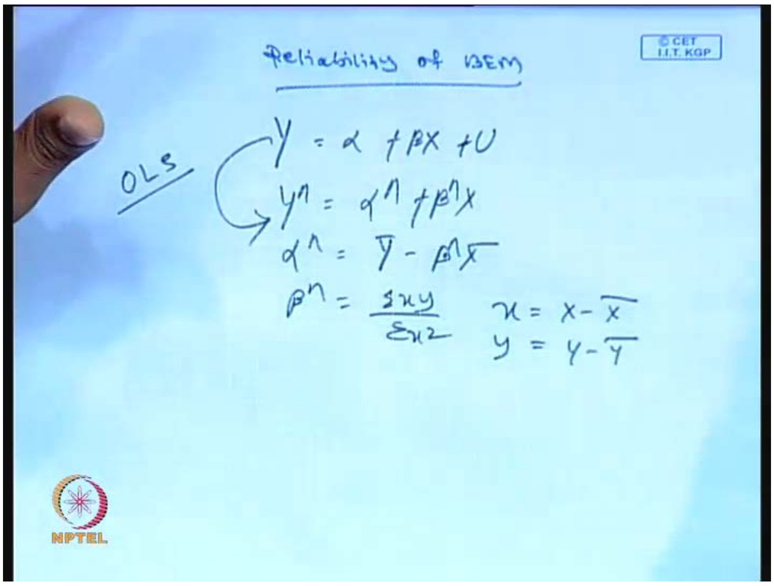
Econometric Modelling
Prof. Rudra P. Pradhan
Department of Management
Indian Institute of Technology, Kharagpur

Module No. # 01
Lecture No. # 09
Reliability BEM

Good morning. This is Rudra Pradhan here. Welcome to NPTEL project on Econometric Modelling. Today, we will discuss the reliability of bivariate econometric modelling. So, in the last class, we have discussed the entire setup of bivariate econometric modelling where the problem boundaries with respect to two variables. So, we we we we have discussed the details about the estimation of alpha hat and beta hats. So, after getting the estimated value of alpha hat and beta hats, so then, our problem setup is completely different, so far, as forecasting or policy implication is concerned.

(Refer Slide Time: 01:03)

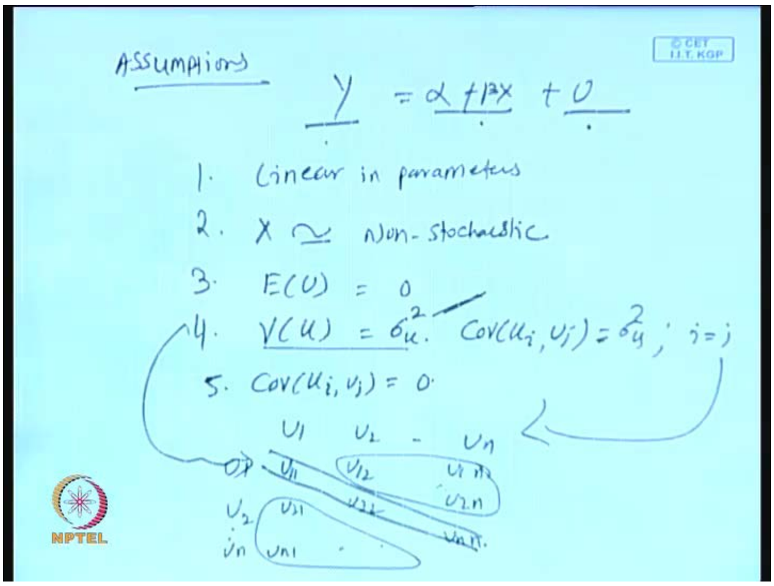


So, let me highlight what is a, what is the last class discussion. So, for bivariate models with respect to Y and X, Y equal to alpha plus beta X and we have to introduced the error terms U, and the estimated model equal to Y, Y hat equal to alpha hat plus beta hat X. And where alpha hat equal to Y minus Y minus Y bar minus beta hat X bar and beta hat equal to summation X Y by summation X squares, where X equal to X minus X bar and Y equal to Y minus Y bar. This was our last class discussions.

But, you know the moment from the first equation to second equation, where the estimated models, we have describes Y hat equal to alpha hat plus between hat x, so, we have applied the technique called as a OLS techniques. Of course, there are several techniques like ordinary least square methods, general generalized least square method, weighted least square methods and maximum likelihood estimator. So, here we start with OLS technique, that is ordinary least square method, because it is a it is very simple technique, easy to understand and easy to apply; that is how, we we start with OLS technique. Then, we subsequently move to GLS techniques, OLS techniques and MLE technique where, we need to have, because some of the problems, it is very difficult to handle through OLS techniques. So, that is how, we have to apply the other techniques like GLS and WLS and maximum likelihood estimator.

But, in the last class, we have little bit problem regarding the OLS restrictions. So, before you applying OLS technique then obviously, we will get the estimated model. But, the application of OLS is based on certain assumptions. So, before we check the reliability part of the models, so, let me first highlight the assumption related to OLS technique because which last class, we have not discussed details, so, we have to little bit finish about this particular common, because the main problems main problems, we will derive from this particular assumptions because that is the starting point of econometric modelling.

(Refer Slide Time: 03:14)



So, you see here. So, what are the assumption regarding this bivariate modelling? So, that is our last class discussion. So, for this particular models, Y equal to alpha plus beta X plus U. So, we have three parts. This is one part, this is another part and this is another. This is called as a total variation, this is explained variation, this is unexplained variations. So, now assumption related to this particular component, this particular component and this particular component and also overall fitness of this particular model; So, the first standard assumption is that model must be linear in parameters, linear in parameters linear in parameters. So, what we have already mentioned here? Here, alpha beta are represented in the form of linear and that too, you can say without any problem. So, linear in parameters.

