Econometric Modelling

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

Lecture No. # 30
Panel Data Modelling

Good afternoon, this is doctor Pradhan here, welcome to NPTEL project on econometric modelling. So, today we will cover panel data modelling. So, in the last lecture couple of lectures, we have discuss dummy variable modelling that to dummy dependent and dummy independent, again a single dummy and multiple dummy and also interactive we felt. So, many way has we discuss means say in this way dummy independent there is a one dependent case, multiple dependent case and is interactive case, similarly in the dummy dependency case.

So, we have discussed engineer probability model, binary chase models, logit model and probit models. And we have also discuss various differences between all this models. Infect we have no discuss a practical problem, because if our luck of time a in the mean times, we will discuss the same thing in the panel data model. Because it is too much integrated with a penal data model, and in that case we will discuss something application and will a touch the data dummy variable modelling also.

So, once will discuss the panel data model, then obviously we can also get touch with dummy variable modelling. So, because it is a means without knowing dummy modeling it is very difficult to handle panel data models. Because it is the basic means panel data for understanding the panel data, dummy variable modelling is essential, so until unless you have complete knowledge on dummy variable modeling. It is very difficult to understand dependent data modelling.

So, last couple of lectures, we have already briefed about the dummy variable modeling, that to dummy dependence in dummy dependent variable and dummy independent variable. So, today will discuss the panel data model, and we will also cover how the dummy can be use in the regression analysis, and also highlight the issue of dummy. So, before we go to panel data modelling.

So, first up all what is panel data. So, in the last few lectures or you can say form the beginning. So, we have discussed the data analysis only. So, that to you know if talk about bivariate or trivariate or in variate or whatever may be the case or you can say multivariate. Are even, if the problem aspects like multicolinearity heteroscedasticity autocorrelation dummy variable modelling etcetera.

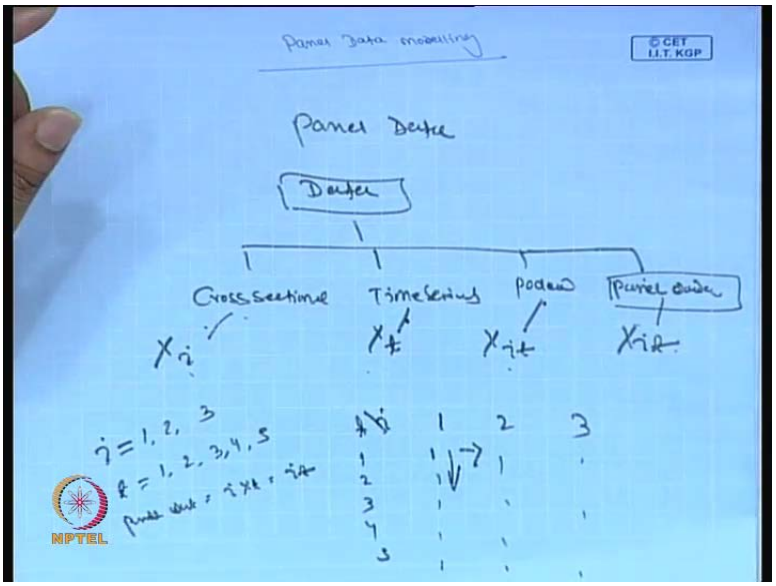
Everywhere so, every problem is associated with data without heaving data. It is very difficult to understand particular problems or we cannot apply it is mean so, per as application is concerned. So, without having particular setup of data it is very difficult to understand it is you know. So, first up all what is data? Data means it is actually collection of information.

So, now once you have some information, then you know that information may be very quantity in nature, in may be very qualitative in nature in may be in both. So, that we have very much you know clearly identify or you can say categorize in the case down variable modelling. So, we mean after dummy variable modelling, we usually handle the quantity form of data. But you know during the dummy variable modelling times show, we have discuss both the quantity aspect of data and qualitative aspect of data.

So, here we will discuss same data you know modelling what with a different setup and different structures. So, basically data means collection of information's. So, collection of information with respect to what so, that is how the basic data setup is a started. So, collection of information, with respect to three aspects time aspect, cross sectional. That is individual unit aspect and also integration of both.

So, that means, data can be collected with respect to individual observation that is what we call it cross sectional unit. Data can be you use under panel you know time series analysis and data can be collected under both the hats. So, like this panel data.

(Refer Slide Time: 04:40)



So, first up all will discuss what is paneled data? So, first we start with the data the data classification then will go to the penal data. So, data is a basically divided into or I can quality for different angels, we can classify one is called as a cross sectional, then time series, then pooled data, then panel data.

So, you know cross sectional means data available with different cross sectional unit that is called as a cross sectional data. So, this is usually you know indicate with i subscript then time series means data is collected over a period of time that is usually represent by subscript you know TS. Pooled data means its combination of both cross sectional and just we are pooling cross sectional and times series data.

So, this is i t representation and panel data is also pooling or you can say globing all this cross sectional times series unit it is also i t, but between i t you know pool data and panel data there is a difference. So, what is panel data? Panel data basically means data collected in cross section, but they are observing periodically. So, means data can be collected a in particular individual or cross sectional units, but can be observe periodically means with respect to time observations. So, that is what is called as panel data setup for when pool data is concerned.

