Exploring Survey Data on Health Care
Prof. Pratap C. Mohanty
Department of Humanities and social Sciences
Indian Institute of Technology, Roorkee

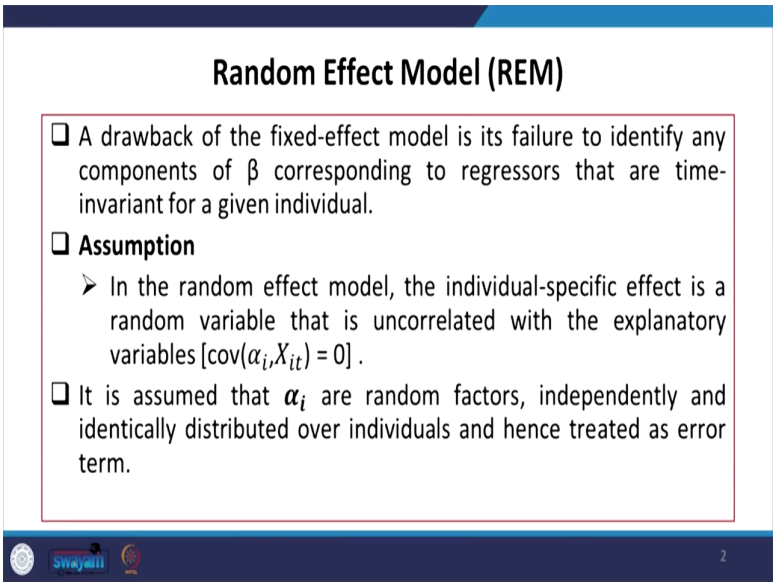
Lecture - 34
Random Effect Model in Healthcare

Welcome friends once again to my NPTEL Mooc module on Exploring Health Care Survey Data. We are in the seventh week of explaining panel survey data. This is the lecture meant for the random effect model.

In the previous lectures, we explain very clearly understanding the fixed-effect model in health care. Without explaining further details in the previous lectures, I think it is time for the random effect model it is based on the distribution of the panel data and usually, there are different forms of data in cross-sections we have different forms you need to taste whether it falls under fixed effect model or under random effect and accordingly we take the appropriate decision.

The random effect model is important because of its drawbacks to the fixed effect model. One of the drawbacks of the fixed effect model is its value to identify any components of beta corresponding to the regresses that are time-invariant for a given individual,

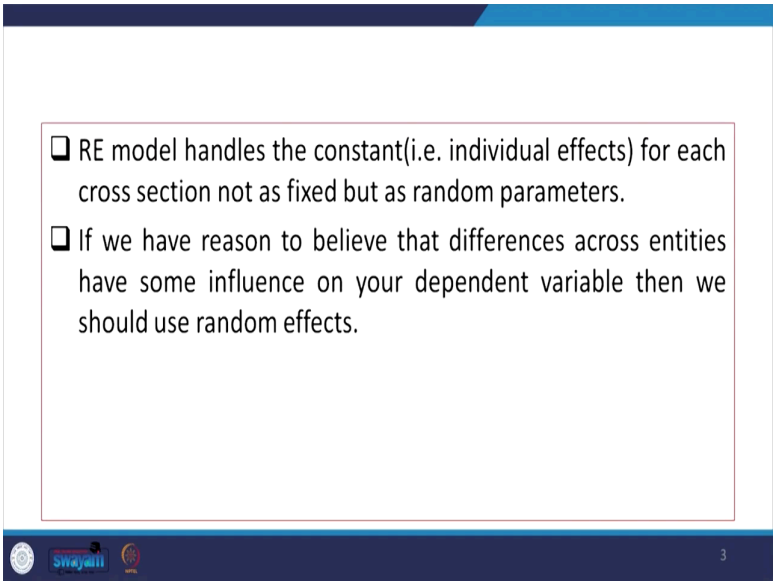
(Refer Slide Time: 01:36)



so, the beta which is actually also, having issues of time invariance are actually not discussed correctly, in the earlier model. so, accordingly, we take off the estimation of beta based on the time component. So, there are assumptions of the random effect model the assumption here is that the individual-specific effect is a random variable that is of course, uncorrelated with the explanatory variables.

Hence the covariance of the covariance between alpha I and the X it the explanatory variables are actually equal to the covariance is equal to 0. so, it is assumed that alpha i are random factors independently and identically distributed over individuals and hence treated as the error term of the distribution.