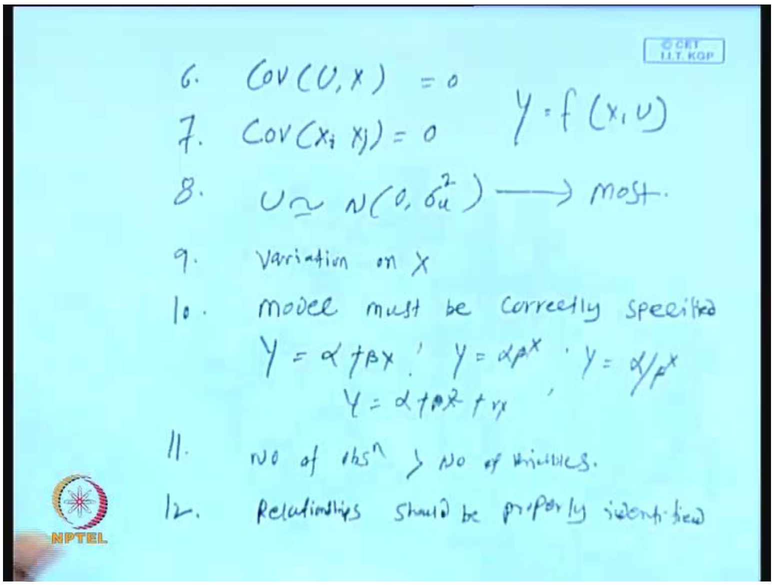
Second, our second assumption is that X should be, your independent variable should be non-stochastic. This should be non-stochastic. So, third item related to the error component. So, what we have discussed last class? Mean of error term should be equal to zero. Then, fourth assumption is the variance of error terms should be equal to unit, that is sigma square U. And fourth fifth assumption is that covariance of U i U j should be equal to zero. So, here this variance of U means, it is nothing but covariance of U i U j is equal to sigma square U, provided i equal to j. If i not equal to j and then, this will come down to this stage. So, now this particular this particular structure is called as a homoscedasticity. If the variance of error term is exactly equal to 1 or constant or you can say unique or this particular presentation is called as a homoscedasticity issue. So, now, if that is not the case, then this problem called as a heteroscedasticity issue and that is the serious problem under econometric modelling. We have a special component on a heteroscedasticity. We will discuss detail about that particular issue. If there is a heteroscedasticity, how to detect it? And how to solve this particular component? Until, unless, you solve that particular component, this model estimated model cannot be used for forecasting.

So, then covariance of U i and U j. Covariance of U I U j, it should be equal to zero. If it is not equal to zero, then there is problem called as a autocorrelation or sometimes, it is otherwise called as a serial correlation. So, now, if we have estimated models, then of course, with the help of estimated model, we will get the error terms. So, now, once we have error term, we will create several other variables related to error terms like U 1, U 2, U 3. So, now we have to track the relationship between U 1, U 2, U 3, like this. So, if these relationships by default cross correlation, if not equal to zero, then it will lead to serial correlation. For instance, like this: we have U 1, U 2 up to U n. So, this side U 1, U 2 up to U n.

So, now, this is we will correlate with U 1 1, U 1 2, U 1 n. Similarly, U 2 1, U 2 2, U 2 n. So, U n 1 up to U n n. So, now, this particular problem is called as a homoscedasticity and heteroscedasticity issue. This particular problem is called as a homoscedasticity or heteroscedasticity problem. If that is, if all these other items like this, these these particular you know diagonals, this particulars, this is called as a off diagonals and on diagonals. These are these are should be equal to zero, if that is not the case, then this particular problem is called as a serial correlation and autocorrelation. It has a serious problem again in the case of modelling particularly converted modelling.

So, when there is serial correlation; obviously, when we will go for any typical problem or any estimated model, there will be definitely autocorrelations. There is standard techniques are there, standard statistics are there to know the exact value of autocorrelation or serial correlation. There is certain limits, if it will cross that limits, means upper limit or lower limit, then it will problem for estimated model. So, we have to find out how that particular problem can be solved and that model can be used for forecasting.

(Refer Slide Time: 08:09)



So, this particular structure is called as a autocorrelation issue. Similarly, this is fifth standard assumptions, then there is another assumption called as a this is sixth assumptions. Sixth assumption related to covariance of U and X should be equal to zero because our model is equal to Y is function of X and U. So, there are three variables. We start with two variable initially but ultimately, we will get a variables called as a U. So, that is difference between U Y minus Y hat, that is estimated model. Covariance X must be equal to zero, otherwise it will create problem. So, I will detail discuss next next items.

So, covariance covariance upon you know X i or X j is equal to zero. In fact, this X I X j is not actual problem in bivariate model. So, when we will go for multivariate model, then one of the standard constant is that there are multiple number of independent variables like X 1, X 2, X 3, like this. So, now all these independent variables should be independent, means the relationship between all these variables should be independent. There should not be any again association between these independent variables. If there is such problem, then it is called as a multi-coordinate issue. So, we will discuss detail when we will go for multivariate modelling because we are in the bivariate models and this particular problem may not be there, means it should not be there in the case of bivariate model.

So, that is how there is a covariance between U and X. Now, since U is in the right side of this particular equation, obviously, there are two variables X and U. So, now, if we will integrate with this multi-coordinate issue X and Y should be totally independent that should, that means, there there should not be any association between X and U. If it is there, then it will be problem.

