Econometric Modelling

Prof. Rudra P. Pradhan

Department of Management

Indian Institute of Technology, Kharagpur

Module No. # 01

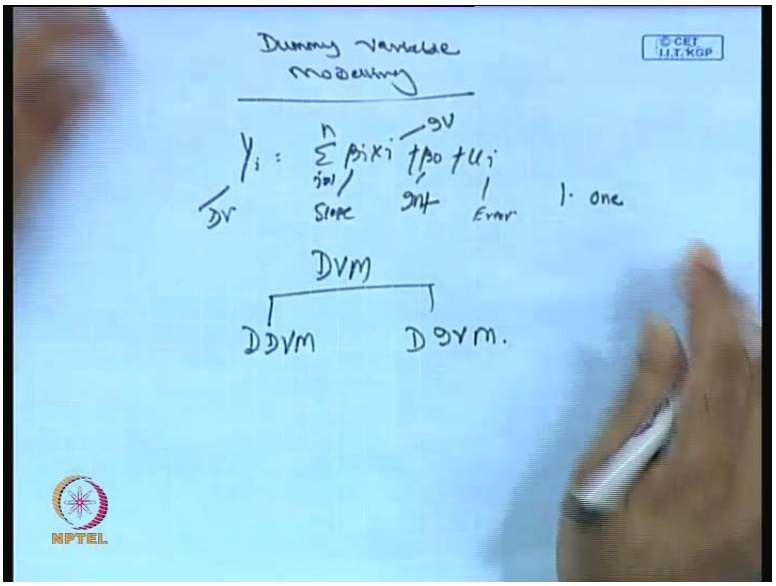
Lecture No. # 26

Dummy Modelling

Good morning, this is Doctor Pradhan here. Welcome to NPTEL project on Econometric Modelling. So, today we will discuss problem called as a Dummy Variable Modelling. So, in the last couple of lectures, we have discussed the entire structure of econometric modelling that too bivariate, trivariate, multivariate and we have also discussed various problems with respect to econometric modelling, that is, you know, multicolinearity autocorrelation problem, heteroscedasticity problem etcetera. So, today we will discuss the concept which, we usually call as a dummy variable or you can say, binary variable or you know categorical variable, there are many names all with respect to dummy variables.

So, it is a very interesting technique and it is very useful and most of the, you know, areas, it is, it has a lots of applications. So, let us start with, what is the exactly the structure of dummy variable modelling and then we will highlight some of the application, where we can apply this term variable.

(Refer Slide Time: 01:31)



So, the starting point is that, we have an we have an econometric modelling say like this Y, Y i equal to summation beta i X i, i equal to 1 to n plus beta 0 plus u I, so here i equal to 1, this is intercept, this is slope coefficient and this is error term, and this is independent variable, this is dependent variable, this is how the model is classified.

The problem is here, when we discuss something about to, you know, there are many ways econometric modelling can be presented. First things, we have discussed bivariate, trivariate and multivariate where, the structure is that one dependent variable and there are several independent variables, starting from one independent variable, two independent variable, three independent variable, you know, four independent variables like this, or you can say, k th independent variables. So, that means, keeping one dependent variable constant, we are extending one after another independent variable, then we are saying, that it is bivariate, trivariate and multivariate like this.

So, whatever may be the situations? So, when we have, means, one structure of econometric modelling is that where, there is a one dependent variable and one or multiple independent variables. This is one way of, you know, highlighting the econometric modelling issue. Another way to represent the econometric modelling is that, we have several numbers of dependent variables and also in the same time there are several number of, you can say, independent variables. So that means, that is called as a

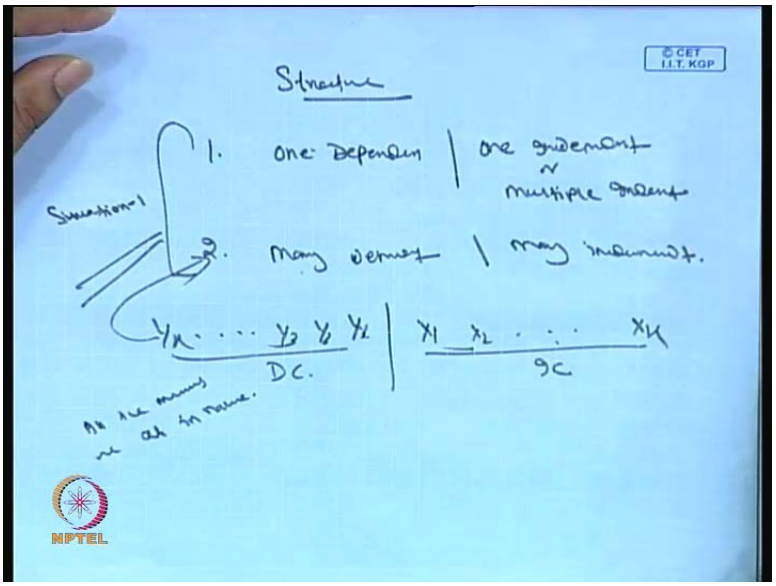
other, in that scenario, there is a, means, that is a situation of, interdependence among various variables.

So, in that structure there is no such exact classification of, you know, dependent and independent but, to at the particular point of time, we have to specify, what are the dependent variables and what are the independent variables? Of course, one dependent variable sometimes one is independent variable to somebody else and says, you know, same thing can be also independent variable to some dependent case. It is a very, that is how it is called an interdependent, that means, if Y y is a function of you can say, Y 1 function of Y 2 then, obviously, Y 2 may be function of Y 3, like this. So, it is a chain type situation, so this is another structure of econometric modeling, that is called as a usually simultaneous equation modelling or you can say structural equation modeling.

So, we will discuss little bit later on that aspect, what is all about the simultaneous equation modelling and structural equation modelling. So, so, this is the second setup of econometric modeling. In the first setup one dependent with one or many independent variable, in the second structures, with the many independent and many dependent variables, so that means, it is a interdependence among the various variable. So, best technique is a simultaneous equation modeling, structural equation modelling or you can say there is an advanced technique called as a factor analysis etcetera, etcetera.

Another structure of modelling is that means, another structure we, usually find in this econometric modelling is that called as a dummy variable modelling. So, dummy variable can be can be divided into two parts like this. Dummy variable modeling, dummy variable modelling basically divided into two parts, as per the dependence structure as per the independent structure, so that is called as a dummy dependent variable modeling, dummy independent variables, dummy independent variable modelling and dummy independent variable modeling. So, that means, you see here, in the third structures, means, in the first structures. Let me highlight here is in the first structures, 1 dependent; let me put it in other way.