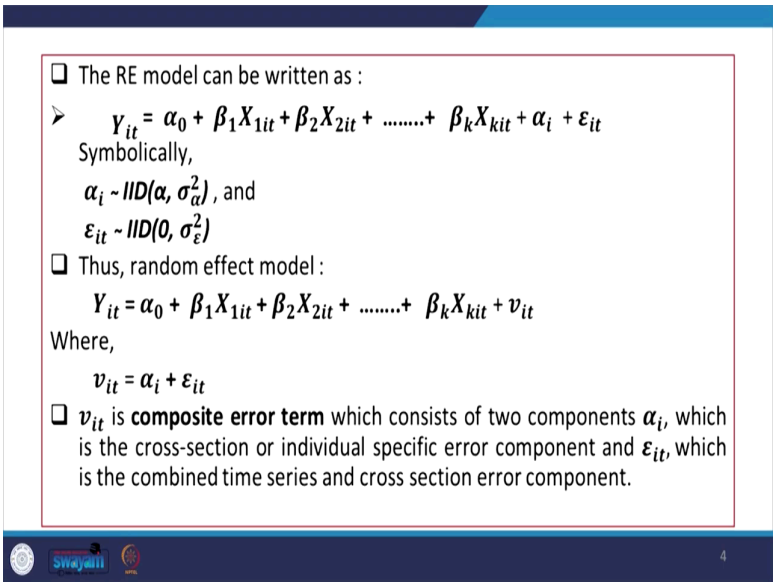
(Refer Slide Time: 02:42)



so, we will also, combine those alpha components in the error term while estimating the random effect model. That is going to be shown in our slide.

The random effect model handles the constants that are the individual effects for each cross-section not as the fixed component but rather as a random parameter. If we have reason to believe the differences across entities have an influence on your dependent variable, then we should use random effects. Instead of a fixed effect.

(Refer Slide Time: 03:14)

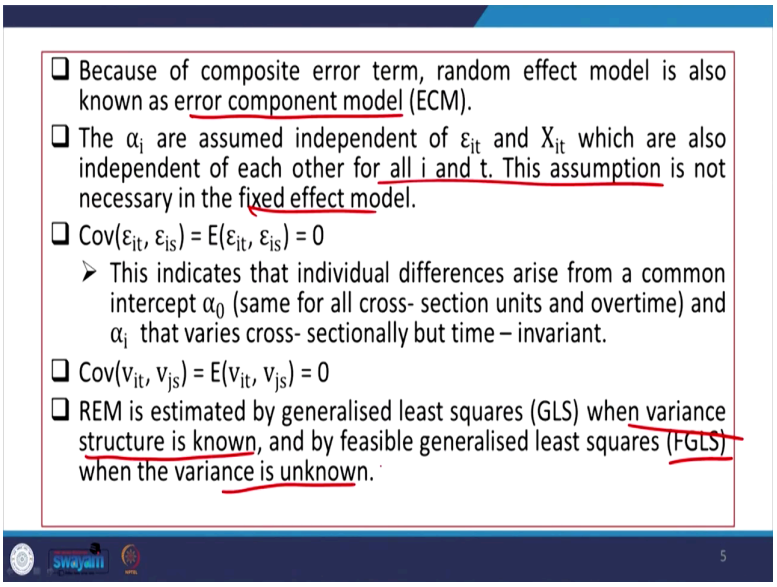


The standard model of random effect is presented here that is the dependent variable is with i Y it is equal to the constant term, with the beta coefficients and its explanatory variables.

And at the end we have two components, that is the alpha i component and the error component. so, both are we have already said that they are actually independently and identically distributed. so, alpha is distributed with its mean alpha and standard deviation, sigma square. And the error term is actually distributed within its mean 0 and standard variance.

Thus, the random effect model can be actually composed of this error term. Since they are independently and identically distributed. so, the alpha i content and the error term both the components are actually capsuled with vit, and the rest we are going to estimate. so, vit is in fact the error term that is presented here.

(Refer Slide Time: 04:31)



And that consists of two components as alpha i and epsilon it, which is the combined time series and cross-section error component.