So, now, if we will integrate this, all these equation in structurally with respect to error term, then error term should be followed by normal distribution zero mean and unit standardizations. Now, zero mean and unit standard. This is the most important one, most important one for this particular econometric modelling.

So, then Ninth there should be variation on X, there should be variation on X. Of course, if there is variation on X then there should be variation on Y. For instance, we have Y series and we have X series, forget about Y in the mean times. So, now, for every Y there is X items. So, now, if the X items are very much similar, then obviously, we will get our model cannot be better fitted. For instance for every Y in the X values are 2 2 2 2 say or 5 5 5 say, then obviously, there is no such variations. So, in this particular context, this model cannot be better fitted. So, this is by default we have to replace it, otherwise you cannot get a best fitted models. These are the inspection, before you go for you can say estimated models. So, because to get the estimated model, you have to put lots of labor and efforts. So, after that, if the model is not reliable, then obviously, you have to redesign and re-estimate to get again better fitted model, but there are certain clues before you go for estimation. So, you have to clarify all these detail and one of such clue is like this, there should be variation on both Y and X, but X is very important here because X is a, X influence Y only.

So, there should be some kind of variability in X. So, this is another assumption related to variance of X and model must be correctly specified. Model must be, model must be correctly specified. Model must be correctly specified. What is what is mean that? For instance, we are just representing Y equal to alpha plus beta X. So, there are many ways we can we can represent this particular relationship between Y and X. For instance, we can also write Y equal to alpha plus beta to the power X or we can write Y equal to Y equal to alpha alpha beta by X, like this or we can also write Y equal to alpha plus beta X square plus gamma X, like this. So, there are many ways many ways we can represent the relationship between Y and X, but for a particular problem or particular instance every every equation may not be very fitted for the econometric modelling.

So, we like to know, what is the best mathematical relationship, we have to use to get the best fitted models. So, the for that you have to do lots of homework and you have to test, there are certain procedures of test or graphically, you can plot it and you like to know what is the exact lesson, whether it is linear one or non-linear one. Then, accordingly you have to proceed further. In fact, if it non-linear one, then it may be very complex problem

So, what we have to do in that linear setup, you have to first transfer to the linear format by using transformation rule. There are various transformation rule through which you can transfer the non-linear program to linear one. Then, you have to go for estimation. So, if you do not go for you know transformation, then it will create lots of problem again and in typically, it is problem with OLS technique.

So, you must be very careful about this. Then, then next is number of observations number of observations should be greater than to number of variables number of variables number of variables, number of observations should be greater than to number of variables. For instance, you see this this particular problem for bivariate econometric it is not a serious issue, because we have two variables only. So, obviously, most of the problems is more than, you will get more than two sample. So, if it is more than two samples, then obviously, there is no issue at all.

But, when there is multivariate problem where the problem setup consist of you can say so many variables like, say ten, twenty or thirty or hundred, then that time sample size is very typical issue. So, you have to very careful how many variables you must be putting in the system and how many observations are there. So, by default it should be whatever variables you are using in a particular system, your number of sample size should be five times than that. So, that means, if the variables are 5 in numbers in a particular system or in particular modelling setup, then obviously, sample size at least should be 5 into 5, so that means, at least Twenty five, but Twenty five is also very small one. By standard rule, your sample size should be greater than to Thirty. If it is greater than to means, Thirty two, exactly, if it is greater than Thirty two, then obviously, there is, you can say we have two different sample structures, small sample and large sample.

So, minimum number of sample must be greater than to Thirty two or for a particular problem, whether it is bivariate and multivariate. And if it is multivariate, then obviously, then there is second criteria. So that means, the second criteria is number of variables and there sample size should be multiple to five. So, now, if it is ten variables, then obviously, sample size should be at least fifty. If it is less than that, then there is problem of modelling or estimation process.

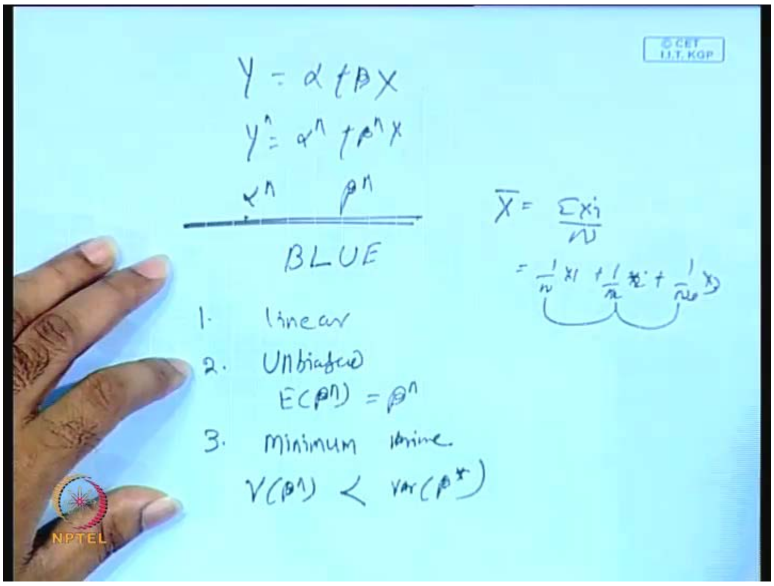
So, if it is more than that, then no doubt about it. The model accuracy is very high when the sample size is very high. The model will be less accurate, if the sample size is very small. So, that is how, you have to go. Always go for higher and higher sample size. So, that, the model accuracy will be very very perfect. So, that is how the reliability part is concerned because reliability part, the objective of reliability part just to check whether the estimated model can be used for forecasting or can be, we call it that it is best fitted model. So, that for that, we we have different test structures. By this process of test, we like to say, we like to be in a position to say that this particular model can be used for forecasting or you can say policy use. So, this is eleventh assumptions. So, it is well number of observation greater than to number of variables.

