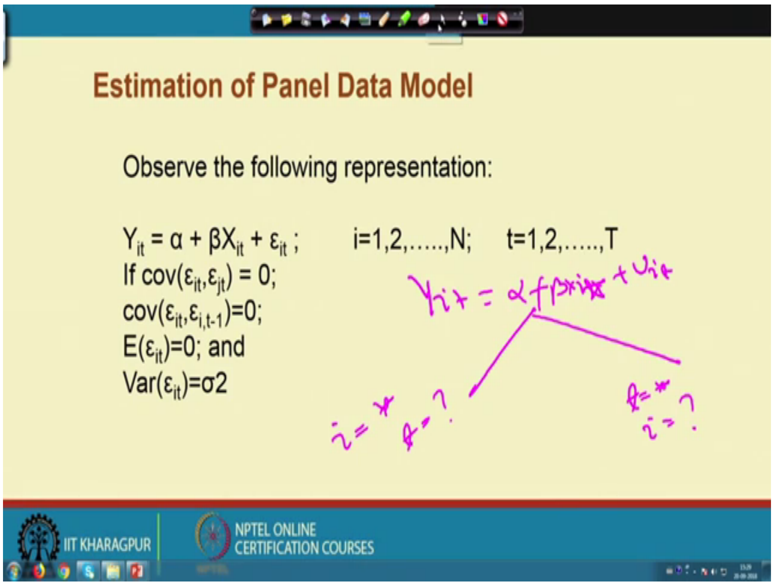
Engineering Econometrics
Prof. Rudra P. Pradhan
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

Lecture – 55
Panel Data Modelling (Contd.)

Hello, everybody and this is Rudra Pradhan here. Welcome to Engineering Econometrics. Today, we will continue with Panel Data Modelling and that too the coverage on fixed effect model random effect model and the method of GMM. Impact in the last two lectures we have highlighted this particular you know concept we have justified the importance of panel data modelling and here we have to start with the various models that is attached with panel data in that to the fixed effect model random effect model and the generalized method of moments. So, we like to check how is this kind of you know concept.

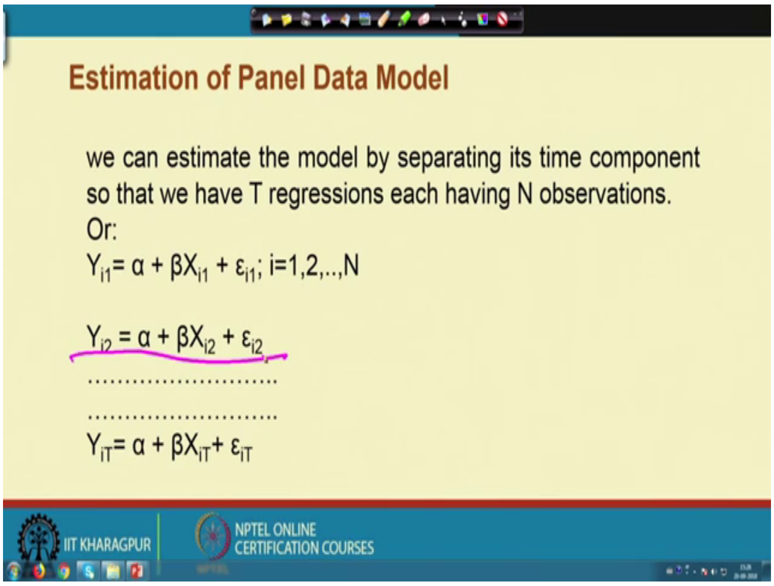
(Refer Slide Time: 01:43)



So, let us see here the usual format of panel data model is like this we have dependent variable Y and then we have independent variable X and the simple mathematical structure is Y i t equal to alpha plus beta X i t and i t by default represents the pulling up data by integrating both cross sectional data and the time series data.

And, there is a requirement of some of the basics like you know the covariance of these two error term should be equal to 0. And covariance of error terms with respect to two different time periods must be equal to 0 and again some of the error terms that is with respect to both cross sectional and time will be equal to 0 and then finally, variance of the error terms should be equal to 0 which we have already highlighted in the last lecture.

(Refer Slide Time: 02:33)



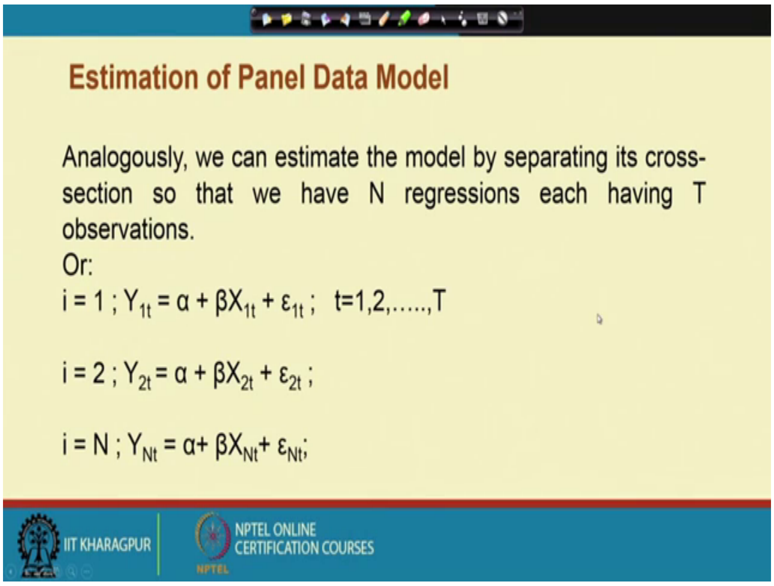
And, in this lecture just we extend this kind of you know structure and then we slightly connect to fixed effect model then will move to random effect model and the kind of you know GMM.