(Refer Slide Time: 05:46)

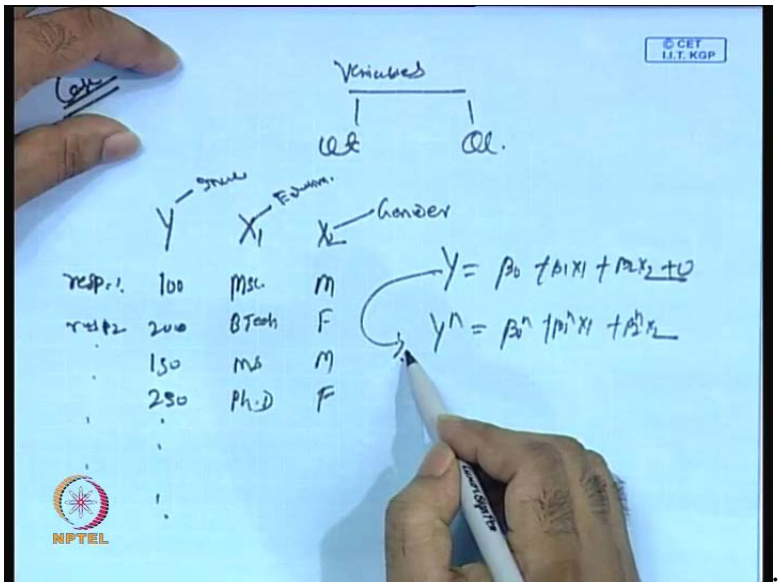


Let me, first highlight this structure, then we will go to this dummy technique. This structure is that, so first case, one dependent, you know, other side, either one independent or multiple independent. This is one structure and second structure is that many dependent and many independent. Of course, this is you can say, situation 1, why I will call it situation 1 because, here one of the standard assumption is that, whatever variables, let us say, I will put like this way. So, it is nothing but, you know Y 1, Y 2, Y 3 upto, you can say Y k; this is X 1, X 2 up to say X k, so this is independent classification, this is dependent classification.

So, this is how we can represent, so now 1 1 then, this is simple, then 1 with 2, then it is called as a bivariate, 1 with 3 multivariates, so this is how the structure is all about. So, now, in this particular structure, one of the standard assumption is that all the variables, all the variables are quantitative in natures, quantitative in natures; that means, it can be measured, it can be expressed in terms of substantial quantity. For instance GDP, FDI, forex, these are all quantitative information. If I will say, what is the GDP of a particular country, what is the GDP of a particular country, for a particular year then, it is some quantitative figures.

So, that is the way, which, we can discuss this, you know, bivariate, trivariate, multivariate and this simultaneous equation modelling or structural equation modelling.

(Refer Slide Time: 08:01)



So, in addition to this particular pictures, there is the another case or we can call it a situation to, or case 2, so here, there is a another type of classification, means, assumption is there, some of the variables may be qualitative in nature, this in the first, in the first case our standard assumption is that, the variables are very much quantitative in natures. So, there is no concept of qualitative, but basically variables basically variables are divided in to 2 parts, one is quantitative and another is qualitative.

So, quantitative and qualitative, quantitative means it can be expressed some number, some percentage, some ratio etcetera etcetera and qualitative means it is just like a categorical variable or binary variable or you can say dichotomous variables like this. This is how we have to represent the situation. It is usually called as a proxy variable, so that means, there is a no direct quantitative information with respect to that variable. For instance gender is a particular variable. If I will ask, what is the gender impact on something else then, obviously, gender cannot be calculated in quantitative terms.

So, if I will say gender then, obviously, the structure is male and female. So, that means, only information if you have collect in a data then, obviously, I will ask you what is the gender? So, the respondent will simply, you know, reply either male or female. So, I will record, I will record accordingly, on means in that particular structure. For instance I have Y something and I have X something and I have a X 1 something and X 2

something X 2 say the variable gender, then, I will ask you suppose, this is respondent 1, respondent 1, this is respondent, respondent 2 like this it will continue

So, in that structure suppose, I will ask respondent 1, what is your Y say, it is let us say it is income levels, as if your X 1 say, it is you can say yes educational levels, educational levels. Then, obviously, I will ask you first, what is your income then, obviously, you will we simply say that something else, say 100 dollars or you can say 200 dollars or you can say 150 dollars or you can say 250 dollars, so like that it will continue. So, then I ask the respondents again, what is your educational qualification then, obviously, we have standard structure of educational qualification. Let us say under graduate, post graduate or you can say matriculations or you can say PhD etcetera. So, accordingly you have to record here.

So, let us say, this is a, this is guy is called as an Msc. So, this is guy, called as a, you can say b. tech. Then, this is guy you can say, PhD. This is called as M.S. So, this is how, we have to collect a data, similarly genders. If I will ask the gender then, we will simply say male or female, male or female like this so. That means, now if I will fit a regression like this, Y equal to beta 0 plus beta 1 X 1 plus beta 2 X 2 then, obviously, this is original, original model. So, we need to have estimated models, Y hat equal to beta 0 hat plus beta 1 hat X 1 plus beta 2 hat X 2. So, now, to transfer 1 equation, equation 1 to equation 2 we needs information.

So, information must be, till now, whatever we have discussed, information must be in quantitative sets. So, that accordingly, we will put this formula and we will get the beta 0 hat, beta 1 hat and beta 2 hat then accordingly, we have to go for you know the testing of you know, for significance of this parameters and overall fitness of models and all those similar type of problems like multicolinearity, heteroscedasticity, autocorrelation etcetera etcetera, but you know if variables are not in quantitative way then, obviously, it is very difficult to, you can say calculate beta 0, beta 1, beta 2, because it is, just you know, way of mathematics..

So, we have beta 0 formula, we have beta 1 formula, we have beta 2 formula. So, accordingly we need to have some information like you know we need summation Y square, summation X square , summation Y X 1 summation Y X 2, if something is

variable, in qualitative, these are all called qualitative information. So, now, I cannot just multiply this one.

So, to multiply this one, then I have to transfer this qualitative information to quantitative information. So, the way we will transfer the qualitative information to quantitative information is called as a, you can say one way of representing, the dummy variable modeling.