Then, last but not the least, then relation should be very identify one. So, relationships relationship should be identified, properly identified, should be properly identified. For instance, you see here, this particular problem for two variable model, it is not a issue, but when you will go for multivariate models, particularly there is, we have a component called as a simultaneous equation modelling or structural equation modelling. In that context, you know model identification is very very serious issue. If you do not identify properly that model, then the process of estimation or its interpretation or reliability will be get affected.

So, what we have to do first means, you must have a very sound knowledge, that is theoretical knowledge before we go for you know fitting the relationship. So, because, so far as a interpretation part is concerned, so, the theoretical knowledge or theoretical background will give you lots of ideas, how to interpret this particular model or when you will go for forecasting issue.

So, these are the standard assumptions through which OLS technique is practically feasible. If this assumptions are on other way around, then obviously, these particular technique is you can say, the model you will get through this particular technique, cannot be used for forecastings or you can you cannot deal for any use because it will give you wrong signals or you can say wrong indications.

(Refer Slide Time: 18:14)



So, now now what is the best idea is the, so, when we have a model, so, Y equal to alpha plus beta X, then we have a received the models called as a Y hat equal to alpha hat plus beta hat X. So, now, we have two standard estimators. So, this is called as a alpha estimators and beta estimators. You know, this particular estimators should follow certain principles. So, there are certain principles behind this particular estimators. This particular this particular principle is called as a represented as the term called as a BLUE; Best Linear Unbiased Estimator. So, that means, whatever estimators we are receiving; that is, alpha hat and beta hat, that should be best, that should be linear one, that should be unbiased one.

So, now, you see here what is linear? Linear means its linear combination of variable. For instance, X bar is nothing but summation X I divide by n. So, that means, 1 by n 1 into X 1 plus 1 by n 2 plus into X 2 into X 2. So, 1 by n n you know n 3 X 3 like this. So, these are all linear combination of x.

So, what I like to say is that, whatever estimators value we have received, alpha and beta hat, so it should be linear in nature. So, that you know the theorem can be meaningful. So, that is how it is called as a best linear .With the classical linear regression models, the estimators what we have received by the process of estimation, should be linear, unbiased, and you can say it must have a minimum variance.

So, what is the what is that minimum variance. So, before that you must have linear in natures then second then there is called as a minimum variance or you say first start with unbiasedness. Unbiasedness means, what is first biased? Biased means difference between expected value of you can say let say we start with beta parameter only. So, estimated beta hat should equal to beta hat. So, this is true beta.

So, now if the estimator beta is exactly equal to true beta, so, there is no gap at all,but we are expecting something what you know actually, something that gap should be very very minimum. If that gap should be high, then obviously it will create lots of problems. So, that means, for every problem ,whatever estimators we are receiving, that should be unbiasedness. So that means, the estimated value should be exactly equal to truthful. There should not be drastic difference between this. As long as the difference is increasing, then the model accuracy will be getting affected.

So, this this should be taken care. So, this is second properties of this particular estimator. Third is the minimum variance. third is minimum variance minimum variance. So, what is minimum variance? Now minimum variance is, you have to find out, you know variance of variance of beta hat is should be less than equal to variance of variance of another estimator say beta star.

So, that means, whatever estimator you have received, that should be minimum variance. You see, variance is variance is an indicator, through which we can just the accuracy of particular theme or particular item. If the variance is very high then obviously, obviously, that model can be you can say cannot be consider as the best fitted model.

So, just like you know we have discussed details in the univariate data setup, where this particular component is called as a dispersion issue. So that means, the variation from the center point should not be so drastic. So, it should be uniformly distributed and whatever the variance we have received, that should be the minimum one. If it is minimum one, then the model can be treated as the best fitted one. If it is not minimum then obviously, it will get affected.

So, variance should be at the lower level. Then forth one is, it should be consistent. Consistence means, it is integration about minimum variance properties and unbiasedness property; that means, its e beta for estimated beta hat so, e beta hat should be equal to beta and variance of beta should be less than to another variance, say beta beta star.

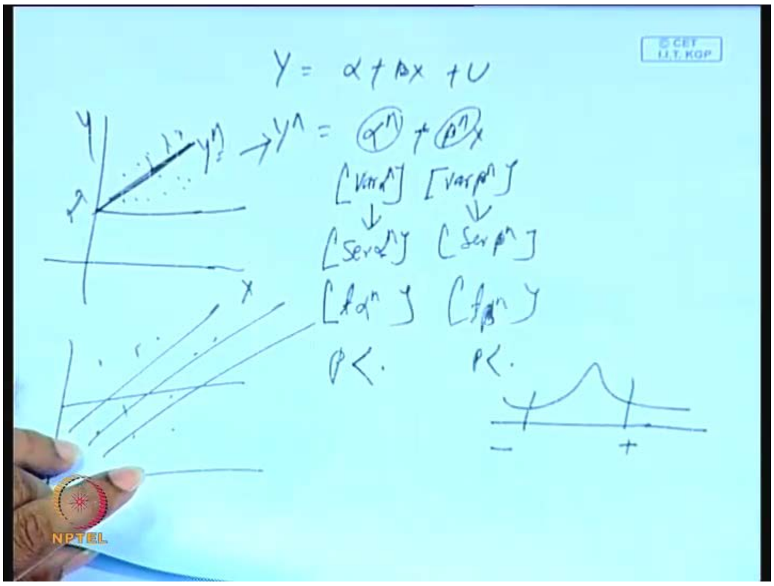
So, when we will go for unbiasedness property only, then there is no link with minimum variance. But when we will go for minimum variance, then there is no no point to go for you can say unbiasedness property. So that means, there should be there should be, you can say, two things are completely different, but so far, as consistency property is concerned, this integration about unbiasedness property and So, unbiased property is that expected value estimated beta value is equal to true value that is not beta hat. In fact, that is beta only.

