation, different structures. But sometimes there impact more or less in fact may be some way contradictory or you can say other way round.

For instance, you may find some kind of relationship between g d p and per capita g d p because per capita g d p is derived from g d p. If per capita g d p is derived from the g d p then obviously g d p is derived from per capita g d p because, per capita g d p is nothing but g d p by population. So, this is how means I am just highlighting one examples how multicollinearity can be serious issue.

For instance, another case you take what is the impact of f d i on stock price and another side I will take another variable say x 1 which is represented as a f d i as a percentage of g d p on stock price. So f d i is one part of 1 variables and f d i as a percentage of g d p is another variable because, the game plan is completely different. F d i there is variables and f d i by g d p then the game plan is completely different but somehow if we integrate f d i to f d i as a percentage g d p means in a particular system if I will use stock price as a dependent variable and f d i as a percentage of g d p is another two independent variable. Then I definitely find there is some linearity between f d i and this f d i as a percentage of g d p.

So, that means in that system will be perfectly. So, means if you use only f d i or you can use f d i as a percentage of g d p either one you can use then the model will be perfectly ok. But if you go with 2 variables then obviously the complexity will start and it will be serious multicollinearity, because for that at beginning I have mentioned that your theoretical knowledge must be very high or external knowledge must be very high with

respect to various theories and various statistical knowledge, mathematical knowledge, then you will enjoy this econometric modelling.

Otherwise it is very contradictory this type of issue if you have no such theoretical knowledge or outside knowledge then you cannot detect particular variable whether they are correlated. Because we know the definitions and we know the setup that is why we are saying. For instance, if there is two variables say I will call money supply and I will use the term m 3 for India point of view.

So money supply m 3 is money supply because money supply has different groups. So m 3 is one of the such groups so, if you will take money supply and forget about data you can artificially generate data. But if you put money supply and this m 3 then obviously there will be some kind of multicollinearity either you use or instance, if you will go by money supply then obviously you will find there is some datas with respect to m 2 and some data with respect to m 3. But both will come under money supply.

So, now if I am fitting any models with respect to m 2 and m 3 then, obviously there is a some connection between m 2 and m 3. Either, you will go by m 2 or separately go with y with m 2 and y with m 3 separately and I will check it which one is the best for the best fitteds. So, then accordingly I will consider that variables. Otherwise, if we will go together with y m 2 and m 3 then obviously there is a serious multicollinearity. This means this is one of the strong cause for issuing multicollinearity problems.

So, another problem is that introducing lag in the systems. So we will discuss detail when we will go for time series modelling so, when we will introduce large in the system then by default it will generate multicollinearity, which I have highlighted very clearly in the last class. So, then mathematical manipulation, mathematical formulation also creates multicollinearity problems. So, you must be very, very perfect to go for mathematical formulation of this particular problem. So, then inclusion of a unnecessary variables in the system will also have a multicollinearity problem.

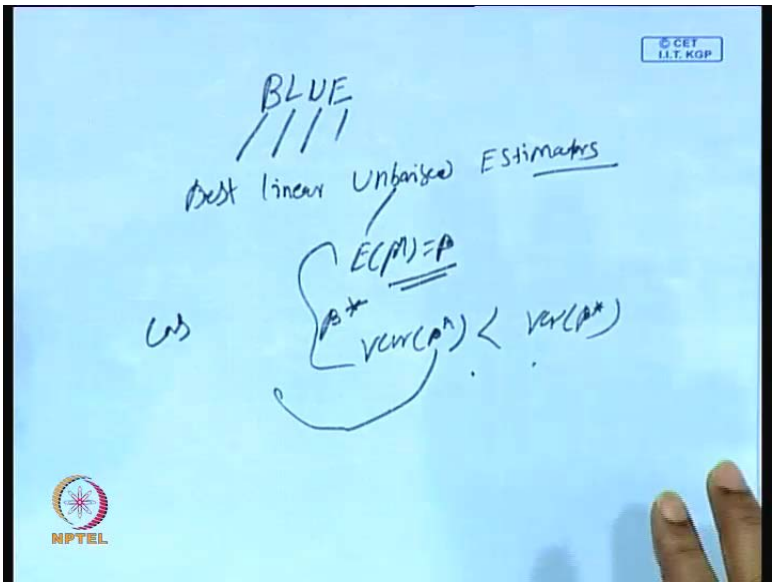
So, again with respect to sample size a lower or high can also multicollinearity problem. Econometric modelling is sometimes it is a fussy game for instance; take a case of cause of multicollinearity with respect to size of samples. If the size of sample then sometimes there may be some collinearity. But later stage these two variable will still diverge themselves then obviously there is no such multicollinearity.

But, there is another interesting problem. First they will diverge, then after certain point of time they will converge. So, that means if you will take total time length then you will find there is some multicollinearity. But if you will see if you will classify this structure in to 2 different sets for instance, for a particular point of say 950 to you can say 2000.

For instance, I have 950 to 2000 data point. Now, first they diverge like this they diverge like this, then after particular point of time they will converge. So that means you make the division here. You take this one, one side and this one another side, then obviously you will find there is no such multicollinearity. But if you will you take total sample periods then obviously there is a serious multicollinearity problem.

So, that means sample size itself will make the game interesting or it can also generate problems. So, you must be very careful how you have to choose this sample size. Whether you have to go for high sample size or whether you will go for low sample size or whether you will go for optimum sample size. So, it is depending upon the structure how you have to coming with their pictures this is how the causes of multicollinearity. Then what are the consequences of multicollinearity?