So, we are just pooling times series cross section, times series cross sectional data, but we are not discussing the inside pictures in between globing the times series cross sectional data for instance. So, when will glob a times series with you cross sectional unit

is times series units, then definitely there is a some kind of error in between. So, the variation will start and that variation will affect the individual parameters as well as the error component so, as a result. So, there is a some kind of you know problem is a happens means a it can be raise theirs to handle the you know pooled data.

So, basically with the you know whether you will use purely time series data, cross sectional data or pool data or panel data absolutely depend on your you know objective specification. Suppose, if you are objective is not study about the variation of individual aspect and time series aspect.

Then obviously, you go had with the pool data, pool data is a just like a simple regression analysis like cross sectional, cross sectional observations and times series observation it is just your globing. So, that means, a data features are not there just your globing and increasing in the sample set that is all. But the moment you will useful data there is a lots of limitation associated with you know the calculation involvement.

So, now basically cross sectional, we will indicate i observation times series t observation and pool data you know globing of times series cross sectional unit. And panel data you know it is also cross globing of cross sectional and times series data. And also in the same times, we are studying the variation of individual units and also individual or you can say group units and also time components.

So, now for this is i t i t i t or this is subscript, but if you have part of for a particular variable then it becomes x I, then x t, then x i t, then this is x i t. So, that means, here is so, there is a information means. So, we are targeting variation of intercept term and also we are targeting the times series term, times series effect.

So, that is how it is called panel data that means, panel data is means broad concept and you know it is complex problem. It is for a it is most practical more you know feasible in the case of handling the you can say regression modelling or econometric modelling. So, now, what is this specialty of you know panel data modelling. So, that means, we can we can solve a particular problem through cross sectional analysis, you can solve a times series analysis you can solve the time series analysis, we can solve through pool data analysis.

So, now the question is what is this you know panel data it everything is you know observe very easily then y will go for panel data. Because yes pool data is a part of the panel data, but pool data we are not just considering the problem aspects of a you know pooling cross sectional times series. Just we are pooling and you are increasing the sample size. So, the moment your increasing the sample size. So, within that a you know integration, there are certain problems which will take care through intercept concept and you can say time series concept, that means, cross sectional issues and time series issues.

So, now what is the first up all the, what is the advantage or what are the learning lesson will lesson from the handling panel data modeling. First up all there are basically three specific factors through which we like to enter panel data modelling or we like to learn panel data modelling. So, first thing, so, we like to observe a particular variable with respect. It is seen individual you needs and with respect to it is a you know periodically issue that is means times series observation for instants lets a there are three airlines 1 2 3.

So, now, we like to know what is there production capacity or you can say cost involvement with respect to three airlines a with three different time periods say or you can say five different time period say. So, that means, a i is a here represented as cross sectional units i stands from 1 2 3 like this i stands for three cross you know times. Then I will put 1 2 3 and five different which time periods will take 1 2 3 4 5. So, this is how will represent. So, that means.

So, if will glob then for panel data, we will glob this 2. So, this is i into t. So, this is i t observations. So, this is i t observations so, that means, we will have a fifteen observation in this particular setup. So, that means, if will go by cross sectional analysis then you have only three observation, it will go by times series analysis then you have go, you have only a five observation.

So, now when will glob all these two then you will get 15 observation, that means, in the very beginning we have mention typically, we have highlighted for any specific modelling. So, your sample size should be substantially very high. Higher the sample size, higher is the accuracy of the model or higher is the feasibility of the model. If the sample size by any chance it is low or you can say it is go against the system for instants

there variables are very high set an and number of sample is (( )) then, obviously, the system itself is inconsistent you cannot solve the problem.

So, if there is proper a means, you have to for a particular model, you have maintain the upper accuracy of sample size or data structures without having, such a clarification clarity, it is very difficult to handle the very difficult to go for estimation. So, before going to estimate particular models your sample size and model specification must be very feasible one.

If this is system is inconsistent then, obviously, you know the estimation will be go other way around, that means, whatever estimation results, we will get all the estimators will go against the blue theorem. So, that is how a these are all means you must be very careful and must integrate in the econometric modelling systems.

So, in this panel data modelling so, we are just observing data you are collecting the data at a cross sectional label and observing the same mean periodically. So, that means, with respect to time series that is how a you know a it is very interesting. So, that means, one of the special advantage of panel data is it will observe a particular variable with respect to it is cross sectional dimension and it is time series dimension that means it is within the effect and between the effects.

So, this is how, we have to observe for instants i equal to 1 2 3 means three different airlines then five different time periods. So, now, for i equal to once and first air lines will like to know it is impact over the years, what is it is a cost involvement over the year is similarly, like this you see here is. So, if will put i 1 2 3 here is then t 1 2 3 4 5 so, that means, here is you will here observations you have will a observation.

So, now I can go for this way we can this way. So, this is how it is called as a within effect between effect. So, this is one of the special advantages, we will observe in the case of panel data setting. Second observation is that it will, obviously, by default it will increase the sample size, if you go by only cross sectional analysis or if will go by times series analysis. Then the sample size is will be very small, but if will be integrate with proper you know panel setup. Then, obviously, it will be give you high you sample size and that is more practical and more reliable for the econometric modelling. So, this is the second advantage, we have to use for panel data setting.