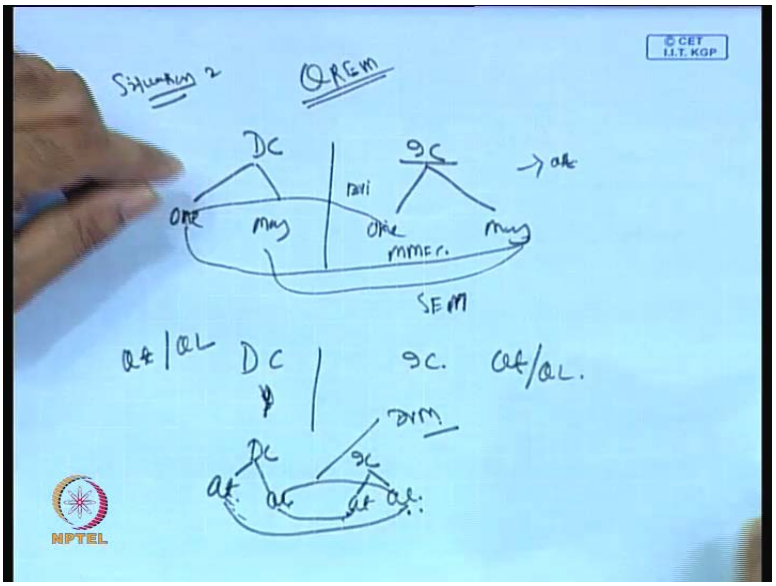
So, dummy variable means, it is a proxy variable, it is called, it is used as a proxy variables and where the variable cannot be expressed in terms of some quantity. So, it is purely qualitative variable. So, we use, the first structure is, we have to collect the qualitative information, then we transfer the qualitative information into some form of quantitative information, then we can go for its estimation.

For instance, if I will here male female items are there. So, instead of male female, male female, I denote, you know, male is 1 or female is 2 or male equal to 0, female equal to 1 then, obviously, I will transfer, instead of male, I will put, you know, I will just replace you know, control f then put to m, then transfer, replace to you can say 0 then, all m will be transfer, replace to you can say, zero.

Similarly, female, I will put again control f, you know, f then, I will replace to 1. So, then, obviously, this particular column will be transfer the entire qualitative informations to some form of quantitative information where, the information is more, more or less binary in natures. Similarly you know education also. So, there is 4 degree here. So, I will put 1 2 3 4. 1 for Msc, B Tech, B Tech 2 for B Tech, 3 for MS, 4 for PhD and accordingly, I will replace instead of all Msc, I will put 1, instead, of all b tech I will put 2, instead of Ms I will put 3, instead of PhD I will put 4. So, then you will transfer into qualitative a quantitative information. After that, estimation is as usual the standard form of the form standard process which, we have discussed long back.

So, now this, the way you will have transfer this qualitatively, qualitative variables to quantitative variable is very top class job. So, the way we will design that structure is called as a dummy variable modelling so, that means…

(Refer Slide Time: 14:41)



So, in situation 2, situation 1 which, we have discussed is more or less more or less, means, in that particular case, one of the standard assumptions is that, variables are quantitative in natures. So, here, means, the moment you will go for, this is otherwise called as, a qualitative response econometric modelling. So, this dummy modeling, otherwise known as qualitative response econometric modelling or you can say qualitative response regression modeling, qualitative response econometric modeling.

So, that means, there's some kind there is some kind of qualitative information we will see here is. So, now, this is dependent classification, this is independent classification. Dependent classification can be one, can be many; independent classification, classification can be one, can be many. So, we have already discussed these four steps, if we many many then it is called as a, structural equation modeling. If you know one, one then it is simply, bivariate econometric modelling if it is 1 many then it is called as a, multivariate econometric modelling

So, this is how the structure, we have already discussed. This is case 1 structure where all these information are quantitative,, but this same same problem can be transfer in to qualitative response econometric modelling if any one of the particular structures are few of them particular, in that particular setup is qualitative in natures, it means, the variable cannot be said as a quantitative variables, it is a qualitative variables. So, because, the way we will transfer this, you know statistical model in to estimate means, original

model to estimated model. So, you need to have information the way, the way will process this information then ultimately, we will get the estimated model

So, that how the information must be quantitative in nature because, ultimately we will, we process it in mathematically, so that, we will get the estimated value or estimated equations. That is how you need all these quantitative informations. So, computer will or software will not read all this qualitative informations. So, you have to, by hook and crook, you have to transfer this qualitative information to quantitative information, of course, there are certain variables which, cannot be expressed directly in terms of quantity, that is how the researcher, has to blame and roles. So, he has to find out some criteria, how to transfer this qualitative variable to quantitative form.

Because, most of the problems and where the variables cannot be fairly quantitative. So, there is a qualitative variable. So, in that particular context, the problem is very complex and in that context researchers rule is very tough or very high. So, he has to be, means, he has to perform very well, to transfer this quantity, qualitative information to quantity information. So, now, in the situation 2, the situation 1 which you have discussed already in that case, there are several structures again. So, that means, here we have in the situation say dependent classification and independent classification.

So, now let us assume that this is Y is on the only the one structure then, it is here you can say, in both the side it may be quantitative information it may be qualitative information, this may be quantitative information and qualitative information; that means, you will find a dependent classification is a quantitative and independent classification is a both quantitative and qualitative. It may be also quantitative and qualitative only, it may be quantitative quantitative, quantitative quantitative which we have already discuss about, quantity quality it is also quality, then quantity quality, then quality quality or quantity quality, these are all called as a dummy variable modelling dummy variable modelling

So, that means, the agenda is here is that, whether, means there there must be at least one variable which should be qualitative in natures. So, that, that, we have to we use artificially transfer in to some quantitative form. So, in that structure, means when there is a question of dummy variable modelling. So, there is a a either means there is a one dummy variables or you can say several dummy variable. If it is one dummy variable

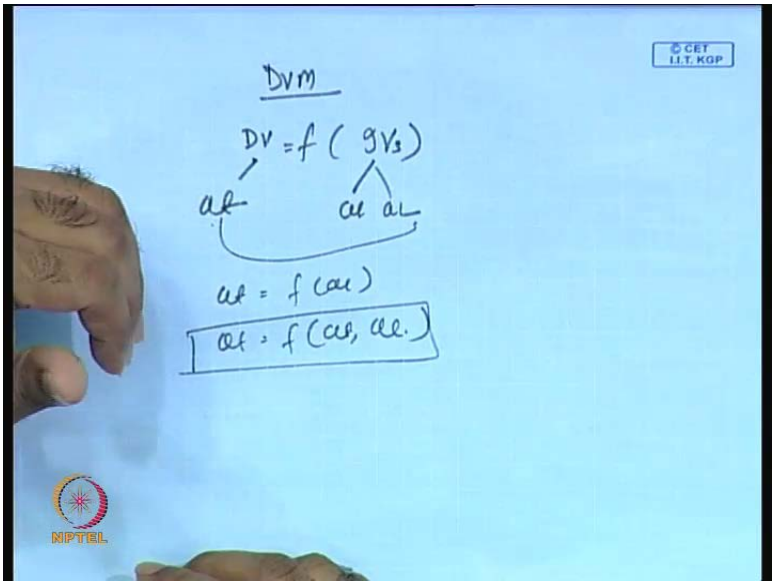
then you must be very careful whether it is a dummy dependent or dummy independent. If there are many dummy variables then, again you have to see how many are dummy dependent, how many are dummy independent.