So, corresponding to this you know model Y i t equal to alpha plus beta X i t. So, if you know the structure is like that. Here we can estimate the models by separating its time component, so that we have T regression and that to each having an observations for instance we can start with like this the extension Y i 1 equal to alpha plus beta X i 1. So, that means, here keeping time constant T equal to 1, if you allow i to vary then the first model will be like this and again when we fix equal to 2 then the model will be the model will be Y i 2 equal to alpha plus beta X i 2. Similarly, you continue then Y i 3 equal to alpha plus beta X i 3.

So, that means, every times we are you know estimating the particular model with to with different time you know you know time period. So, that means, every years we have a estimated model where the sample observation will be cross sectional type. In the you know compare to this you know cross sectional structure so, we can have also time series structure where keeping i equal to you know 1 or i equal to 2 you know keeping cross sectional components constant T can vary.

So, that means, we may have a structure where for every cross sectional unit we may have you know estimated models where you know the sample will vary with respect to different points of you know time. So, now, what is happening here?

(Refer Slide Time: 04:47)



So, if you go ahead with this kind of you know structure then will you find the general the general structure is like this. So, here i equal to 1, Y 1 t equal to alpha plus beta X 1 t a similarly Y 2 t equal to alpha plus beta X 2 t, so; that means, here i remain constant then t will be vary.

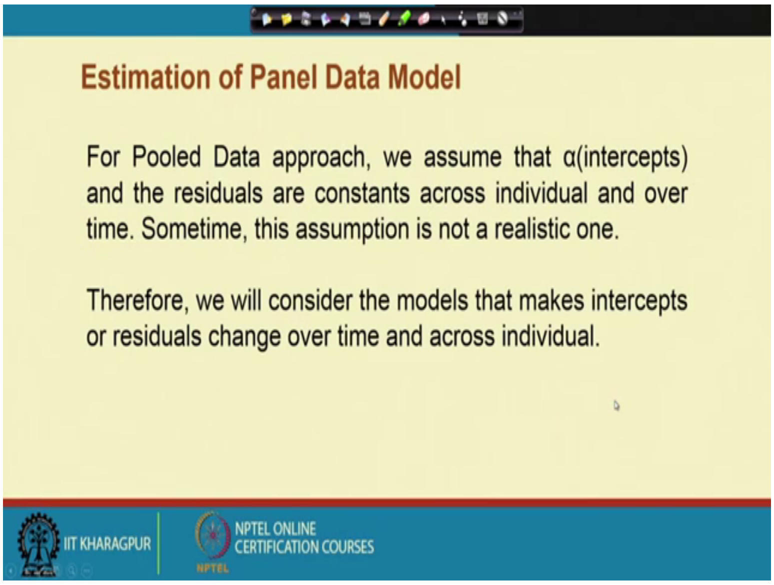
So, that is technically you know corresponding to original model Y you know this once Y i t equal to alpha plus beta X i t. So, we have a two different flow, for instance like this you know we start with like this you know Y i t equal to alpha plus beta X i t. So, then we have of course, there will be u i t or error term. So, this you have to extension. So, where i will actually a keep constant and t will be vary then you know they here t remain constant then, i we like to vary, ok.

So, this is how the different structures. So, corresponding to this ones so, we if you if you go ahead then you will find so, this is the case where i you know vary t constant. So, that means, this part. So, if I call a t here part v is a this is what actually we call which we.

So, this is what we call actually a this is a the case where i fixed, t will be vary i fix t vary means this site and then the other one here t fixed, so that case then I will be vary.

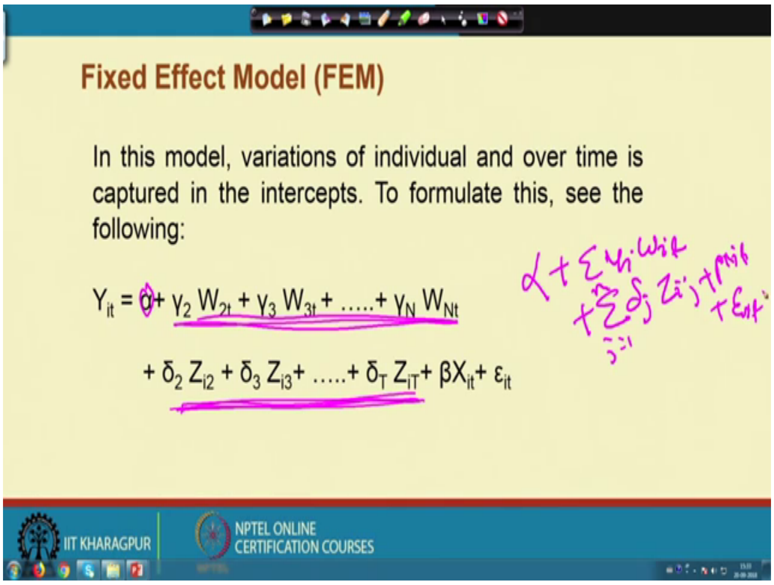
So, this is how the you know two different structure and a accordingly we can have different levels of you know output so, that means, having with pull of that term with respect to cross sectional units in time series units. So, keeping i remain constant t can vary then we have a various sets of you know econometrics output that is with respect to we know every cross sectional unit and again for every time a means time periods we have estimated model where are i will actually vary. So, that is how you know we have lots of you know flexibility with respect to panel data modelling.