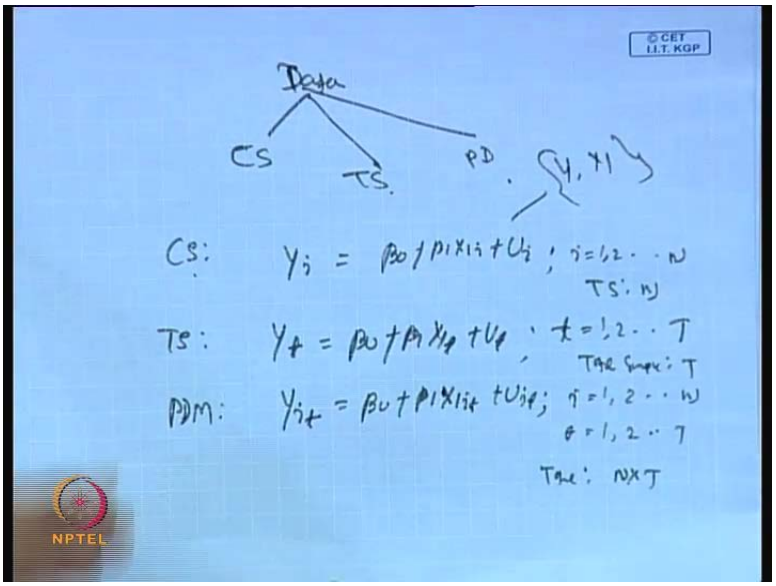
So, third advantage that we need special treatment for handling the panel data analysis. Because the moment you will glob you know i with t in between there is a lots of problems you know variation may be you know it can be raise. So, that is how you have to be careful how you to handle all this problems to before integrating cross sectional observation time. So, these are the three major issues, we have to you had before you wanted into the panel data modeling.

So, now what is all about this panel data setting, generally before I go to this particular setup? So, I will briefly highlight what is all about the modelling structure under this you know panel data modelling. So, we have already mentioned. So, data is specifically divided into three parts time cross sectional time series and penal data.

So, panel data pool data is. In fact, same you know there is no means typically when will handling penal data then there is no issue of the pooling concept. So, it will come in between automatically .So that means so, we will have the three different structures cross sectional times series panel data. .So, obviously, there is a pool data so, will go to go to that aspect separately.

So, when will discuss about separate aspects of panel data, that times pool data as image or you can say as an impart right know for comparing, you know cross sectional tome series with pool and obviously pooler panel is same group. So, if will study it is variations of individual aspects time series aspect, then obviously, it will be penal data otherwise it is pool data. So, then, obviously, there are three forms of the data. So, times series cross sectional and integration of time series cross section. So, that is how is called as a panel data.

(Refer Slide Time: 16:19)



So, there are three accordingly, there may be three models we can develop for cross sectional. The model will be means if this is system will be with respect Y and X then we can write like this Y equal to beta 0 plus beta 1 X a X 1 i a plus u plus u i. So, here i equal to 1 to up to n. So, this is a how is cross sectional modelling. So, the system will be here y upon a x 1 and we are setting with a cross sectional observations.

So, keeping time remain constant, that means, we are not using a anytime aspect. So, we are just integrating cross sectional items on this. So, time series the model will be instead of Y i we put Y t. So, which is nothing but but beta 0 plus beta 1 X 1 t plus u t, t equal to 1 2 up to capital T. So, this is how the time series setup all about.

So, that means, when will set the time series. So, the variation will start with respect to time keeping a cross sectional it in remain constants. So, we are not bothering the cross sectional unit. So, just like you know three are lines, we are just targeting i equal to 1 or i equal to 2 or i equal to 3. Then we for each particular component say i equal to 1, then we have to see what is the variations or what is the structure within the that particular system when t equal to 1, t equal to 2, t equal to 3, t equal to 4, t equal to 5.

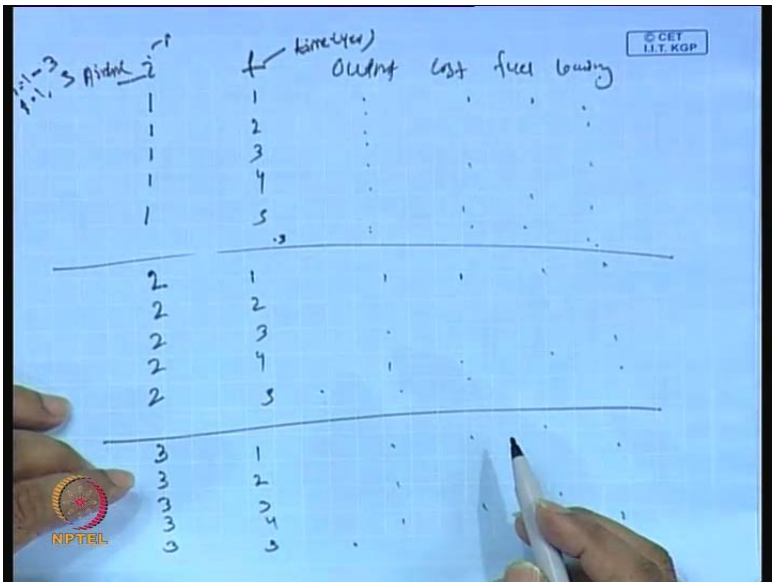
So, similarly, for i 2 i 3 you can analyze, but there is no connection between i 1 to i 2 and i 1 to i 3. So, that is what the time series all about similar in the case of cross sectional we are discussing row wise. So, how what is the variation i 1 i 2 i 3, but we are not discussing what is the a variation of i equal to 1 2 3 with respect to t 1t 2 t 3 t 4 t 5. So,

this is how it is constant. So, we either using i equal to 1 2 3 with respect to t equal to 1 or i with respect to 1 to 3 and that has to integrated with a t equal to 2 only then the impact means the variation of i 1 2 3 with respect two t equal to 3 like that means, it a particular year. So, we are discussing a cross sectional variation so, for cross sectional observation.