Because, the structure or you can say mathematical setup is little bit different. of course,, the theoretical representation is very much certain,, but you know mathematical derivation is little bit different. So, as a result, so you have to very careful about it. So, that means, in the moment will we handle this type of modeling, in the first hand you have to see, what are the variables are dummy variables and the second question is that how many variables, how many dependent variables are dummy dependent and how many independent variables are dummy dependent dummy independent.

So accordingly, you have to proceed. But, we start with a simple model; the simple dummy variable modelling is that. So, there should be you know dependent variable which, is very much quantitative in nature and there is at least one you know, independent variable which, is dummy in natures. So, that means, this is the basic or easiest way to represent the dummy variable modelling. So, the complexity will start when, there is one after another dummy variable, dummy independent variable introduced in the system, the way we have discussed in the case of multicolinearity.

So, that means, the starting point is that.

(Refer Slide Time: 20:43)



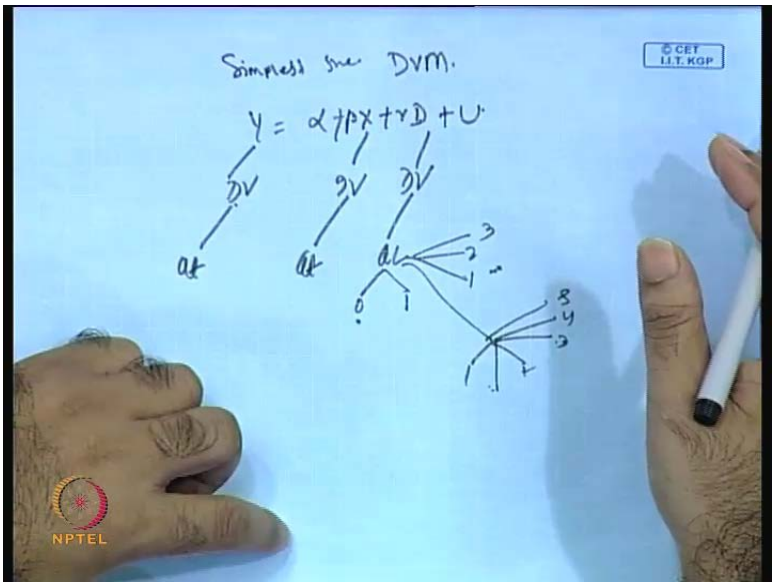
So, like this. The simplest structure of dummy variable is that, so you know, dummy variable modeling, a simplest structure is that, it is a dependent variables and function of independent variables. So, now, here we will assume that, this is quantitative information, independent variables. So, some are quantitative and some are qualitative even if quantitative, qualitative it is also. So, that means, quantitative as a function of qualitative it is or quantitative as a function of quantitative and qualitative this is also ok.

But the quantitative, qualitative this is somewhat actually, a little bit also complex, but the simplest or you know best procedure; best structure of dummy variable modelling is in that particular format. Most of the problems you will handle in this particular issue where, there is a one dependent variable which, is quantitative in nature, and another side there is at least one dummy independent variable. So, that means, in the right side few variables must be quantitative in nature and few variables are qualitative in natures. Among the few quality qualitative variable it may be one, it may be multiple, so if it is one or in the case of multiple, the structure is accordingly designed. So, we will discuss in details how is that particular setup.

So, now first thing is what are dummy variables? So, dummy variable is usually called as a binary variable which, usually use as a proxy for other variables. So, that is how the definition of dummy variable. So, there are various names related to dummy variables, you can say, binary variable, categorical variable, response variable, dichotomous variables. So, there is a so, many ways we have to represent the dummy variables. So, in any case the problem is more or less same. So, you just know the methodology or structure then, everything will follow accordingly. So, now, what is all about this dummy variable modeling?

So, the standard, what we will discuss here is, we start with, a we start with this concept called as a, dummy independent structure first, that too, you know, one dummy independent where, one dummy, one dependent variable only..

(Refer Slide Time: 23:06)



So, this where the simplest structure, the simplest structure is that, you know, Y equal to alpha plus beta X and plus gamma D. So, that means, this is the simplest structure of simplest structure of, you can say, dummy variable modeling. This is dummy variable modeling, that is independent dummy and this is dependent variable, this is independent variable, this is dummy variable.

So, this particular measurement is a quantitative, this particular measurement is quantitative, this particular measure is a qualitative. It may be 0, 1 it may be 1, 2, 3 or it may be like this 1, 2, 3, 4, 5 like this. Let us say, this is generally the best cited example. This is a seasonal effect, is another cited example or you can say, time priors, different time priors, is another cited example and this is with respect to different colors, different regions, different countries, different state. This is how it can be classified. So, that means, you see here, is let us say, I will see here, what is the sales volume with respect to you know, product say this pen?

So, now what is the sale of, this sales volume of this pen? So, that means, we have to know what are the factor which, it can influence sales volume. Let us say X, x is a one variable, obviously, the amount of sales depends upon price of sale or you can say advertising involved in this particular pen. So, that means, it is quantitative in nature. Then, there is variable called as a dummy variable. So, dummy variables, that means, we like to know, you know, when a particular company sales is forecasting, your agenda is

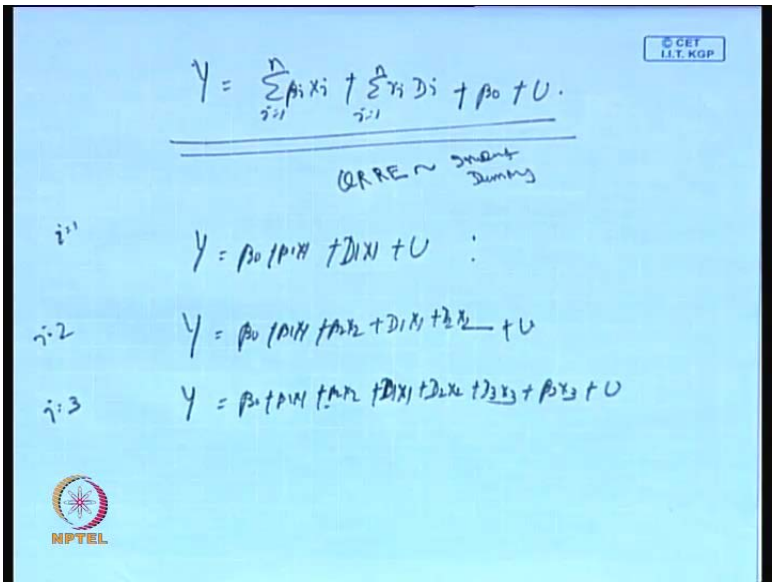
to forecast the sales of a particular company and, obviously, let us assume that your company is a company has a several branches across the country

So, that means, you like to know which particular city is more effective and more significant than the others. So, in that case, so you have to accordingly introduce a dummy variable that is you can call as a regional effect. So, now, suppose there are 5 different cities, are there your product is going. So, which particular city is more more significant with respect to your sale. So, in that case you have to use 4 dummy. So, suppose sales from south then you will put called D equal to 1 then, you know sales from you can say north you will put D equal to 2 then, sales from you can say west then, you will put D equal to 3, if sales from east then, you will put you can say D equal to 4. Like this, you have to transfer the entire sales component.