(Refer Slide Time: 07:26)



So, now, taking this clue we will proceed for the panel data modelling. So, before we start the panel data modelling that to with respect to fixed effect models random effect models and the GMM so, we let us you know go through same kind of you know pooled data structure. In pooled data structure we assume that alpha intercept and the residuals are constant across individual and overtime. Sometimes this assumption is not the realistic one therefore, will consider the models that makes intercept or residuals change over time and across individual.

(Refer Slide Time: 08:09)



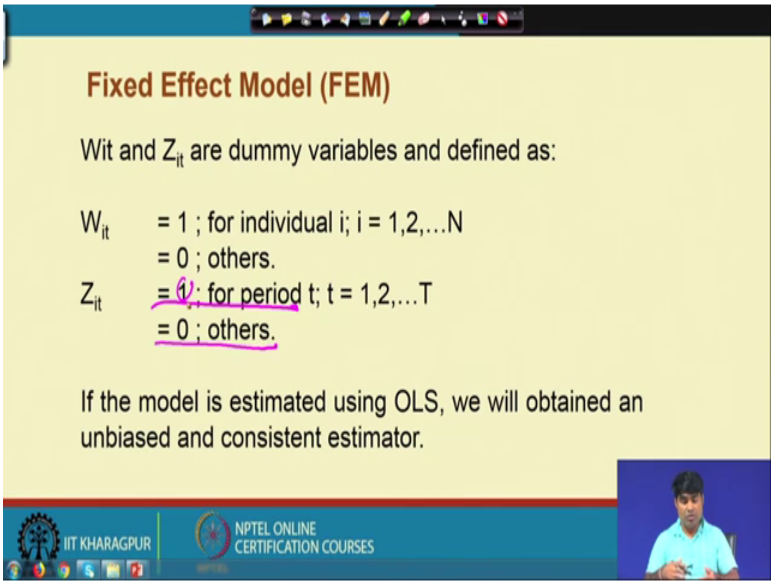
So, what is happening here we have the concept called as you know fixed effect model and then we have actually a random effect models and we have GMM. In the case of you know fixed effect model the variations of individual and over time is captured in the intercept that to you know the entered impact to the go to intercept where the error component will be remain constant. So, in that case we have the structure like this, for instance so, this is intercept and this is one set of you know variation that to with respect to you know different you know time and then this is another structure with respect to process in a you know types.

So, this is again with respect to dummy application and this is with respect to again dummy application and that to dummy with cross sectional unit and dummy with time series in it is. So, that means, technically if I like to generalize then the model can be rewritten like this alpha plus summation let us say Y i W i t. So, that is that is one kind of you know case or then another case is a something like you know W you know summation delta j then a delta j then Z i j. So, j equal to 1 2 you know ns plus beta X i t and then you know error term. So, this is this is how we can actually we can start this particular you know concept.

So, technically we are actually a bringing the a you know time component and the cross sectional component into the systems and then we estimate the a model. So, that means, we like to check the between dependent variable and independent variables how the cross sectional unit can you know bring the kind of you know impact again how the time series you know component can bring the impact. So, if that is the case then every times dummy can be used for cross sectional unit and then time series unit. So, a that too know whether the cross sectional unit has a impact over any time series unit has an impact.

So, every times the variation would be with respect to 0 1. So, when the times particular time period will be consider then the wait will be 1 and if not then the by default it will be 0 for others. Similarly, for the cross sectional unit, then by default the entire import to the go to the intercept that is what the concept of you know that the concept called as you know fixed effect model.

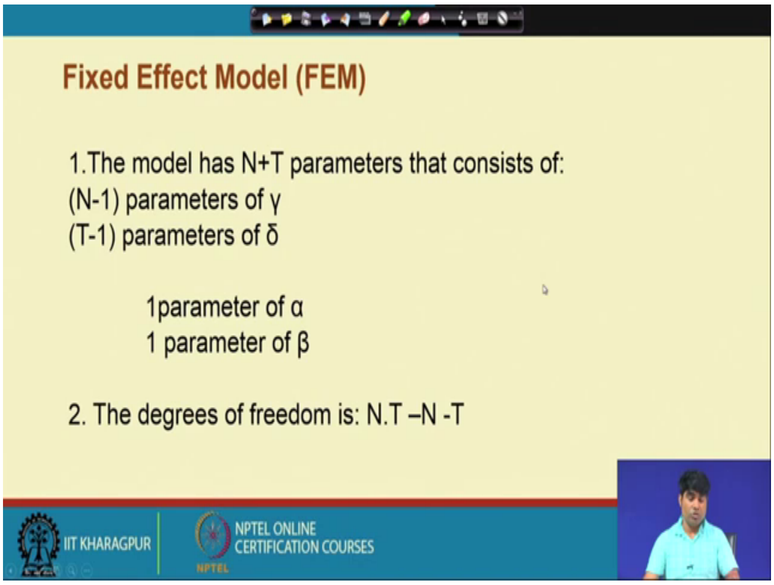
(Refer Slide Time: 11:08)