(Refer Slide Time: 42:30)




So consequence is that we have already discussed when we will go for overall fitness of the means goodness fit of the models are best fitted models. Then obviously we will go by significance of the parameters. But the parameters should follow some principles which we call as the blue best linear unbiased estimator. Until and unless you receive

these properties that means, the estimated parameters which you have received in the systems should be unbiased, should have minimum variance and should be very, very consistent. What is unbiasedness means, the different biased should be equal to 0 that is the difference between expected value and true value. Unbiased means e beta hat should be equal to beta.

So, now we will take another parameter say beta star hat. If variance of beta variance of beta hat is less than to variance of beta star hat then, beta is considered as the best parameter than the beta star hat. Similarly, consistent this is called as a minimum variance property. This is unbiased property similarly, if you will together if you (( )) together then it is called as a consistency property.

That means when m stands to infinity then, that means if you will increase sample size then there is enough chance that expected value of particular parameter should be equal to true value that means for this instance e beta hat equal to beta and in the second instance variance of beta hat should be less than to variance of another beta hat. And variance of beta hat should be close to 0 when n stands to infinity that is what consistency property.

If all these properties are maintained then estimated model can be considered as the best fitted model. But if all these properties are going against means if it is in not unbiased, if it is not biased, if it is not having minimum variance, if it is not consistent. Then obviously the model has to redesign and one such problem is multicollinearity. If, multicollinearity is there then it will affect the unbiasedness property, it would affect the minimum variance property, consistency property and so many other things are there the estimators should be linear in parameters etcetera.

So, this is how you have to observe. So that means so far as a consequence is that involvement of multicollinearity, or the presence of multicollinearity will lead to make the system deviation from the blue. So, it will take diverge from the blue components so, as a result if it is diverging from the blue then obviously the model cannot be considered as the best. So you have to find out the solution how they will converge to the blue? That means you have to choose a system where all these parameters should follow the blue property, the best unbiased estimators. Then you can use for forecasting and policy use.

So, this is how the major consequence of the multicollinearity problem so then, next is how to detect? so far as the detection is concerned. I have already mentioned the way we have discuss one of the standard tricks how to detect the multicollinearity issue correlated among the regressor. Suppose there are 5 variables so accordingly you have to find out x 1 x 2, x 1 x 3, x 1 x 4, x 1 x 5.

Similarly, x 2 x 3, x 2 x 4, x 2 x 5, x 3 x 4, x 3 x 5, x 4 x 5 these are the possible correlations then you check whether all these correlations are 0 or not, if it is 0 then it is free from multicollinearity. If few are 0 and few are not 0 then obviously there is a some kind of multicollinearity is there.

So, if some kind of multicollinearity is there then you have to check what is the degree of that multicollinearity or degree of association. If it is high, then it has a strong negative impact best fit of the model and if it has a minor degree is very minors or very low; then obviously the impact is very less on that feasibility of the models.

So this is how we have to detect the structure of multicollinearity. This is one of the standard detections, standard measure how you have to detect the multicollinearity. As a statistician or econometricians there is a lots of various clues how you have to detect the multicollinearity. For a standard reason is that very beginning in my first lecture of this multicollinearity I mentioned that.

So we look for two specific objectives, that is significance of the parameters and significance of the overall fitness of the model. Now, for significance of the parameters, we are looking for you know our objective must be all parameters should be highly significant R square should be high and F should be statistically or highly significant. But that is the requirement but ultimately we have 2 different other options, that is few parameters may be significant and others are not significant and R square will be high and F will be very high this is one situation.

And another situation having low R square and having few parameters or statistical significant; then it is also a problem of multicollinearity. So in that context what you have to do? You have to go for lots of permutation combinations and you have to find out whether there is a multicollinearity in what extent the multicollinearity is all about.

So, this is how you have to detect the multicollinearity issue. Once you know very standard thing is that if your parameters, all parameters are not statistical significant and R square is coming high then this is one of the inspection method how you have to observe or how you have to expect that there should be multicollinearity.

But you have to go again every time by detection criteria how to detect the multicollinearity. In fact, we sometimes econometricians sometimes use the partial correlation coefficient is the detection measures you know if high multiple correlation and low partial correlation coefficient. Then it will also indicative means indicates you that there is a problem on multicollinearity. This is how you have to go for detection criteria.

So now so far as solution is concerned there are number of solution you can find out these multicollinearity issues. And one of the standard solution is either you go for optimum sample size or if still there is multicollinearity problem so, either you increase sample or decrease sample till you get the models which is free from multicollinearity is one of the standard trick how you have to go for solution.

And second thing is, you check whether the variables are closely related by inspection or by theoretical knowledge, if you know that few variables or two variables are closely related to each other it is better to drop that variables. Then with rest of the variables there is enough chance that model will be best fitted where all parameters will be statistically significant and overall fitness will be very high.

So, this is another way how you have to get the solution. Besides, standard there are so many tricks how you have to get the solutions with respect to the current cause then solutions will be multiples in nature. But you one standard criteria is that so you have to first point out why this problem usually happens and accordingly you have to apply the solution criteria. So, that is how you have to solve these multicollinearity problems.

Now, sometimes there are different techniques if will we apply the problem of multicollinearity can be solved. I have mentioned that econometric is of very fussy game. Sometimes, by the use of different techniques the multicollinearity problem may be automatically solved, because the technique itself will take care of this multicollinearity issue. This is very interesting and in fact today's world too we are in process of various softwares so this is software (( )) So if you apply different softwares with different

application, different techniques then the problem of multicollinearity will be automatically solved and you will get the best fitted models.

So, now before ending this particular class so, I will like to highlight two things here. So far as multicollinearity is concerned you start with your objective specification, if your objective is predictions only, then obviously multicollinearity is not a serious issue. But if your objective is reliability issue then obviously multicollinearity is a serious problem and you have to first solve the multicollinearity then you have to use for forecasting. If that is not the case, then obviously it will go other way around. With this we will conclude this particular session thank you very much have a nice day.