So, e beta hat equal to beta and variance of beta hat should be less than to another variance of estimator. So, now, if that is the case, then, this particular theorem is called as a BLUE; Best Linear Unbiased Estimator.

So, now every model has to be tested with respect to all these indicator. These are all called as a means these are we are assuming that, these are the indicator, through which we we just the reliability of the particular models. If whatever estimator we have received, that should be taken care with respect to all these indicators. If these indicators are going on the right track, then the model accuracy will be high and we can say that model is reliable one.

So, now we have to know what is exactly this reliability part of this modelling. So, this reliability part of this modelling basically see here is. So, what is exactly the proper setup here is..

(Refer Slide Time: 24:11)



So, now, the model is like this. Y equal to alpha plus beta X plus U. All right

So, we are getting Y hat equal to alpha hat plus beta hat beta hat x. So, now so, far as a is reliability is concerned, so you we have to integrate so many things, because we like to know, whether this alpha hat and beta hat are you can say reliable one. So, that is, with respect to this particular theorem, BLUE- Best Linear Unbiased Estimator. So that means, alpha hat and beta hat has to be checked so that the theorem can be satisfied. So that means, the properties of all these estimator should be as per the rules or principle of this econometric modelling. So, that is what we call it best linear and unbiased and obviously, it will go by minimum variance and that a consistence property.

So now, how do we go for that? The when you will get the estimator models, then basically the estimator model will be represented by like this. This is followed by variance of alpha hats variance of alpha hat and with respect to beta hat,we will receive variance of variance of beta hat. So, our standard assumption is that variance of alpha hat should be very very low and variance of beta hat should be very very low.

So, with respect to variance of alpha and variance of beta hat, we like to receive the standard error of standard error of alpha hat and standard error of beta hat. So, which is derived through variance of alpha hat, which is derived through variance of alpha hat? Then, there is there is term called as a stat .Then we have to apply this statistics actually to check whether this particular parameter is statistically significant or not. Ok.

So, now to know the statistical significance of this particular parameter,that is what another part of the reliability, it is not like that whatever estimator we have receive,that is just you know follow the principle of blue. In the same times, these particular items should be statistically significant and for that, we we have lots of test statistics, like you know, t test, f test, j test, like this.

So now, for these particular standard problems, we usually help take the help of t statistic and f statistic and for this particular angles, so, means, so far as the you know significance of the alpha hat and beta hat is concerned, we have to apply the t statistic and through which we like to make a judgment whether this particular parameter is statistical significant or or not. If it is statistical significant then obviously, this particular model can be considered. If this parameters are not statistical significant then it will be problem it will be problem definitely.

But you know, here in this particular bivariate setup, we have only two items. So, that is alpha hat and beta hat. But when we will go for multivariate model, there are several beta's like you know beta 1, beta 2, beta 3. Alpha is just supporting component that is represented as you know sorry you know constant item and there is slope items. This slope items are you can say beta 1, beta 2, beta 3 like this and you know constant term we call is intercept. This is what we we when get the supporting factor through which you have to start this issue like this you know.

So, this particular representation is like, this is this is alpha hat this is alpha hat and through which we have to start the Y hat component that is nothing but this particular item alpha hat and beta hat x.

So, now this side is X axis and this side is Y axis. So, now, you like to assume that this is particular estimated model; in between there is true value, these are the true value. So, we like to know, how the you know best fitted line can be derived through this you know true values or true points. Because it should be you know by default or you can say by inspection it should be in the middle of all these points. So, the way you have to design, so that the you know path will be in between the two so that the model accuracy will be very high.

So, now you know, when you have plotted points, there are many ways you will get to know mid points. It may be you know like this, when we have several points like this.

So, it can be like this, it can be like this, it can be like this. So, all this things cannot be possible to you can say consider.

So, we have to choose one. So, that is one must be very best. So, by inspection it is very difficult, of course you can make a judgment, but it is not accurate judgment. So, that is how you have to go through statistical procedures only. There is statistical procedure, that statistical procedure is called as a reliability check. So, that reliability check will indicate whether this particular path is best one or this particular path is best one or this particular path is best one.

So, So, as far as significance of the parameter is concerned, then standard error of alpha then we have we use t statistic t alpha hat similarly we have to see t beta hats. So, then since it is issue about the significant, then significance for significance level we have to see we have to the compare with probability level.

So, in this particular system, so, the significance level we usually consider it you know 1 percent level, 5 percent level and 10 percent level. So, maximum limit we will go up to 10 percent. So, now, 1 percent, if this some item is significant at 1 percent, then obviously that means, 99 percent chance is you know fact and 1 percent is not supporting. Similarly if it is 5 percent; that means, 95 percent are supporting and 5 percent are not supporting. Again 10, if we will say 10 percent, then 90 percent are supporting and 10 percent are not supporting.