So, now, now taking this clue if you go ahead then, the entire structure can be written like this W i t and Z i t at the dummy variables and defined as you know W i t equal to 1, for individual i and i will vary from 1 to N and for then equal to 0 for others. So, that means, when you consider a particular sectional unit that time the wit will be 1 and for other cross sectional unit by default it will be 0. Similarly, Z i t is representative for the time series component. So, when we consider a particular time period that time that time period will be use as a proxy one and at the sentence the other time periods will be treated as you know 0 impact. So, that is why so, the structure can be can be editing actually Z equal to 1 and Z equal to 0.

So, for instance if we use consider t equal to 1 so, for the time period you know 1, so, it will be 1, then the other time periods the effect will be 0. So, like this we have actually a you know extensions; that means, the simple model can be extended with you know lots of dummy variables by the help of you know cross sectional impact and the time series impact. And, if the model is estimated using OLS, then we will obtain an unbiased and a consistent estimator and, that is that is not the big deal. So, obviously, we like to check how is this particular you know structure.

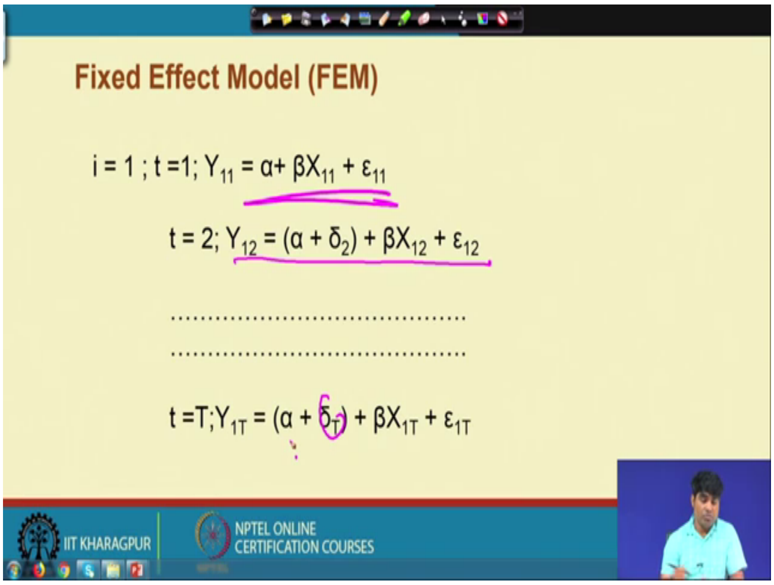
(Refer Slide Time: 12:41)



In the case of you know fixed effect model a model has actually N into T parameter; that means, N represents the number of cross sectional units and T represents the number of you know time elements. So, obviously, we have N minus 1 parameters of Y and T minus T minus 1 parameters of you know delta. So, one parameter of you know alpha and one parameter of beta that will also part of this system and by default the degree of freedom for the entire model will be N T minus N N minus T.

So, N T represents the a cross N represent the cross sectional you know items and T represents the time items. So, so, if you multiply these two and that will be the total sample of this you know panel data and since we are taking care of you know degree of freedom that with lots of you know dummy with respect to cross sectional unit and trans unit by default the entire degree of freedom will be finally, N T minus N N minus T. That is what the structure.

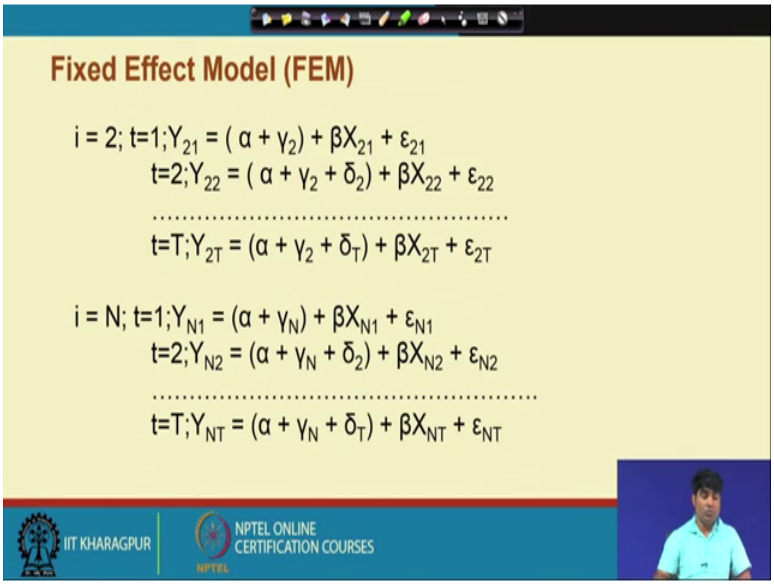
(Refer Slide Time: 13:52)



So, now, will just you know elaborate this fixed effect concept. So, let us start with i equal to 1 and t equal to 1 and where you know you know i and t a has a some kind of you know connection. So, that means, when i equal to 1, t equal to 1 then the model will be simply represented or like this. So, this is what Y 1 1 equal to alpha plus beta X 1 1 and then error term. So, when i equal to 1 and t equal to 2 then this model can be started like this. So, in that case delta 2 will be you know you know new impact that will come into the intercept.