So, this is how time series modelling all about here times reach modelling. So, this is pear cross sectional modelling. So, now, then there is called as a panel data modeling. So, in the case of panel data modelling so, our model will be integration of i t. So, Y i t equal to beta 0 plus beta 1 beta 1 X 1 i t plus u i t here i equal to 1 to n and t equal to 1 to t and total sample is a N into T.

So, here total sample is a T and here total sample size is a N, total sample size is a at you know N for cross sectional modelling total sample is a capital T, for time series modelling and for panel data modelling a total sample size is N T. So, this is how the panel data mean this is basic of this panel data structures. So, now how do I represent this panel data setup? So, let me highlight here the structure of panel data.

(Refer Slide Time: 19:57)



So, generally I have mention how you to setup this panel data structures for instance I have already mentioned for i and I will put it here t, then you know let us say the variables say output then cost then let fuel well amount loading can capacity. So, these

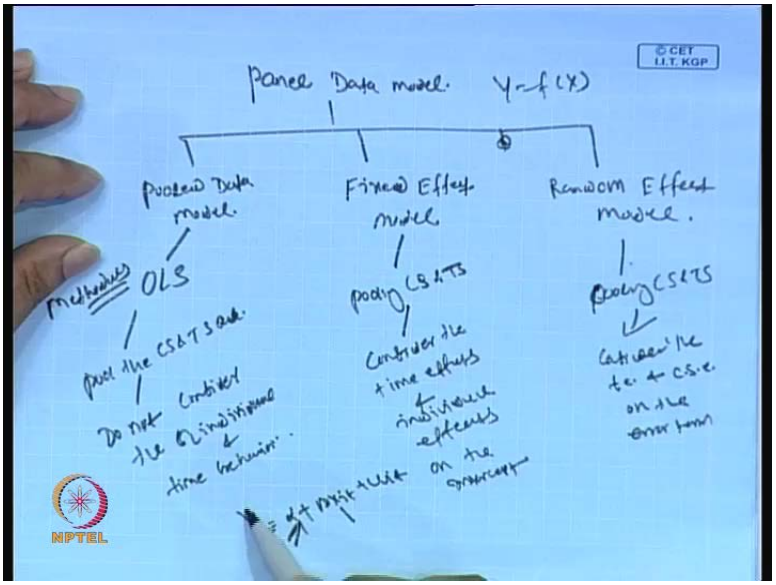
are the observations. So, we have taken i equal to 1 2 3 N T, we have taken 1 to 5. So, then how will represent the, the sample setup.

So, this is how this is very interesting. So, i is there t is there. So, we are integrating i t how is the sample setup. So, let us see is. So, for you know this is for airline, i for airline and t for time our it is, year wise analysis so, then, output with respective t and i. It cost with respected t and i t 12 with respect to t and i and loading is with respect to t and i. So, now, how will setup the decision? So, let us a i equal to 1then it is 1 1 1 11 1 so, t equal to 1 2 3 4 5 so, 1 2 3 4 5.

So, this is one observation. So, accordingly it will be analyze. So, keeping i equal to 1 t is varying here. So, now, when i equal to 2 let us say 2 2 2 2 2 then obviously, t equal to 1 2 3 4 5. So, accordingly it has to be filled up then similarly. So, when i equal to 3 then 3 3 3 3 3 3 then 1 2 3 4 5 so, this is how proceeds. So, similarly, when i equal to 4 i equal to 5 then obviously.

So, now here we are getting 5 5 5 then 5 plus (( )) a if will add then you will get more number of observations. So, this is how this is term will operate so, that means, this is how the structure of the panel data modelling. So, now, with the basic frame work can panel data setup. So, will discuss the, you know the accurate estimation process of panel data modeling.

(Refer Slide Time: 22:17)



So, how do I start with this panel data modelling let see here? So, now, panel data specifically, as a three different angel. So, one is called as pooled data angels then another is called as a fixed effect models, it is pooled data model. Then it is fixed effect model then there is called as a render effect models. So, there is another model called as a random effect model.

So, that means, we have specifically highlighted a four different structure of data setup a time series cross sectional then pool and penal. So, surprise a panel data is a concerned. So, we are talking about the pool component and the panel component. So, we are not very much treat about the time series, individual time series or individual cross sectional.

So, now here we are just pooling and we are observing the issues. So, now, for pool data modelling so, the sample setup is we can use in simply ordinary list square method. So, for as a methodology is concerned so, means you see here is.