Again in between, there are, this test procedures are usually divided into two parts called as a, one tailed test and two tailed test. If it is one tailed test, then it may be consider as a 1 percent level, 5 percent level, 10 percent level. If it is two tailed test, then you have to divide by 2. So obviously, two tailed test at 1 percent means you have to go 0.05 then 0.05

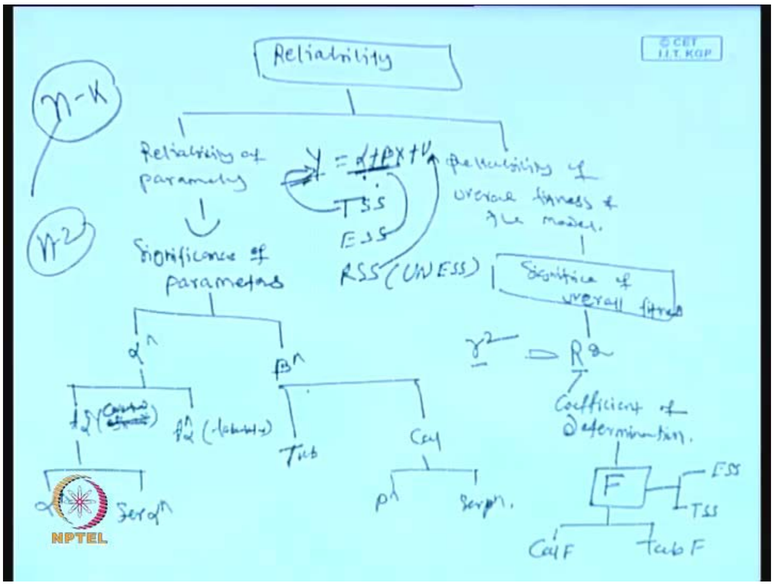
Similarly, 5 percent means, you have to go for 0.1 and 0.1.Similarly, 10 percent you have to go for 0.5 and 0.5. So, this is how you have to proceed. So that means, this particular item is normally distributed. So, this part and this part has to be considered, this is plus side and this is minus side.

So, this is how you have to; that means, 50 percent is this side and 50 percent is this side. So, if it is one tailed, then then it is q distribution that is positive least square or negative

(()). So, this is how you have to go through it. So, now, so, probability level you have to decide and probability level you you have to decide. And with this particular system, we have to check the standard, you know best fitted model or reliability model.

So, now I am giving you the exact procedure of reliability. What is all about this reliability of this estimated model?

(Refer Slide Time: 31:45)



So, reliability, so far as a reliability is concerned, reliability of estimator is concerned, we have to check two things; first is with respect to reliability reliability of parameters and second is reliability of the reliability of the overall fitness of the model overall fitness of the model overall fitness of the model ok.

So, first is a reliability of estimator; that means, the initial starting point is you must have Y information axis, information about your problem, your your basic relationship between Y and X must be known to you, then you have to make a functional relationship, that is what we will call it in mathematical model, then you transfer into statistical model by introducing the error component, then you have to estimate the models by the use of OLS technique , then you like to have Y hat equal to alpha hat plus beta hat x.

So, the moment you have Y hat equal to alpha hat plus beta hat X , then we are going for reliability. So, So, the reliability has a two parts. One part is with respect to significance of the parameters that to alpha hat and beta hat so what we have just explained detail.

And second is the overall fitness of the overall fitness of the model. So, the overall fitness of the model is that that can be also standard technique through which you have to judge the overall fitness of the model. Just like you know, see here there are three parts you know, alpha plus beta X, alpha 1 part, beta X another part and you know Y is another part. So, this is one side and this two part is another side. So, this should be, you know this this particular reliability means, so what is the overall impact of this particular models, so that is means explained side then unexplained side, then obviously, the reliability of the individual parameter; that means, section a and section b that is, alpha hat beta alpha hat alpha plus beta X is one part and you know U is another part.

So, we like to know what the judgment of explained item is and what the judgment of unexplained item is. So, reliability has two parts with that is with respect to parameter significant check and this is with respect to overall fitness of the model ok.

So far as this reliability of the parameter is concerned,so we like to know this significance significance significance of parameters. So that means, if parameters, so, since; that means, if your objective is to check the reliability of parameter; that means, the idea is very clear, we like to know what is the significance level of the parameter. If that item is statistical significance and close to and percent, then that the model accuracy is very high. If it is if, that item is not statistical significant and 1 percent at one tailed test or two tailed test, if it is totally failed, then you have to go for again 5 percent. Ok.

So, now if 5 percent, again you have to go one tailed test or two tailed test. If it is not significant, then you have to go to 10 percent level. Again 10 percent level, you have to check it at one one tailed through one tailed test and you can say two tailed test. If not you know, after that 10 percent, 10 percent experiment, if it is not statistical significant, then we can conclude that, this particular you know item or variable is not at all statistically significance. Yes there is relationship, but that relationship is not strong enough to use for you know forecasting. So, that is how the reliability you know rule is concerned.

So, now what is the significance of parameters? Here the significance of parameters, we have two aspects, that is, you know, alpha alpha aspect and beta aspects. So, we like to know what is the significance of alpha hat and beta hat. For that, you know we have to know two things.So alpha hat is known to us, so, we like to know t alpha hat and t alpha hat, t alpha hat you know tabulated value tabulated value, and this is estimated value, this is estimated value t alpha hat and t beta hat.

So, you know this is not estimated. In fact, this we call it a calculated value. Calculated calculated t statistic and tabulated t statistics. So, now, we have to compare the calculated t of alpha hat and tabulated t of hat. Similarly we have to calculate calculate the or you have to compare the calculated t of beta hat and tabulated t hat beta hat all right.