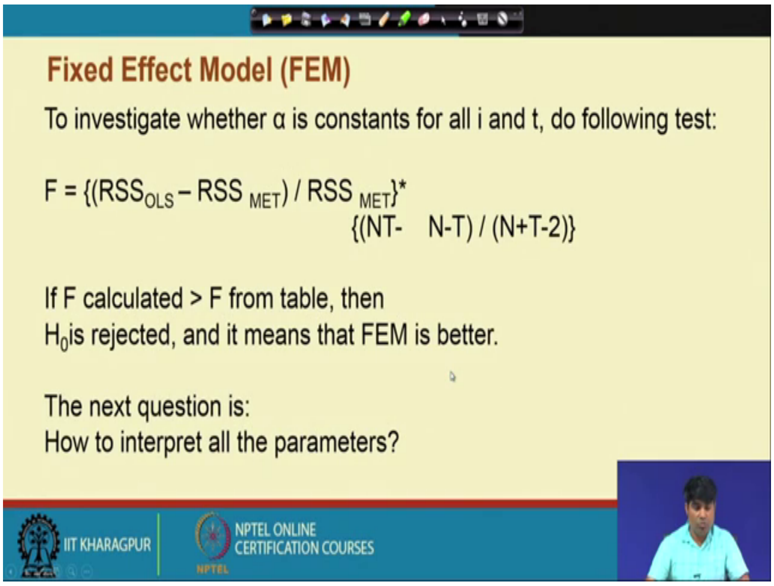
So, likewise if you extend t equal to T then the delta T component will be coming into the pictures. So, that means, for the first time period so, alpha will be the kind of you know proxy for the second time period. So, the impart will be alpha plus delta 2, third time periods then it will be alpha plus delta 3 and so on. So, ultimately the impact will be different with respect to different time periods. So, there is a high chance in a t may be t equal to 1 the intercept impact will be high, where t equal 2 the intercept impart may be low or may be high, it exclusively depends upon the sign of this particular you know delta 2 and the a alpha. So, ultimately so, this is how the kind of extension.

(Refer Slide Time: 15:31)



And then we can move into the a further extension, where i equal to 2 and t equal to 1, then the model extension would be like this and Y equal to alpha plus these are the intercept; that means, actually we are moving to second cross sectional unit where you know t will be again vary. So, that means, so, alpha then the first you know intercept term would be a added then a you know t will be vary again. So, then that that times you know the entire you know impact I can come to the intercept. So, likewise you can you know extend this models so; that means, there is a huge pool of you know extensions.

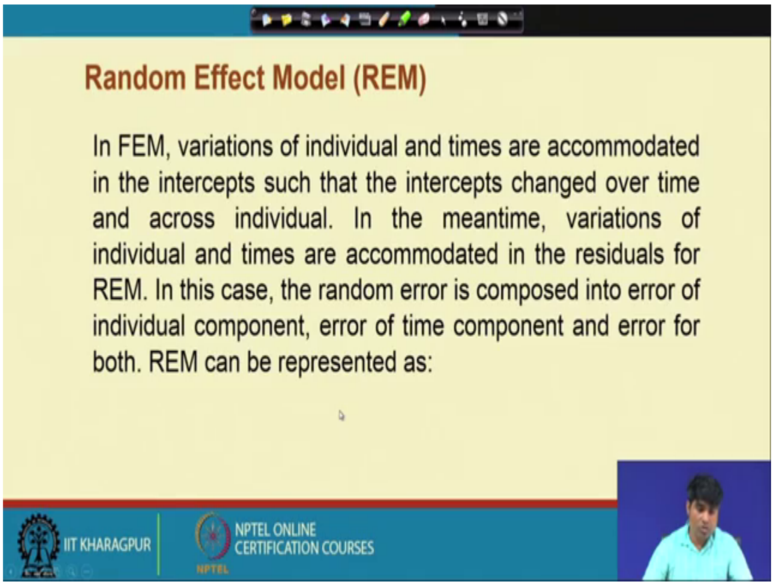
(Refer Slide Time: 16:08)



So, now, to summarise the fixed effect model where alpha is the constant for all i and t. So, we need to actually test the following so, that is actually of tenth way statistics the a statistic actually a you know derive through the a residual sum of the squares that to which you know OLS mechanisms and then the kind of you know restricted mechanism. So, then you know you will find over there F is becoming statistically significant or not. If F is becoming statistically significant then the default the particular model will be very effective for the particular you know for the particular requirement.