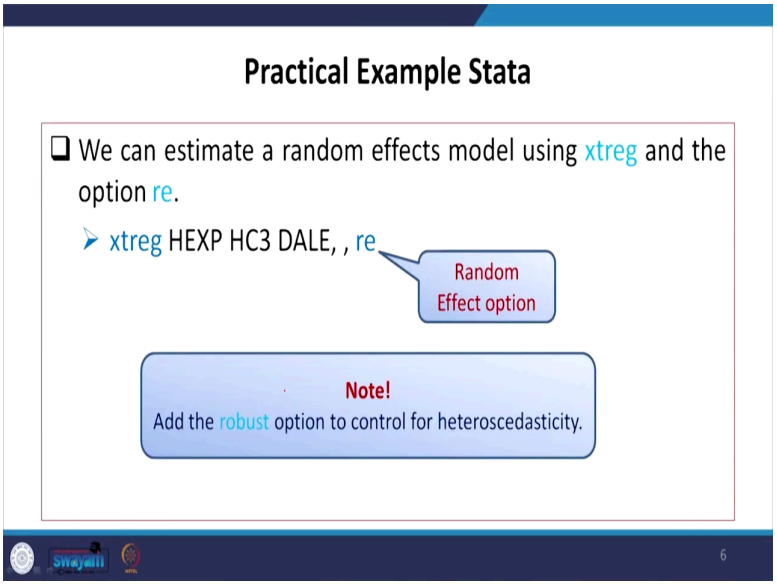
Because of the composite error term, the random effect model is also, known as the error component model. This is also, known as the error component model. In short, it is called ECM, alright. And the alpha i are assumed as independent of their term and the X it which are also, independent of each other. For all i as well as all t component this assumption is not necessary for the fixed-effect model. so, this is how it is different as compared to the fixed-effect model.

so, the covariance of alpha i t and alpha i is equal to alpha i_t, and the covariance of epsilon i t and the epsilon s the is nothing but equal to 0 and we indicate that the individual difference arises from a common intercept that is alpha naught is same for all cross-sections, cross-section units and also, over time. And the alpha varies cross-sectionally, but not by time or time-invariant.

so, the covariance after composing both this term error with its time component error with its cross-section component. so, the covariance of vit and v j s is equal to v i t and v j s should equal to the expected value. Basically, when we find out the covariance, we take the expected value of these two are actually equal to 0 or is equal to 0. so, the random effect model is estimated by a generalized least square technique in sort called GLS.

When variance structure is actually known that is one of the important aspects. When the distribution is known, then we explain it through the GLS model and by feasible generalized least square technique as well inside that is called an FGLS when the variance is known to not know. so, there are two approaches one is when variance is known and when variance is not known and accordingly GLS or FGLS models are actually used. We are also, going to show you practical explanations using the given dataset.

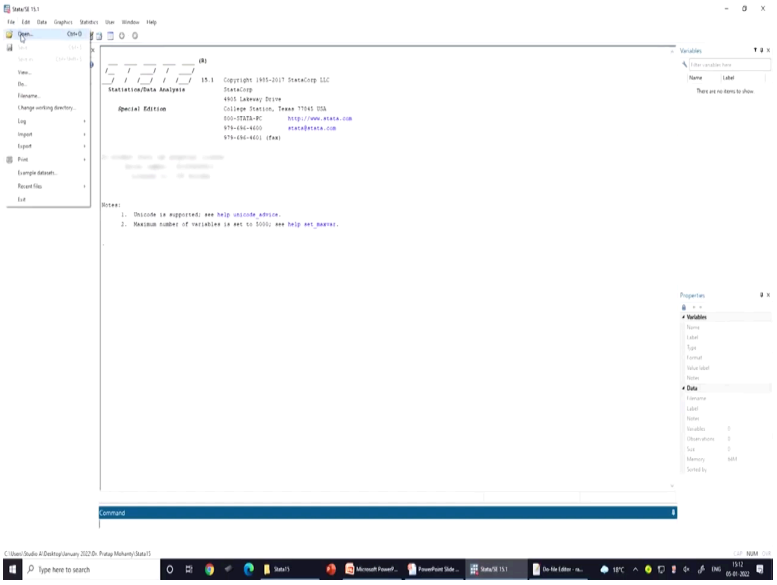
(Refer Slide Time: 07:38)