So, accordingly if will we go by, you know, estimation. So, these variables cannot be say south, north, east, west cannot be, you can say estimated directly. So, what you have to do, you have to transfer all this information to quantitative information. For instance, you just denote south means 1, north means 2, west means 3, east means four. So, accordingly, transfer all this south, west, south, you know, north, east, west in term, in term, in terms of 1, 2, 3, 4. Then, you go for estimation. Obviously, you will get a better result. So, this is the simplest structure of dummy. So, now, I will put you in the general framework of dummy. So, then we will come to a specific call.

So, the general framework of dummy is that. Before, I highlight this general form of dummy, let me first highlight here is.

(Refer Slide Time: 26:45)



So, let Y equal to, Y equal to summation beta X i, i equal to 1 to n plus summation gamma I, D i, i equal to 1 to n plus beta 0 plus U. So, this is the generalized format of qualitative response, regression modeling, qualitative response regression modelling or econometric modelling with independent dummy with, independent dummy. So, that means, it is it is the dummy, independent variable econometric modeling. A dummy independent variable econometric modelling

So, that means, here Y is always remain same. So, now, you are just adding one after another variable in the systems in the form of quantitative or in the form of qualitative . So, obviously, r square will be substantially high because, Y is remain constant and now the moment you will we introduce one after another variable whether, it is in the form of direct phase or in the form of dummy, then, obviously, your r square will approach to very high and f will approach to very high. But, in the same times, where will we introduce 1 after another variable like time series analysis then; obviously, by introduction time series analysis with a log introduction.

So, similarly here is, if you will enter X or corresponding X or we can say, dummy then, obviously, r square, by default r square will increase, f will increase. That means, model model overall fitness, will fitness, will be high or increase because, you are increasing simple model to, you know, multivariate model. So, when we will handle multivariate model then, obviously, most of the cases r square will be substantially high and you

know, f will be substantially high,, but, the main problem is that, whatever variables you are introducing all these variables will be significant. Otherwise, it is unnecessary to introduce any particular variable.

What we have learnt from multicolinearity. So, let us say that there are 5 different variables. So, 3 are you can say, quantitative variables and 2 are qualitative variable, so now, out of all these variables. So, we have to first go through original regression introducing all these variables,, but by default it will go by all these variable without to initial test, initial any testing, then, obviously, some variable may be significant and some variables may be a non significance. So, in that case,, but r square will be very high and f will be very high. In such a situation, what you have to do? So, we have to check what is the highest percentage a influence on means particular variable which is highest influence on y

So, that means, in that case you if will you follow the path multicolinearity then Y has to be regressed through this first a very high impact variable then subsequently you will introduce a second impact important variable then every time we will check it by introducing 1 after another r square must be significant r square must be very high substantially increase or adjusted r square will substantially increase f will start increasing substantially and in the same times whatever variables you are introducing that must be significant. So, yes of course,, the significance level may vary. So, it may be instead of in the first case if it is at the 1 percent in the second case it may be 5 percent or you can say it may be 10 percent.

But, we have to go up to 10 percent then you see after introducing or one after another independent variable in the system, then, two things must be you must be taken care. First thing is you, whatever variables you are introducing in the system, one after another, then, that variable must be significant and this and it should not see, effect the significant of other variables.

This is the number one important thing and second thing is that the overall fitness of model will be substantially high and adjusted r square will be substantially high, f will be substantially high. By any chance, if will we introducing one after the another variable and you know significance level of the parameter is get affected and in the same times,

you know, f statistic is also getting effected, then, it is better not to introduce that particular variable.

So,. So, one after another variable will introduce as long as r square will increase, f will increase and parameter whatever parameters involve in the particular system current system, so it has to be a it it has to be significant. So, now, when will we put here quantitative response, qualitative response econometric modelling and general format. So, this is how we have to represent. So, now, with this particular structure. So, you can go for very simple structure, you can go for very complex structure

For instance, I will, I will take you know, just i equal to 1, i equal to 1. In fact, sample size,, but, here I am putting different way. So, Y equal to 1 means, say it is nothing,, but, beta 0 plus beta 1 X 1 plus, you know D 1 X 1 plus U. So, this is the simplest structure of, you can say dummy variable. So, now, if we put i equal to 2 then; obviously, the system will be beta 0 plus beta 1 X 1 plus beta 2 X 2 plus D 1 X 1 plus D 2 X 2. So, this is how it can be extended substantially. So, if I will put i equal to 3 then; obviously, the system will be beta 0 plus beta 1 X 1 plus beta 2 X 2 plus beta k a plus D 1 X 1 plus D 2 X 2 plus D 3 X 3 and obviously, here we will have beta 3 X 3 plus U, plus U. So, this is how it will extend substantially.

So, that means, this is generalized formula. So, mathematically if will we put to allow 1, 1, then, it is this structure, if we allow, 2 2 means, 2 independent variable quantitative, independent variable and 2 qualitative independent variable. Similarly i equal to 3 means, we are introducing a system with three quantitative variables and 3 qualitative variables like D 1 D 2 D 3. Similarly beta 1 X 1 X 2 X three. So, now, you know if will we introduce one after another variable, then, one thing you must, means, two things you must be very careful. The variables which, you included in the system, it should be significant and by the way, by the process of this significance levels. So, others should not affect their significance level.