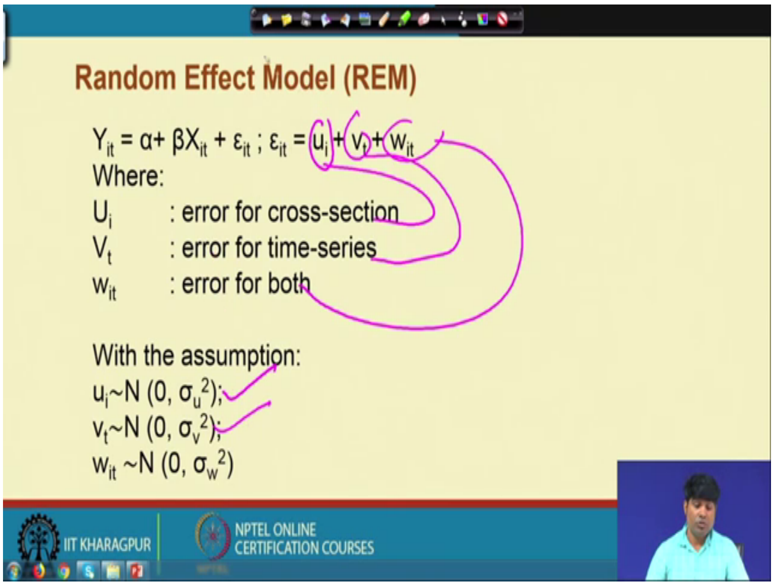
So, now the next question is how to interpret all these you know parameters? Obviously, the interpretation is the very simple because you know ultimately the impact will go to the intercept. So, we can check with that you know it is a high weight or you know in a low weight and then the coefficient of you know a X i will be as usual like you know ordinary regression you know milling. So, we can actually highlight the details by taking a particular you know examples.

(Refer Slide Time: 17:28)



So, let us first you know you know moving to random effect model compare to the fixed effect model in the random effect model variations of you know individual and times are accommodated in the intercept such that the intercepts changed over time and across individual. In the meantime variations of you know individual and times are accommodated in the residuals for random effect model, in this case the random error is composed into error of individual component and error of time component and error for you know both.

(Refer Slide Time: 18:13)



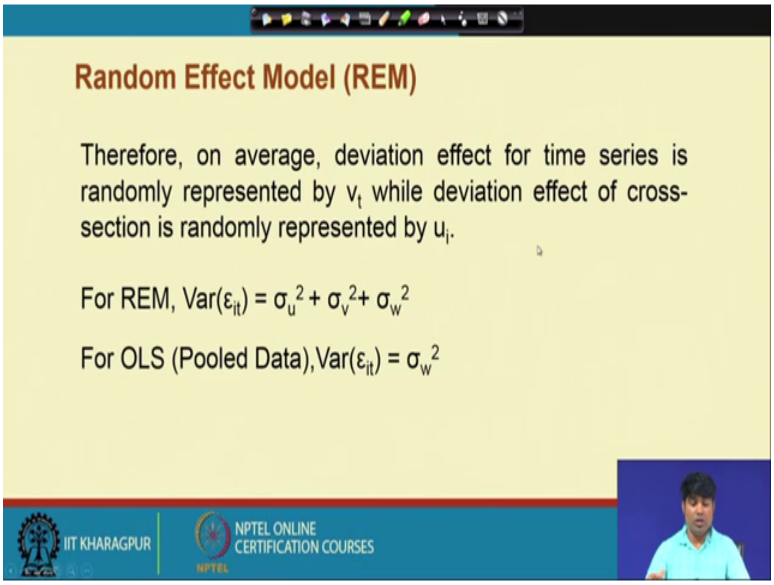
And a random effect model can be represented as you know where you know Y i t equal to alpha plus beta X i t and the error terms will be final actually move into three different forms. One particular form is the cross sectional type, another form is the time series type and then the pool type. For instance we have here u i that will be the cross sectional error and v t that is actually time series error and w i t is actually by pulling back cross sectional and time series.

So, that means, in the case of you know fixed effect model the time impart and the cross sectional impact will be move into the intercept that is a taken care of through alpha and then plus minus depending upon the a you know negative and positive impact of you know time element and cross sectional elements. And, then finally, in the error component particularly for the random effect model same way it will move into in into the final error terms that is with respect to the cross sectional error then time series error and the pool error.

So, now like as usual you know regression modelling since there are you know three error terms so, you expect the error term you should be actually mean you know normally distributed with a mean 0 and unit variance and again error term v t which is actually a you know due to time component here and it will be normally distributed with the 0 mean and unit variance and then w i t which is actually having in N 0 with you know unit variance.