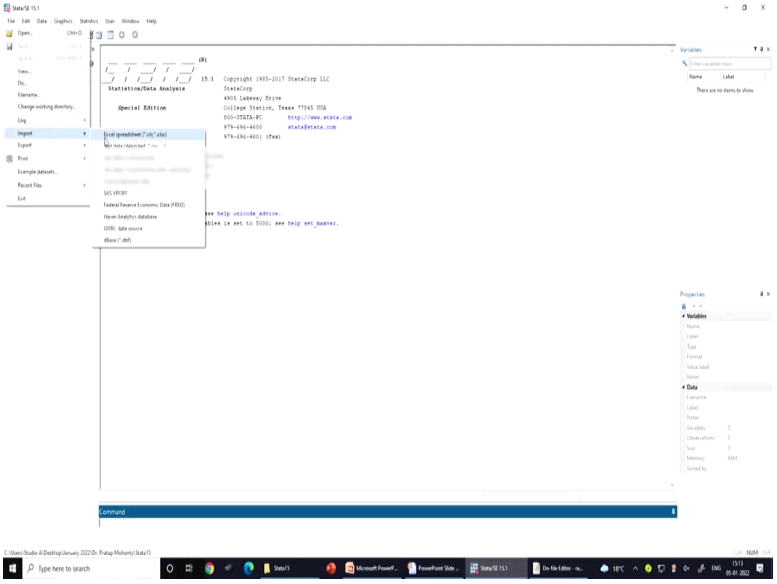
We will first estimate the random effect model using the command xtreg command xtreg with the options we will give it as re random effect and then we will find out whether that actually fits or not,

so, we will go to the practical session, alright. Now here is our Stata.

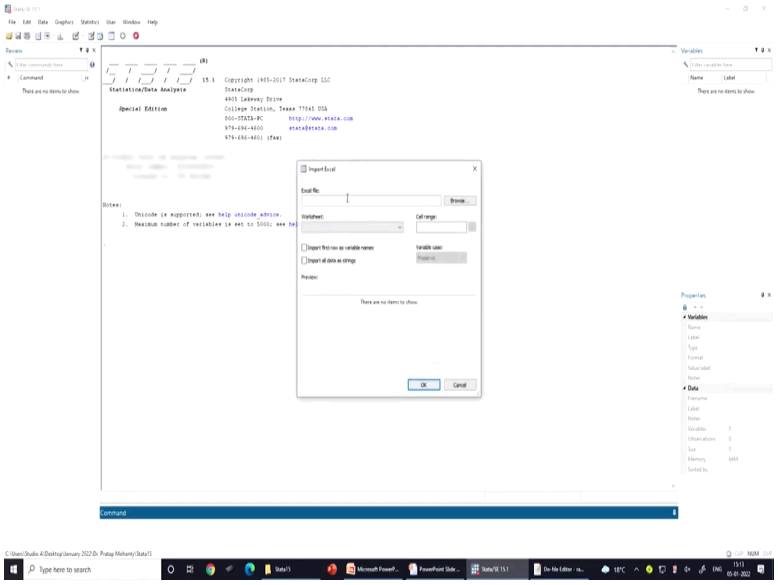
(Refer Slide Time: 08:00)



(Refer Slide Time: 08:08)

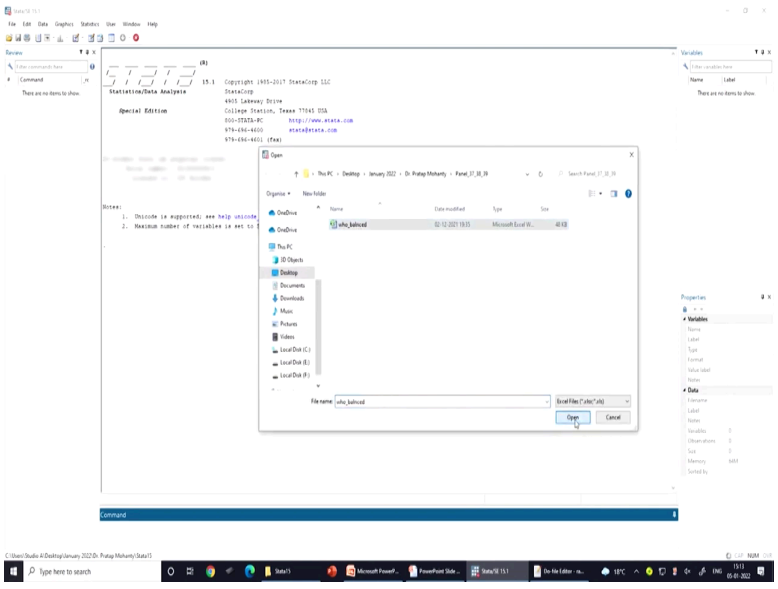


(Refer Slide Time: 08:09)