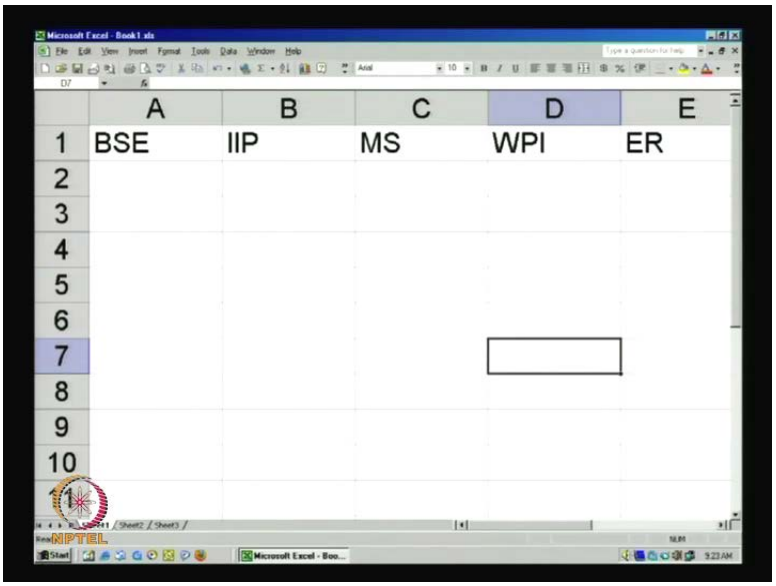
If a introducing of another variable, new variables, if the overall significance of the variable is increasing and significance of individual parameter is also increasing then you are in the right track,, but if you are, after introduction of a new variables, if the other variables are drastically in the negative side and getting effected the overall fitness of the model, then, in that case it is better to stop that variable. You should not introduce that

variable in the system. So, this is how we have to extend the structure of bivariate you know dummy variable econometric modeling.

So, now obvious question is why it happens? This particular structures and is it essential or is it that means, we like to know what is the nature and consequence of this particular modelling. So, the nature I have already explained. So, the nature means a, obviously, it is a system where few variables are quantitative and few variables are qualitative.

So, that is the nature. So, that is how the way, but regression will not effective, if a system consist of about quantitative information and qualitative information. So, of course, initially you have quantitative information, qualitative information, but, when will you go for its estimation, the moment you will enter to the estimation process. So, that times you must, you must have a file where, you know, all these informations will be quantitative in natures.

(Refer Slide Time: 35:25)



If not, then, your first job is to transfer this qualitative variable to quantitative variables, the way like this. I have given a problem which, you have discussed earlier, BSE, IIP, money supply, wholesale price index and exchange rate. So, I will put one factor with seasonal effect, let us say, there is another factor, we introduce seasonal factor or else this is the time series analysis. So, let us say, the time priors which, the information which we have right now is 1990 to 2009. Then, I like to know, what is the impact of global financial crisis or you can say what is the impact of asian asian financial crisis on

this particular inflation rate or India's money supply. So, in that case. So, you have to see what is the impact means, so, you have to introduce the dummy variables.

So, that means. So, if we say Asian financial crisis in 1996, then, you know, keeping 1996, remain constant or one level, so, you have to change accordingly. For instance, so up to, up to you know 1996, you have to put use a 1 1 or 0 0 then, other 1 you will put 1 1 so, that means, it will give you this signal of the impact of you know Asian financial crisis. Now if that particular coefficient is a significant, so that means, there is such, there is a such variation with respect to that dummy variable. Otherwise, if not significant, that means, there is no such impact.

So, if it is significant then; obviously, Asian financial crisis has an impact on you can say money supply or you can say inflation, otherwise it is not. So, just we are theoretically assuming yes theoretical assume means, there is some minor level of impact. So, if there is minor level of impact, it should not be considered at higher level, that too, at this statistical way. So, you must be very careful about all these issues. So, now, basically when will we go for, you may have any form of model. So, whether it is you know causes some modelling or time series modelling or panel data modelling or you can say regression modeling.

So, variable, two things are very important. One is variable choice, variable classification and another is the informations so, that means, so the information must be always with you, so that you can go for estimation. So, now, if the variables are some variables are proxy and some variables are quantitative in nature, that times, you have you have to do lots of things. Otherwise it is very difficult to, you can say, go to a particular conclusion or sometimes it cannot be possible to activate until and unless you transfer this qualitative information to quantitative information. So, this is the fundamental issue of, you can say, this dummy variable econometric modeling.

So, that means, it does not matter that dummy variable modeling, it can be time series problem, it can be cross sectional problem, it can be panel data problem. Obviously, the introduction of dummy may be with respect to cross sectional unit with respect to time series unit. For instance, the way we have highlighted this, is you know time series dummy. Sometimes you know, cross sectional dummy also there. Like say, you know, original effect state level effect or you can say something other way, gender effect

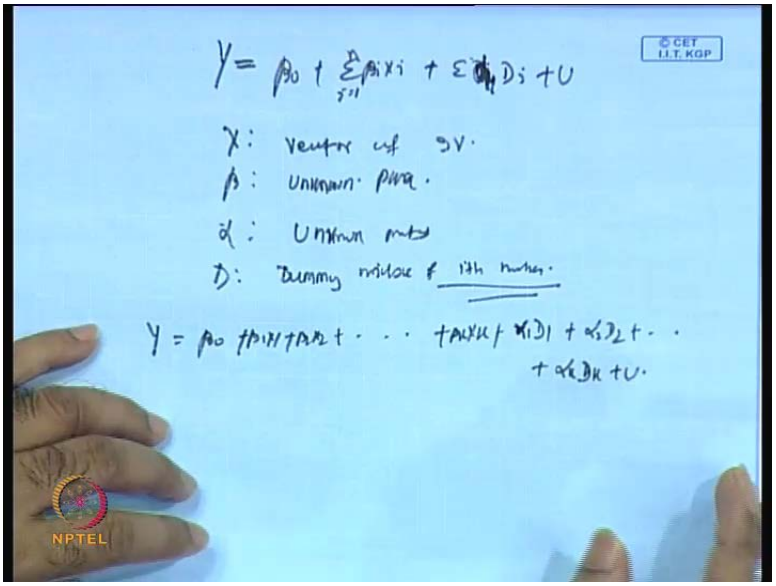
etcetera. These are the things which, we dummy variable has to be introduced and the estimation can be captured through dummy variable technique only.

So, this is how the basic starting point of this, you know, dummy variable econometric. So, that means, what we like to means, why, what are the things we like to know? That means, the variables which, you are collecting usually may be, quantitative may be qualitative. If it is quantitative then, it is very good for you. So, you need not put any extra effort or extra labour, but if it is, you know, some qualitative informations are there, some variables having quantitative informations are there, then that context you have to play fantastic role.

That means you have to design the transformation appropriate. Transformation, then after that, you have to estimate the model. Otherwise, computer will not directly read, if will we use the software. Again, it you know, go by manually or if you exam point of view, it is very difficult to handle this type of problem. Because if I will say, you know, what is the gender is you say X variable X 2 variable and you know X 1 is you can say income variable. So, I cannot just multiply these two. It is very difficult, means mathematically, you cannot multiply anything with you know qualitative information.

So, it should be very quantitative in nature. So, that means, that, that means, we have to transfer the entire structure in to quantitative problems so, that the structure can be built properly.