So, our m is to have the data whether it is a times series data or processional data or you can say panel data or pool data. So, now, once you have a data then; obviously, we have the variable information because without having you know, what is the data analysis structure? Data analysis is the middle process of the econometric modelling. So, initially we will start with the theory, then we you design problem then with respect to problem, we have to specify of objectives.

And with respect to objectives, we have specify hypothesis and with respect to objective hypothesis, we said the model and once we will said the model, that means, it is mathematical form of the model since hypothesis there. So, accordingly it will transfer to statistical form of the model. So, once we have the statistical for model for instants let say there are variables y and x. So, now this is y as simply as a function of X like this, let us say Y is simply function of X. So, now, this is implicitly formats.

So, I will equality purely mathematical model. So, now, you know you before you go for proceed further. So, you have to put this model in explicitly format. So, for as corrects specification must be concern for instants in the models here we have discuss. So, these particular models so, we have discussed the cross sectional effect times series multivariate panel data modelling system.

So, now when there is a you know mathematical form of the model then error must not with their then, obviously, we have to just explicit, we have to put the explicitly format of a just y equal to alpha plus beta x. So, then similarly, we have to apply i subscript or t subscript then also i t subscript. So, that is not a problem so, that means. So, you have to correlation with the objective hypothesis then mathematical form of the model then you to transfer into statistical form of the model by introducing the artificial error term and final, we have to minimize the error term get the best fitted model.

So, this is how the usual procedure all about. So, now when you feed the statistical form of the model then, obviously, you need to estimate statistical form of the model. So, for that you need information or data. So, with the help of data or information you have to process or estimate the model to get the known means, we to understand the exact when nature of the relationship among the variables.

So, for that we need to have a information and it is process. So, when you need to have information process and obviously, the system will be vary system will be very efficient with the information is especially very high. But most of the problems in the real wise real life situation it cannot be possible to have a huge sample problem, sample size. So, what have to do? So, since it is a problem or constant for econometric modelling. So, econometrician has develop the concept of panel data.

So, that a cross sectional times series can be glob and that can be used for econometric modelling to verify particular statistical models. So, the way will integrate then, obviously, that structure is called as a panel data modelling. So, when will integrate then there are several serious issues are coming. So, that issue we have to take care very carefully.

So, now accordingly so, when will glob all this variables means data with means that is with respect to cross sectional times series and obviously, you will be find there are two forms of models. So, one is called as pool data models and another one is called as a pure panel data models. So, when there is pure panel data model. So, this pair pure panel data model is further divided into two aspects that is called as fix effect model and random effect model. So, I will know highlight what is fixed effect model and random effect model.

So, for as OLS is concerned pool data model well is concerned this particular structure is just you know, it just pools the cross sectional and times series data. That do not consider any you should not any consider any aspect of cross sectional and times series, we have issue, that means, you do not consider or there is no consideration of the individual and time behavior individual that is cross sectional and time behavior.

So, that means, you know if there is no such consideration of individual and time behavior, then you know you just pool the data and as usual the application of OLS in the case of cross sectional modelling and times series modelling separately. So, you can also float this same apply this OLS to this pool data analysis. So, it is very simple, very easy, very reliable, but there is inside there is lots of complexity. So, that will complexity can be a you know evaluated you can say investigated through the panel data modelling that two fixed effect model and the random effect model.

So, now, in the case of fixed effect models so, there are what is there? So, here means we are just pooling first condition is a satisfied pooling cross sectional and time series data. And in the same times you have you means you are considering the time effects and individual effects, we are considering time effects and individually effects on the intercept.

So, that means, so, when you are using panel data model. So, you have means you have three different formats. So, one is called as pool data model then fixed effect model and random effect model. Pool data model in the case of pool data model, we are just pooling the data without having any connection with the individual aspect and times series aspect and for that region will OLS is the technique, which is usually use to estimate that particular model and to have the parameters.

When will go for fixed effect model and that times, we are just pooling the cross sectional and time series data. But in the same times, we are observing the time series effect and individual effect on the panel data. That means, this is in the case of fixed effect model. So, the individual infect and you know time series infect is studied to intercept.

So, there is term called as a intercept, we have set Y equal to alpha plus beta X plus you i t. So, alpha is the component where we will make the variation. So, we will apply the variation with respect individual aspect and time series aspect with the intercept concept.

So, you know like this, when will set a simple model Y i t equal to alpha plus beta x i t plus u i t.

So, now when we study the variation or effect of individual aspect or time series aspect with respect to you know intercept that to target here only. So, that means, it may be alpha i or it may be alpha t. So, this is how we have to observe or separately you can say I will create another parameters and alpha i and you can say gamma t. So, this is how the effect of panel data can be observed. So, that means, when we are observing the variation of individual aspect and times series aspect on intercept then it is called as a fixed effect models.

But, random in the case of random effect model is something different you know here means we can make a guess here is that the moment you will, we have the variation impact on you know this intercept. That means, when times series variation and cross sectional variation is observing with the respect to intercept. Then obviously, that impart will directly go to or in directly go to the error components.

So, as a result error cannot be constant. So, error will vary so, that means, how it is connected. So, first pool data there is no connection with the you know intercept. So, intercept remain constant. So, now, when a means, we are artificial keeping constant you know the time impart and cross sectional. Because time is a dynamic and also cross sectional is dynamic, that means, it is not statistical just like i equal to 1 or i equal to 2, but t equal to 1 or t equal to 2 here i varies and t also varies. So, that means, the moment it will very then there is lots of variation or you can say further infect.