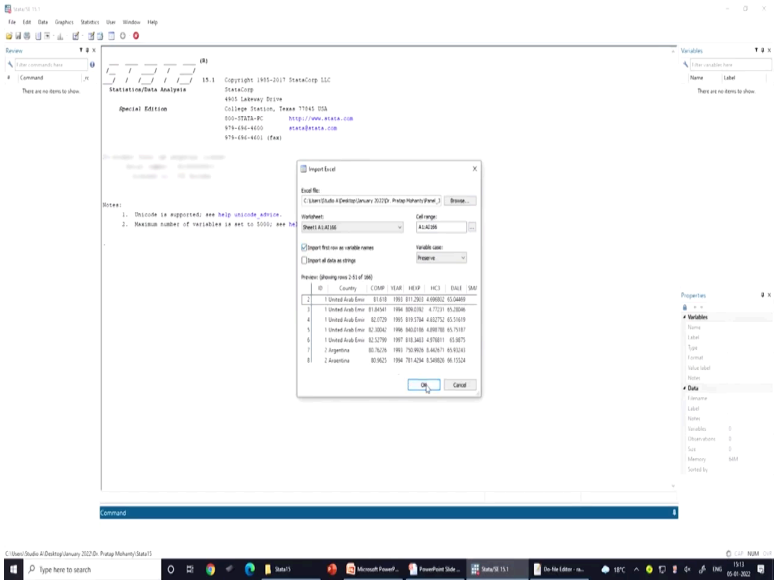


And we will all so, use the same w h o data, which we already experimented in the previous lecture.

(Refer Slide Time: 08:12)

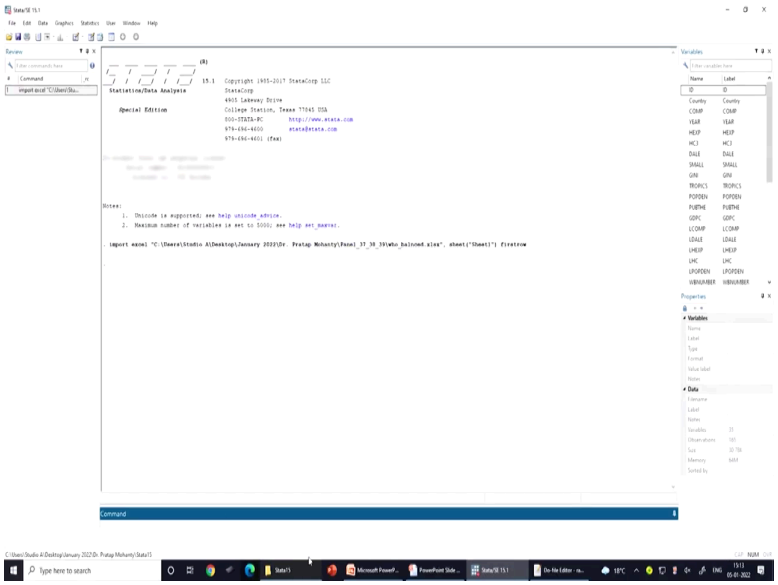


(Refer Slide Time: 08:14)



And it is in our screen and we are taking a WHO balance data set as given.

(Refer Slide Time: 08:18)

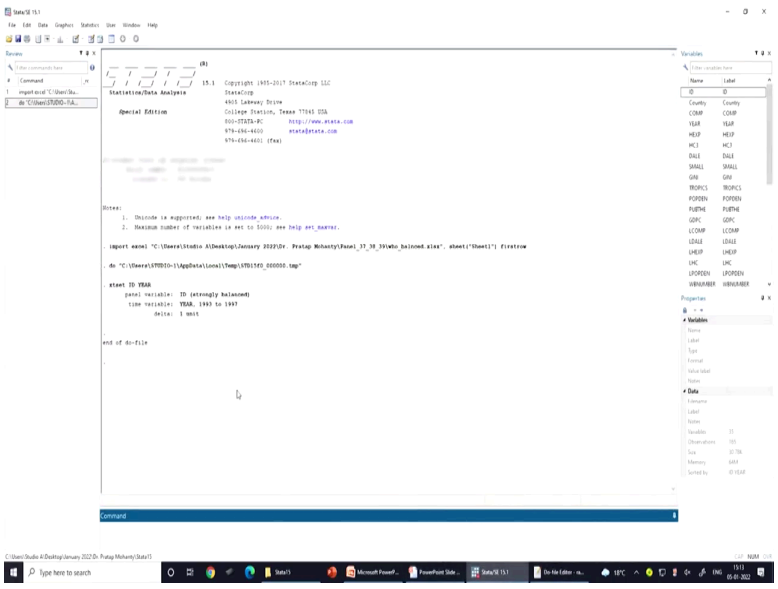


And accordingly, we specified the way we define it.