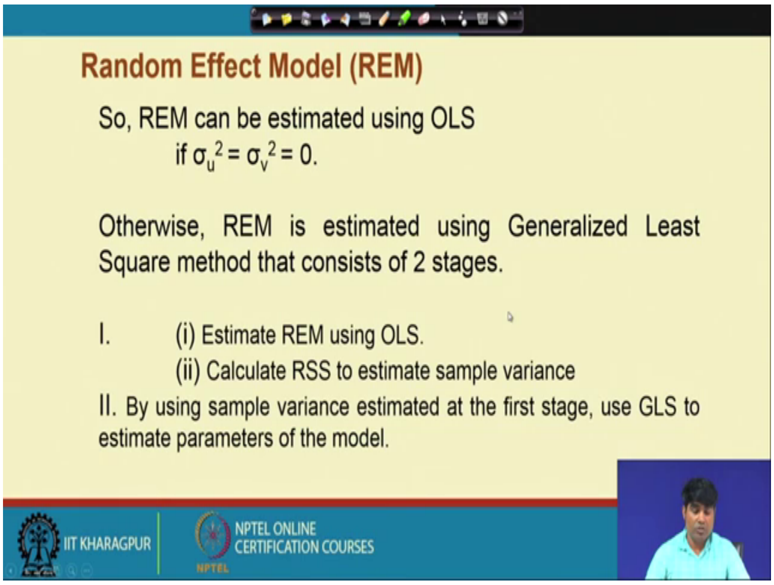
So, now, with this you know structure we like to check a how is the kind of you know deviations.

(Refer Slide Time: 20:11)



So, moving into the kind of you know estimation process what we can have you know in the deviation effect for time series is randomly presented by v t while deviation effect of cross sectional randomly represented by a u i. So, ultimately for a random effect models the ultimate variations of the error terms will be sigma square u that is actually through cross sectional you know impact and sigma square v i t is through actually a time series impact and then sigma square w that is with respect to actually pool data. So, as a result so, the error a total error will be decomposed into cross sectional error time error and the pool data error. So, that is how the partition.

(Refer Slide Time: 20:59)

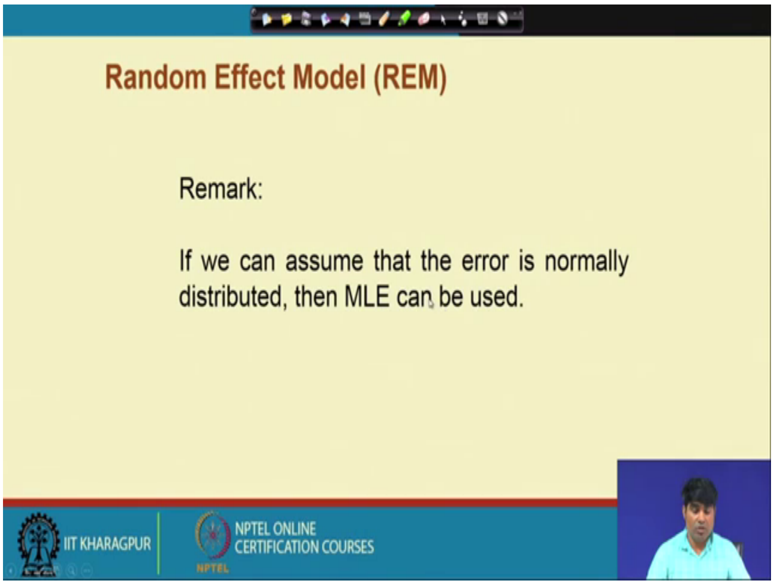


And, in the random effect models you know you know the it means it can be estimated this the particle model can be estimated using OLS provided if sigma square u equal to sigma square v and that to equal to 0. So, that means, both you know error variance for time and for cross sectional should be equal to 0, otherwise the random effect model is estimated using you know generalized you know that means, GMM mechanism instead of you know OLS we have to use you know GLS you know like least square methods. So, that is actually with that is with respect to two stage process.

The first one is the estimate the random effect model using simple OLS technique and then calculate RSS to estimate sample variance. And, then by using a sample variance estimate at the first stage use GLS to estimate parameters of the models, that is what the a structure of the random effect model where the error variance of the cross sectional units and time units are not same and they are not equal to you know equal to 0.

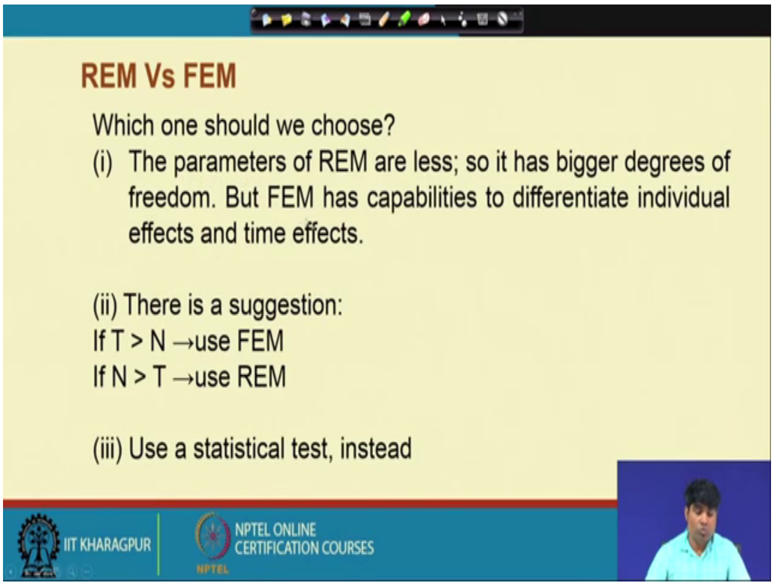
So, in that case we can use GLS and by default GLS follows the two step process by first step is the application of OLS s and second stage by the application of you know GLS.