So, now that further impact is observe through error intercept and that the moment you observe that variation in the intercept that is called as a fixed effect model. So, when we apply that variation on you know intercept and that fixed effect model. And obviously, that a further impact will go to the error terms. And when will be observing the error variance with respect to time dimension and cross sectional dimension than it is called as random effect model.

So, that means, the usual procedure is a you will go with pool data then fixed effect model then random effect model, but is null fixed effect model and random effect model are the two you can say, pure panel data. Model pool data is a not pure panel data model because we are not targeting any variation of times region cross sectional observation.

So, here we are studying you know, we are increasing sample size that is very handy for the estimation process. In the same times, we are studying the variation with respect to cross sectional observation and times region observation.

So, in the case of random effect models so, we are pooling again. So, you we are pooling cross sectional and times series data and consider the time effects and you can say a cross sectional effect, one the error terms. You see here is so, now, we are we initially we have this much of connection.

What we have long from an econometric modeling whether in it is Bivariate, trivariate modeling multivariate. So, ultimately target is y and x dependent to an independent, but you know when will target dependent to independent that times. So, there are certain you know there may be certain problems. So, one such problem is the having in data sample size. So, if sample size is a low then you cannot go for estimation.

So, to solve that particular problem so, we are applying or econometrician usually use concept called as panel data. So, the moment will use, the panel data will increase sample size. What in the same times there is some an additional problem. So, that to means additional problems; that means, a observe with a intercept and with it can be observe error term because X you know between Y and between Y and X t there is other terms are left alpha and Y t.

So, Y i so, that means, the connection will go to this side in connection will go to this side. So, if it will come to these sides this is called as a fixed effect model and if this will go to this side it is called as a random effect models. So, now, that means, a it is the dummy variable which make the difference between a between the 60 effect model and random effect model. The dummy variable which we have discussed in the last couple of lectures here, the dummy means here it is the dummy independent clusters.

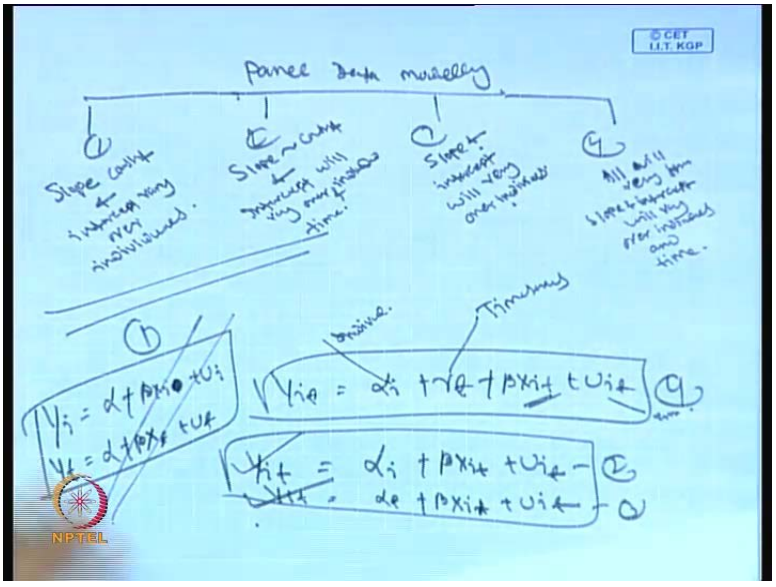
So, that means, here we are using dummy with respect to cross sectional observation and we are you are using dummy which respect to time series observation. So, that means, we are very much interested to know as a cross sectional point of view which particular component is more efficient, more effective, more proactive an another said with respect to time. We like to know which particular is more efficient, more practical and more stable one. So, this is how the enter theme is the all about. So, means that is the enter structure of panel data modeling.

So, now, we will briefly highlight exact difference of fix effect model and random will be because pool data once will go for pure panel data modeling, that time pool data is not in you are a territory. So, it is some of what out of the state territory, because we ultimately together means. We are very much interested when will go for pooling the cross sectional time series data then what are the other imparts will get?

Obviously, the advantage is that we have more and more sample size and we will get the means our estimated model will be more practical, more feasible and in the same times the parameter will follow the blue theorem. But in the same times there is some times of additional problems. So, that means, additional problem is this intercept will not constant and error component will constant what you remember when we are handling the direct OLS techniques. That time, we are assuming that the intercept is constant then error term is constant and also the slope is concerned.

So, there are three forms all together. So, there are all together three form to observe the panel data settings. So, that means, you see here is. So, when will go for panel data settings, we will together find three a means four different setups together. In fact, to we have already mentioned here three different setups. So, within three different setup, two different setup is more practical and more feasible, but above all there are four different is we have to represent the panel data setting.

(Refer Slide Time: 38:16)



So, what are the structures you see here is, this panel data modelling is like this. So, it has four different connections, so,1 2 3 and 4. So, this is how 1 2 3 4. So, what is the panel data all about? So, panel data will be general format is like this Y i t equal to alpha i plus beta X i t beta X i t plus u i t. So, this is how the usual panel data is represented or you know when will observe time series component then, obviously, Y i t equal to alpha t plus beta X i t plus u i t.