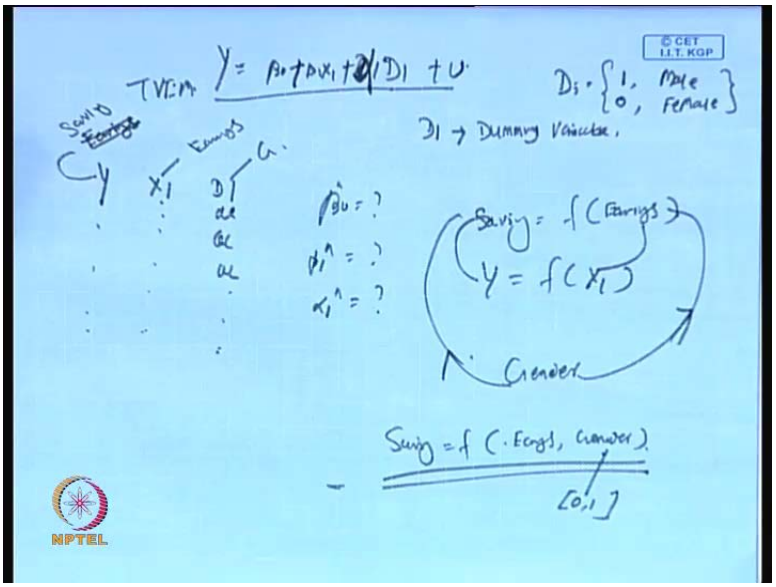
(Refer Slide Time: 40:32)



So, now, there are several structures which, we have highlighted which, we have highlighted then here. I will start with that particular generalized formula then; we will discuss a part of it. So, Y equal to summation beta 0 plus summation beta X i, i equal to 1 to n plus summation delta D i. In fact, I will call it here is, alpha i, let us say alpha i then D I, alpha i D I, alpha I, alpha i D I, this is alpha i D i so, that means, plus U. So, this is the general formalized formula.

So, where or else, let me first highlight X is the vector of X parameter variables, it is X is a vector of independent variables then, beta is unknown parameters, unknown parameters. So, then gamma, alpha is also unknown parameters then, D is the dummy variables, dummy variable of i th numbers, dummy variable of i th number. So, this is how the entire structures. So, now, if I will put in different way, in a explicitly format then, it will be coming like this. Y equal to beta 0 plus beta 1 X 1 plus beta 2 X 2 continue beta k X k plus alpha 1 D 1 plus alpha 2 D 2 plus alpha k D k plus U. This is how the general formula.

(Refer Slide Time: 42:35)



So, we start with a particular case. So, particular case is a let us say, let us say like this. Let us say, Y equal to beta 0 plus beta 1 X 1 plus beta D 1 D 1 or you can say alpha 1 alpha 1 D 1 alpha 1 D 1 plus U. This is how you have to say. Let us say here, D 1 is the, D 1 is the dummy variable, dummy variable. You remember one thing, when will we introduce dummy variable in the system, you must know, the systematic approach

behind this dummy variable or systematic way you are introducing the dummy variables. So, that means, there should be proper a step or structure as a result you are using the dummy variable. So, without having proper structure or get proper knowledge, get proper setup you should not introduce blindly. So, you must have complete or thorough knowledge.

For instance, let us say, I will means the dummy variable sometimes, mathematical it is very difficult to convince properly. So, you can convince if will we highlight what is exactly Y and what is exactly X and what is exactly dummy. So, that means, in this particular structures, in this particular structures, so we have variables Y, X 1 and D 1. These are the variables and our requirement is the beta 0 hat, beta 1 hat and alpha 1 hat. So, these are all requirement. So, we have Y, X 1, D 1 so, that means, by default it is looks like a trivariate econometric modeling, trivariate econometric modelling which, we have discuss long back.

But the way we discussed long back, it is the setup where, all these variables are quantitative in natures, so, that means, only quantitative information are available for Y X 1 and X 2. So, now, here, you have Y informations it is quantitative, X information it is quantitative, but D information which is a qualitative, qualitative, qualitative and qualitative. So, now, it is against the earlier discussion, that means, in the earlier discussion of trivariate where, only 3 variables are there and all these 3 variables have quantitative informations then, you can go for its estimations, but you know when all these information are qualitative then, obviously, it is a very very difficult to handle. So, you have to transfer it to a proper step.

So, let us assume that for simplicity, we you, we you, we would say this is Y equal to say earnings, Y equal to earnings, X 1 represents here, you can say, in common levels, in other way we will put it this way. Y equal to saving and X 1 is a earnings and D 1 is the genders. So, let us, we start with the problem is that we like to know what is a, what is the impact of a earnings on saving the function is a what is the impact of earnings sorry yes earnings impact on savings. So, that means, if I will put Y equal to function of X 1 here like this, so, that means, Y is the saving and X 1 is the earnings.

So, this is how the entire structure. So, that means, we have a earning information, we have saving information. That is purely quantitative in nature. So, we like to know what

is the impact or association between earnings and savings. So, is it positive or is it negative or if it is positive or negative? So, which cause and which effect, but we have already highlighted here which cause and which effect, that means, I think, is the effect and the earnings is cause, that means, we like to know whether earnings is most important factor which, can influence saving, saving of a particular person or particular household or you can say particular economy.

So, now saving depends upon what are the income earnings or income level of a particular person, particular state or particular household. So, it will, whatever way you have to represent. So, so, you must have saving information, if it is say, household wise information so; that means, you take hundred households information household 1, household 2, household 3 then every household you ask what is the income level total annual income then, obviously, you ask what is the total saving that is involved, saving then you start regressing. So, the moment you will regress, then it will give you signal whether, whether you must earnings has a significance impact on saving or not, but most of the instances earning has a significant impact.

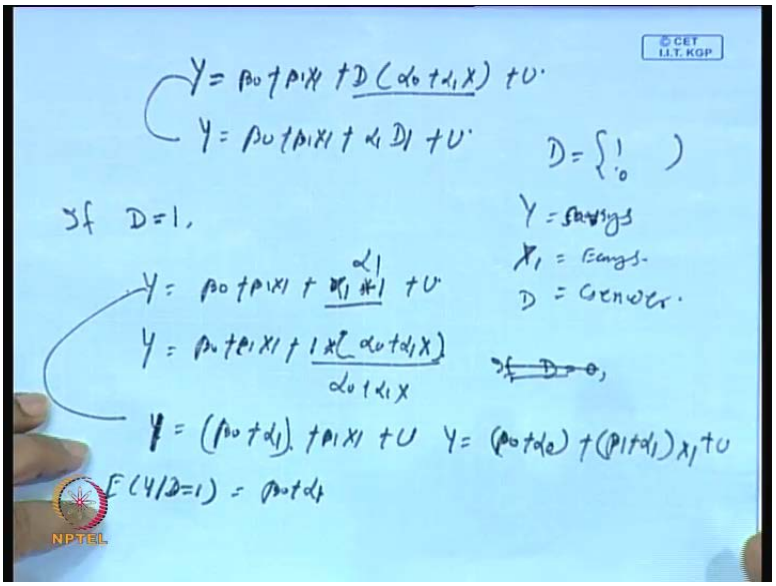
So, this all theoretical knowledge, still, you have to verify through econometric analysis. So, in that case you have to build a model. So, building model is a summation Y equal to beta 0 plus beta 1 X 1. So, that is the information with you. So, you have to just estimate properly. So, this is one way to handle this issue. So, now, I will add another variable in the systems. So, what is the gender impact on you know saving in this particular setup. So, that means, I will introduce another variable say gender. So, what is the gender impact on saving? So, that means, a gender will be go to independent side, right side then and we like to know what the gender influence on saving is.