So, now what is happening here is, so, now, this t of alpha hat depends upon two components. t of alpha hat depends upon the individual alpha hat value and its standard error of its standard error of alpha hat. Similarly, you know for beta hat, we have two aspects, that is, you know, tabulated tabulated t statistic and you know calculated t statistic. Calculated t statistic again depends upon beta hat and standard error of beta hat. ok.

So, this is one part of the story, one part of the story, that is what we called as a reliability of parameter. So, once what is the exact structure of reliability of the parameters? So, we like to know means, we must have alpha hat value and beta hat value, so we must have a variance of alpha hat and variance of beta hat, then we have to get the standard error of alpha hat and standard error of beta hat because this square root of variance of alpha hat is nothing but standard error of alpha hat and square root of beta variance of beta hat is nothing but standard error of beta hat.

So now that means, we must, you we must means we must have alpha hat and variance of alpha hat and beta hat and variance of beat hat then everything will be automatic means not automatic everything can be very smooth to get this work done .All right.

So, now the second part of the model is called as a reliability reliability of overall fitness of the model. So, this is what we called as a significance or otherwise it is called as a significance significance of overall fitness of the model over all fitness of the model fitness of the models.

So, overall fitness of the model for a particular econometric, you know technique solely depends upon and the statistic called as a ((r squares r square)). This is ,what it is usually depends upon r square, this r square represented as a coefficient of determination. We we we have discussed r component, that too small r , that is what we called as a simple correlation.

So, covariance of X Y upon sigma X and sigma Y that is with respect to two variables. In fact, the starting point of correlation is two variables . So, now, we we we must have a you know we must have two variable, then we can calculate the correlation, then correlation coefficient simply represented as a small r. If we will make it is square then it is square of correlation coefficient.

But for bivariate models, this particular small r and capital r model is same. But when we will square it, r square, that is what we will call it its contemporary its small r square. But this r square this r square there is difference. So, this is called as a coefficient of determination. This particular item is called as a coefficient of determination, this particular item is called as a coefficient of determination.

What is the rule of coefficient determination? The rule of coefficient determination is that, it is the ratio it is the ratio between explains to total. So that means, what is the percentage of explained component is information the dependent variables. So, that is what we will call it r square, that is overall fitness of the model.

So, this the moment there are two means two steps here. First is, you have to calculate the r square statistics. So, now, once we will get r square value, so that r square value has to be tested again, because here we are we are having Y equal to alpha hat and beta hat X only. So that means, alpha hat is known to us and beta hat is known to us, nothing is available, but lots of other statistic also simultaneously we have to calculate.

So, for the reliability is concerned, one part of this particular item is called as a r squares. r r square is the ratio between explained sum square by total sum square we will derive details in the next class not today.

So, now. So, this r square structure is like this means, it is the ratio between explained sum by you can say total sum square. So, now. So, the. So, far as the reliability of r square is concerned, then it depends upon its depends upon f statistics, this depends upon f statistics. So that means, for significance of the parameters, we are using all t statistics and for overall fitness of the model, we are using the f statistics.

So, now, like this particular case, here also you like to know , what is the calculated f and what is the tabulated f, calculated f and tabulated f. So, now, we have to compare the calculated f with tabulated f. So, means, what is exactly the difference between calculated f and tabulated f or similarly calculated t alpha hat and tabulated t alpha hat, calculated t beta hat and tabulated t beta hat.

So, now the thing is that, for calculated there is standard you know techniques or you can say formula, through which you can obtain this value, but there is you know by simulation we have statistical tables.ok.

So, statistical table has a standard norms. So, now, we have a calculated statistics and we have a standard norm. So, we have to compare the calculated statistic with the standard norms. If the calculate statistic is the greater than to the standard norm, then the statistic then that item will be statistically significant. But the standard norms are available, you know, at difference significance level. For instance, at 1 percent level, at 5 percent level, at 10 percent level. Again for one tailed and two tailed, one tailed and two tailed, again one tailed two tailed, again one tailed two tailed, that means, with respect to 1 percent, 5 percent and 10 percent.

But some of the standard books there is you know chance of 25 percent, then I go also up to 25 percent. But but you know (( )) you know who for statistics and you know finance finance problem or any other technical problems. So, we usually handle up to you can say 10 percent level only.

But social science problems which is very complicated and when your sample size is very extremely high and high, then that times you can go for, but very extreme case, you have to go for 25 percent. But 25 percent is not at all you know very reliable one. So, in most effective way, you have to go up to 10 percent. So, if it is not up to 10 percent then you have to reject or redesign the structure entirely all right.

So, now for t statistics, you have a t tables and for f statistic you have a f tables. So, again for t table weighted and you know f table weighted, there is another items you need to integrate to get this particular item, that is what we called as a degrees of freedom degrees of freedom. Degrees of freedom means it is the difference between total number of observations and number of you know independent variables represent in the system.

So, we will discuss in detail when we go for that is what we will sometimes call it is a n minus k, n is total number of sample size and k is the total number of total number of variables independent variables in the systems. So, n minus k is or total number of parameters in the system whether instead 2 variables you can say that k represents total number of variables parameters involved in this particular model

So, now this n minus k represents the degrees of freedoms. If the degrees of freedom is very high, then the model accuracy will be very high. But you know for this is true for the t statistic, but for f statistic we have two different. In fact, for this f statistics we have 2 different component, one is called as a explained sum square and another is called as a total sum square. So that means, there are three standard items, one is called as a tss, then ess, then sometimes called as a rss or you can say sometimes it is called as a unexplained sum square explained sum squares ok.