So, this is one format or sometimes what happens so, we can use the third format. So, Y i t equal to alpha i plus you know gamma t plus beta x i t plus u i t. This is actually complete form of the panel data models. So, that means, here it is the individually impact. It is observe by individual impact and it is observe by time series impact and it is observe both the case error component.

So, this is as usual the independent variable and this is the dependent variable. So, now, you see here is some time is when keeping constant and making target on time series and model will be like this. So, now, I am keeping this one constant and the felt targeting this one, then it is called as a means, it is cross observation of cross sectional only. So, that means, if will be a together then obviously, but in addition to that then there is model called as y i equal to alpha plus beta X i t plus u i sorry beta X i plus u i. Or we will put Y t equal to alpha plus beta X t if plus u t.

So, this is another format. So, that means, this is 1, this is 2, this is 3, this is 4. So, there are four different steps of the models when we are talking about the panel data modeling. So, that means, in one case it is just like cross sectional modeling. It is time verses time series modelling. So, there is no connection with time series impact. When will there are observing cross sectional and when we are a observing time series, no connection to cross sectional and the specialty these two model is that.

So, here we are assuming the alpha component and means intercept and error component and the slope coefficient are all constant. So, that means, alpha is constant, beta is constant and error is a constant. It is not a dynamic.

So, now, but we can make it the system dynamic. The way we have discuss you know heteroscedasticity issue and the autocorrelation issue because with once we have error term then, you will create a additional error terms with respect to time series, with respect to cross sectional. So, you know if that is not your objective then the model will

be very much constraints. So, that means, alpha is constraints and this slop is intercept is constants. Slope is constraints then error term is also, me what constant.

So, then that is one setup models. In other setup models, intercept will vary with respect to cross sectional unit keeping time remain constant. And in other case, time will be change with a respect to different time observation where cross sectional will be remain constant. And in the end finally, both will be go together. So, means both intercept and you know means both cross sectional time series will be together; they can vary.

So, these are four different step of the panel data setting. So, let see here is. So, now, in the first case a slope constant. So, slope constant and intercept vary our individuals. So, first is a slope constant and intercept vary our individuals? So, that means out of four different situations, we are not discussing this particular aspect because this is not at all in are boundary of panel data.

So, this is purely you know purely cross sectional aspect; individual cross sectional and individual time series domain. So, there is no to know feature of the panel data setting. So, that is why we are not bothering about this particular aspect. So, now, we are discussing all this three aspect then accordingly, we will represent various steps and structure this panel data modeling. So, that means, if you have a model and if you are integrating cross sectional and time series so, that means, all together we have dependent variable, independent variable and intercept and error term.

So, that means, some time we have a basically the target is to know what is the intercept, that is the parameter intercept and the parameter variables, that is the slope computation. We like to know how is the impact on slope and how is the impact on intercept?

So, obviously, you is ultimately is a variables. So, ultimately U Y X are all together variables. So, we are targeting mostly of the parameters said that is intercept and slope. So, accordingly the model can be design in the means in the case of panel data setting. So, in the first case, slope will be constant. So, as usual slope will be constant. Then a intercept will vary our the individuals.

So, this is first case. Second case, slope is constant, slope and intercept varies our individuals and times intercept will vary our individuals and time. So, this is second

aspects of panel data settings. So, that means, in this first case, slope is constant, intercept will vary with individuals only. So, that is i subscript i only.

So, here is slope constant and intercept will be vary over i and t together. So, this is another setup. Then slope and intercept will vary our individuals. So, now, here we have discuss slope is constant in this particular. In these two cases, slope is constant. One case intercept will vary with individuals and intercept vary with the individual n times. In the third case, slope is constant and intercept will vary our slope and slope intercept there is no slope constant. So, here slope and intercept will be vary our individuals. In the fourth case, all will be very simultaneously. That means, is slope and intercept will very our individuals and time.

So, this is the fourth case of panel data setting. Now, what we will conclude here is that means, when there is a panel data setting, which is means which is advance, then the cross pure processional modelling you know you time series modelling. So, in the panel data setting, again we have to divided into two groups. In one group, slope is remain constant then we allow intercept vary with respect individuals. In another case slope will vary with respect to individuals and times periods.