(Refer Slide Time: 22:30)



Then, what is happening here in the random effect models the requirement is that we like to assume that the error is normally distributed and then next we will like estimated can be used to estimate these parameters.

(Refer Slide Time: 22:53)



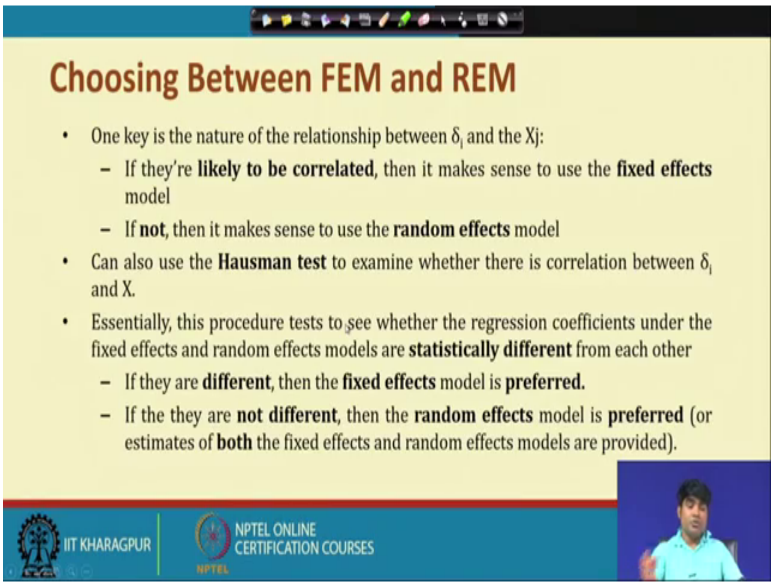
And, there the parameters of random effect models are less. So, it has a big degrees of freedom. But, fixed effect model has capability to differentiate individual effect and time effect.

So, compared to actually random effect model and fixed effect model. So, if the objective is to check actually individual time series impact and the individual cross sectional impact then my suggestion is actually the fixed effect model is the better choice. If that is not the case then the best choice is the random effect model where actually the kind of intercept or the individual time impact and cross sectional impact will not be actually coming into the picture. The entire you know weight will be a taking care of you know that to through actually the error terms.

Now, the issues actually how to make a kind of choice, in what situation you can use random effect model and what situation you can a use actually fixed effect model of course, theoretically I have already highlighted if you have no objective to study the individual impact of a cross sectional unit and the time series unit, then you can go ahead with random effect models if the objective is to know which cross sectional unit is here in high impart and which time series you know it is having high impart or low impart and then in that case fixed effect is the right choice. But, technically a if T greater than N; that means, the time elements greater than to cross sectional elements that to with respect to size then use fixed effect model.

And, if N greater than to T, so, that means, the cross sectional units are you know high than the time units in that time the choice will be a random effect model. So, use a statistical test instead of choosing between fixed effect model and random effect model.

(Refer Slide Time: 24:57)



So, one vital is the nature of the relationship between delta i and the X j; that means, the kind of you know cross sectional units, then the time series units and then the kind of you know independent variables impact if they are likely to be correlated then it makes sense to use the fixed effect model if not then it makes sense to use the random effect models.

Can also use the Hausman test to examine whether there is correlation between delta i and X j essentially the idea is the means the procedure a procedure test to see whether the regression coefficient under the fixed effect and random effect models are statistically different from each other. And, means technically if they are different then the fixed effect models is actually first and choice. If they are not actually different then a random effect model is the next preference.

So, so obviously, so, we have lots of you know clues which can actually justify whether through you know take care the fixed effect models and the random effect model to address the particular you know engineering problem and the that to as per the a requirement of kind of you know situation. So, ultimately we have lots of you know options whether to you know use random effect models or fixed effect models or sometimes you can actually refer both the results, there is no harm at all. Ultimately, if you connect with you know randomic panel data models then some times the intercept by default will be the move in the process.

So, so, ultimately it depends what is your actually the kind of you know objectives and how is the kind of you know problem structure. So, looking actually all dimensions you know then you have to fix what kind of you know model panel data model you can pick up to solve this particular you know engineering problem and as per the kind of you know this is making requirements.

So, ultimately so, we can get to know what is exactly the fixed effect model and what is the random effect models and how they differentiate each other and what situation you can you use fixed effect model what situation you can use random effect model. Sometimes there is no harm to refer both the models output to address the particular you know engineering problem. So, what will you do? So, we will take a live problems and then you like to check how is the kind of you know you know estimation and how the estimated output can be a used by you know by addressing some of the you know a technical issues and the kind of you know requirement.

So, taking the clue of you know estimated outputs we can get to know the particular difference between fixed effect model and the random effect model and we can get to know the kind of you know beauty of the panel data model to you know highlight some of the engineering problems. With this will stop here and will continue in the next lecture.

Thank you very much. Have a nice day.