So, obviously, if I will integrate then this saving is a function of earnings and genders earnings and genders. So, that is how we have to represent. So, now, the moment we will put earnings earnings in gender, as you know factors which, can be considered as a consider for the influence of saving then, we we will be highlight the model structure is like this. So, now, here if will we introduce, you see, the dummy variable by name or you know by indication will give you the setup, the setup dummy in the dummy variable the setup is very important factor.

For instance here gender is there. So, gender means, by default, you must have theoretical knowledge to bring the setup. So, gender means there is no problem at all. So, you know gender, how many ways gender can be represented. So, here gender can be represent in two different ways, so male and female, so obviously, so in that case, so gender range will be 0 and 1, gender range will be 0 and 1. So, that means, we will fit this particular models, so where, D i is a proxy of 1 and 0, say 1 is for male and 0 is for female. So, this is how, the constant, I have to add in this particular system then, I will go for its estimation, otherwise, it is not possible to go for its estimation.

So, I will, I will, I will, I, in fact, I will highlight in other way round.

(Refer Slide Time: 50:12)



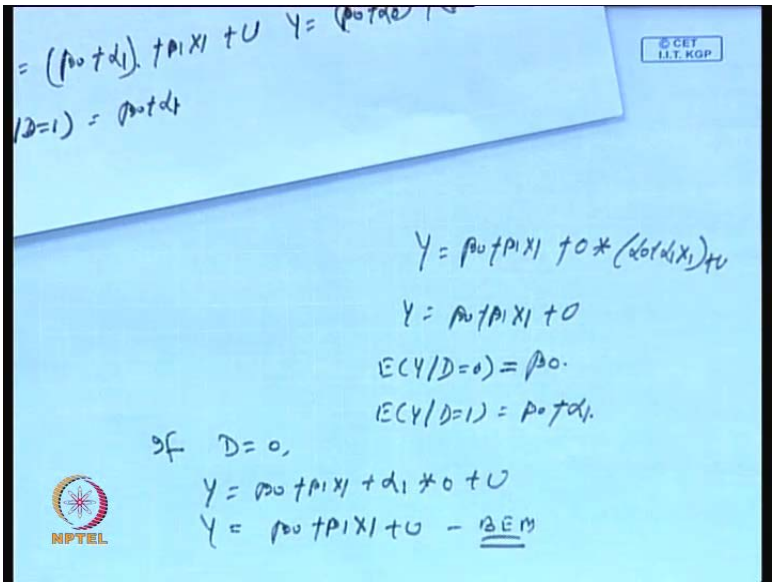
So, let me highlight here once again. Here is Y equal to let us say beta 0 beta beta 0 plus beta 1 X 1 plus D into alpha 0 plus alpha 1 X plus U. So, this is another way we can represent also. So, we have, we have written here is first format beta 0 plus beta 1 X 1 plus you can say gamma 1 D 1. So, now, I am putting like this way. So, this is another type of models, which is used as a dummy. So, anyway, so now, here D is recognized as a 1 and 0 and Y is the earnings. No, Y is the saving, Y is the savings and then X 1 is the earnings and D is the genders, D is the genders.

So, now if D equal to 1 then, the model will be structured as a Y equal to beta 0 plus beta 1 X 1 plus d. So, gamma alpha 1 into 1 plus U. Otherwise, if you will put in that particular format then, it will be Y equal to beta 0 plus beta 1 X 1 plus D equal to 1, 1

into 1 into alpha 0 plus alpha 1 X. So, this is another form of models. Then you see here, is if D equal to 1 then, this this this will come like this. So, now, what you have to do. So, you simplify. So, Y equal to, that means, this is alpha 1 and this is simply alpha 0 plus alpha 1 X, so Y equal to Y equal to beta 0 plus alpha 1 plus beta 1 X 1 plus U. So, this is how. So, E of Y of 1 you can say D equal to 1 is equal to simply beta 0 plus alpha 1.

Similarly, if I will put D equal to 0 then, similarly if I will simplify this one, then it will be Y equal to beta 0 plus alpha 0 plus beta 1 plus alpha 1 into X 1 plus U, is it ok.? This is how the structure is all about then, I will put another another structure here is.

(Refer Slide Time: 53:21)



So, now, for this I will put it here only. So, now, if if if D equal to 0 then, obviously, Y equal to beta 0 plus beta 1 X 1 plus alpha 1 into 0 plus U. So, it will be equal to beta 0 plus beta 1 X 1 plus U. So, this is simple bivariate econometric modeling, this is simple bivariate econometric modelling and in other way if will we put like this way then, Y equal to beta 0 plus beta 1 X 1 plus 0 into alpha 0 plus alpha 1 X or X 1, right.

This is in fact, X 1 right, this is X 1 plus U. So, obviously, so Y equal to simply beta 0 plus beta 1 X x 1 plus U, so E upon Y where D equal to 0 is equal to simply beta 0. So, that means, the the otherwise E Y upon D equal to 1 is beta 0 plus plus alpha 1 alpha 1 or yes alpha 1. So, now, I will this is the case where D equal to 0 this is the case where D equal to 1. So, that means, the influence factor between male and female is the alpha one.

So, that is how the structure of dummy variable, that is the structure of dummy variable. So, dummy variable will give you lots of hints or indication how a variable can be considered as a proxy variable and how that proxy variable can be used to estimate that particular econometric modeling.

So, that is how it is a very interesting technique. So, that means, you start with a variable called iterative way then, you transfer in to quantitative way then, you will have to go for it is estimation then, you come back to original position then, you highlight what is the particular gender means, in this particular case, gender impact on that particular you can say savings and earnings. So, this is how the dummy variable is you can say examined or you can say utilizing the econometric modeling. In the next class we will discuss the detail setup of, you can say, econometric modelling with a some examples and also we will discuss the other part of the dummy modeling, that is you can say, a qualitative response dummy dependent econometric modelling. So, with this we will conclude this session, Have a nice day. Thank you.