So, tss represent total sum squares, which is nothing but the actually we will derive from this structure, Y equal to alpha plus beta X plus U. So, this particular item is nothing but explained sum square and U related to unexplained residual sum square and t represents to Y. So, this is how the all about the integrations. So, but the calculating procedure is something different with respect to total sum square, explained sum square and residual sum square.

So now, so when we will when we have a information, so we have, we have Y information and X information through which you fit the model you get the model, but anyway you fit, but when you'll go for typically means in a smaller version of the problem, then usually we will not derive every times.

So, what we have to do,we have Y information and we have X information. So, we know what is alpha hat formula and we know what is beta hat formula, similarly we have also formula for variance of alpha hat and variance of beta hats.

So, again we, through statistics, t statistic, there is calculate a statistic, how to get this t value. So, now, if we will go sequentially,with respect to the available information, then you can automatically or by very easily you can get this particular value, that is you know, t alpha hat calculated and t beta hat calculated similarly in the case of.

So now, once we have Y information and X information for this particular bivariate setup, so, what we have to do in the first step itself, you have to calculate sum Y, sum X , sum X square, sum Y square, sum X y.

So, now in the same times, we must identify number of sample size,that is you know n, and in the same times what is k. In fact, for this bivariate models, k is usually represented as a 2. So that means, the structure is n minus 2, degrees of freedom is n minus 2, k represents number of parameters in the systems, number of parameters in the system or number of variables in this particular systems all right.

So, now this will determine the degrees of freedom. So, now, once once you have such information, then you first calculate the alpha statistic, alpha hat statistic, then beta hat statistics. First you start with actually beta hat, so that is, summation X Y X Y by summation X square or means it is in a deviation format. If we will go by other way round, then it is n summation actually minus sum X sum Y divide by m summation X square minus sum X whole square. This is the standard formula.

So, whatever step of this particular problem, but you know whether this step of the problem is perfectly or not. That is how we are checking here is. So, that that is how there is no you know hard and fast rule that ,we have to you know check the functional form or you have to derive etcetera. First you do it, but the best procedure is, check it properly, then you fit the model accordingly then obviously, reliability is very consistent. Or else, what you can do to, you start with something, then automatically, at the point of reliability, it will be give you indication, whether this model is perfectly for forecasting or not ok.

So, now let us let us assume that we are you know we have no knowledge about it all these testing etcetera, so, we start by the standard rules, you get the alpha hat and you get the beta hats, then you fit the model like Y y Y y hat equal to alpha hat and beta hat x.

So, now you calculate the alpha hat statistic, beta hat statistic. Then you go by standard procedures, get the variance of alpha hat, get the variance of beta hat, then get the standard error of alpha hat and standard error of beta hat, then you calculate the t of alpha hat, calculate the t of beta hat, then, you compare with it calculated t of alpha t of alpha hat and calculate a t of beta hat.

So, this is the tabulated value available available with you and obviously, you have to calculated value. So, then we have to make a comparative analysis. In fact, you go any statistics books or any econometric books, in the end of the books, you will find standard statistical table, you know with respect to t statistic, with respect to f statistic and with respect to j statistic, etcetera etcetera.

But you know, we need only t t tables and you know f tables. But t statistic is easy to understand and easy to pick up, but f statistic is little bit complicated,why because, when we will get r square, then f is the difference means, in this particular setup, f is the ratio between ess by tss.

So, now ess is a standard standard you know statistics and tss is also standard statistic. So, ess depends upon something, because it is it is degree of freedom is completely different and tss degrees of freedom is completely different. You see here is, this is this is ess part. So, this part, the degrees of freedom, this part is degrees of freedom is and this part degrees of freedom is, because it depends upon two parts. but this is on the one part ok.

So now,so, since the degrees of freedom are are different, so obviously, in the case of ess, the degrees of freedom will be completely different and you know below tss the degrees of freedom is also completely different.

So, as a result,so in the f case, for f statistic, the table is something different. So, that is why you must be you must have thorough knowledge how to pick up the tabulated figure.

So, once you get the tabulated figure, then you have a calculated figure, you make a comparative analysis. If the calculated statistic is higher enough than the tabulated statistic, then the item is statistical significant. Usually you start with at 1 percent, because one statistical significance 1 percent is the best. If it is not best, then you have to go for 5 percent, the second best. If it is not 5 percent, then you have to go for third best that is 10 percent. After that, if that item is not statistically significant at 10 percent, then the better you have to stop there. Stop means, you have to either redesign or you can say restructure, then again you have to go for reliability test, till you get this you know items significant.

But for bivariate structure, it is very easy to pick up for you can proceed. But when you will go for multivariate, its the problem is very complicated. It is very difficult to get all these variables statistically significant or overall fitness. Obviously, when we will introduce 1 after another variable or when we will move from bivariate to multivariate, then obviously, explained sum is all you know with respect to X items and total sum is with respect to Y only.

So, obviously, Y is always one. So, that item will be remain constant. So, when we will add 1 after another item, the numerator will be start increasing. So, as a result r square will be always high, once you introduce one after another variables. So, when r square is exclusively high, then your f statistics will be exclusively high.

So, the overall fitness will be usually very high or you can say you can be get significant. But the parameter you may not be get significant. So, you have to make a compromise. So, the compromise rule we will discuss details when we will go for multivariate framework.

So, with this today we will stop here is and we will discuss the details in the next class. thank you very much have a nice day.