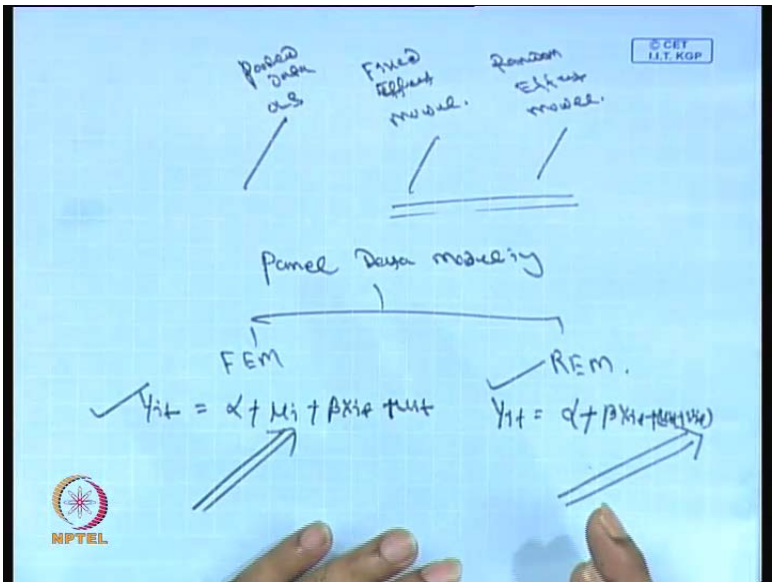
In both the cases, intercept will remain constraint, the slope is constant, one case intercept will be very with individuals. In another case, intercept vary with individuals and times. In the third case and fourth case, both intercept and slope will vary. One case it will be vary with individual, in another case it will vary with both the individuals and times. So, these are the four different setup of panel data modeling. So, now, will come to what is the exact you know estimation process of this panel data modelling. So, there may be different setup.

So, panel data modelling like you know something about cross sectional, you know something about time series, you know something about pool data and you know something. All this case were intercept, slope is constant. Intercept will vary with the individuals and times again slope and intercept both will vary with the individuals keeping time remain constant and both will vary with respect to individual and times.

These are the various structures all together in the panel data setting, but we specifically highlight two aspects of panel data modeling. That is the impact of fixed effect model and random effect models. And that we will just means in the same times will just make

a difference with a respect to pool data setting. So, all together we will discuss three different setups.

(Refer Slide Time: 48:36)



So, pool data model then panel data model that too, fixed effect model then panel data model that random effect model. So, all together we will discuss pool data OLS then and you know fixed effect model and third is a random effect model.

So, now we see here is before we going to discuss this three separately, I like to make some clarification with respect to fixed effect model and random effect models. All though the estimation process on or calculation process are completely different from pool data model to a fixed effect model and panel data model. So, in the mean times I am not discussing the details about the calculation procedure of these three models.

So, will discuss this means estimation process in the next class. So, in the mean time I will just clarify few things here with respect to fixed effect model and random effect model because these two are the core model in the panel data settings. So, generally when will be go when will be make any analysis on panel data modelling. So, random effect either random effect fixed effect or both have to means had to go together. So, it is must when will the moment mention panel the analysis first on panel data modeling? And, obviously, they must be some kind of fixed effect model, random effect model or both models. So, pool data will be come betweens.

So, it is very it is mandatory that we have to know what is the exact structure of panel data? Means if fixed effect model and random effect models? So, now I will make a differential. What is the exact difference between fixed effect model and random effect model? So, let us see here is so, basically panel data modelling in it is original form panel data modelling is looking into two aspects; this is called as a fixed effect models and random effect models.

So, let me first highlight what is the modelling difference between these two. So, the modelling difference between two is a like this. So, here is Y i t equal to alpha plus mu i plus beta X i plus u i t. So, this is how u i t. So, this is u i t. So, this is fixed effect models then random effect model will be so, y i t equal to alpha plus mu i. So, this is you know mu i means it is again variation on intercepts, alpha is a again supporting factor to that alright. So, then and this is alpha then beta X i t plus u i t plus u i t plus b i.

So, now, so, this is fixed effect model and this is random effect models alright. So, now, here in the case of fixed effect models, so, the impact will go directly to intercept and in the random effect model the impart will directly go to the in error term. So, that is the most technical difference between the fixed effect model and random effect model. So, in one case, the individual variation or times series variation is will go to intercept and if it is separate, then you will put mu y a or you can say any other component say it R t. So, this is how or you can say gamma t.

So, this is how one case it will study the individual variation. Another case it will study the a time series variation, but involve the cases it will target the intercept only, but when will be target the variation on intercept, then, obviously the some kind of impart will go to the error component.

So, the moment it will go to the error component, and then we will study the error observation also. So, in the moment will a study the observation, and then in that case, we will assume that the intercept will remain constant. Then what is the impact on error variance when we are making individual variations and times series variation? So, when will target individual variation and times series variation then, obviously, it will go to error terms. So, that means, error term basically so, it as three different segments. So, one segment is the individual segment then times series segment and both this segment.

So, these are the three different classification of this a means error terms s error impact. So obviously, so technically, there is a difference between fixed effect model and random effect model. Some times there is a confusion, because we have not discuss anything about the estimation process which will discuss in the next class. So, in the mean times, after you know a knowing the technical procedures of this two models you get know the exact answer, but always there is question. So, what for type is between the two which particular model is more feasible and which one should be use?

So, it depends upon. There are many aspects. We have to answer these questions. First thing when will go for you know fixed effect models then, obviously, the parameters are too many, but where will go for random means because it is intercept or changing and, obviously, time is changing so, obviously, parameters very high. So, when will go to random effect model then the parameters are very less.

So, as a result, if that is the case then, obviously, fixed to go to the fixed effect model means a go to fixed effect model may be better chase or random effect model may be better choose depending upon exactly depending upon the size of sample. So, means exclusively, it will depend upon the individual observation and its time series observation. So, with the help of individual observation and times series observation then we can get two means. We can come to a point that means, we can decide that which particular model is more effective than others means. The detail will be discussed in the next class. So, with this we can conclude the session. Thank you very much. Have a nice day.