(Refer Slide Time: 08:22)

Now we are all so, giving you the details.

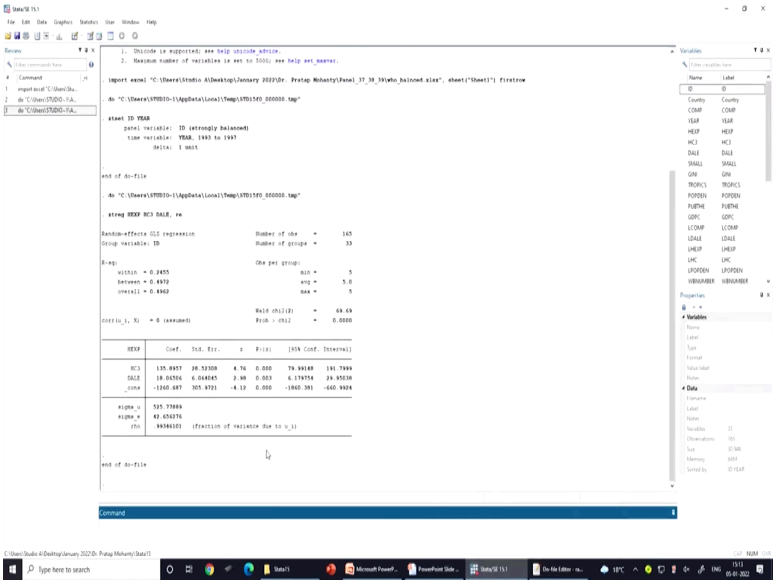
(Refer Slide Time: 08:34)



In on our screen. First, we will operate with will set the ID variable and the year variable in our on our screen and this is already set. The xtreg command has already been identified. That it is a strongly balanced data and the data is from 93 to 1997.

And now, we will go by the second command. The second command is to attach with xtreg the way we do and at the end, we will give the option is re naught fe, not fixed effect.

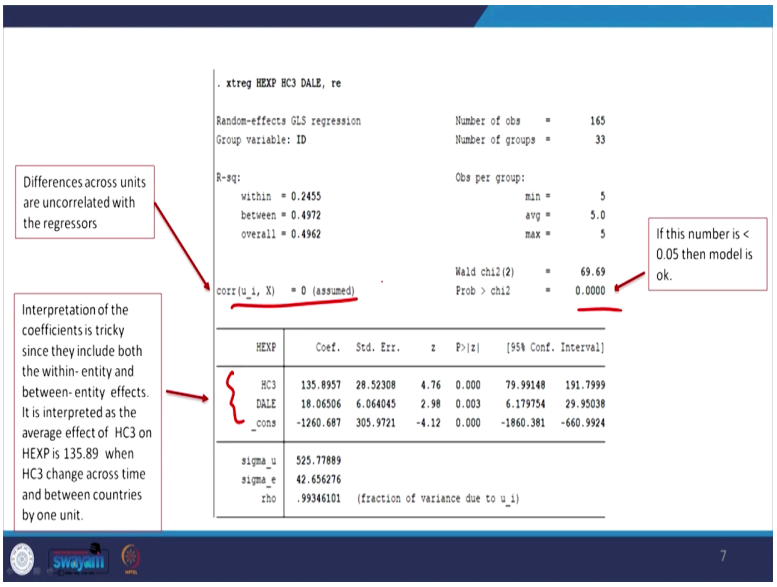
(Refer Slide Time: 08:58)



so, this is our result and now, you can see the model is actually significant and the implications impact of each of the components is defined and its level of significance are, also, defined you can easily read between the line. And these are positively linked to the dependent variable, so, and I am not explaining much I have already explained earlier in our previous lecture.

Now, another aspect is that you can also, add the robust option to control for the heteroscedasticity. If there is any, technique wanted to go for to check this heteroscedasticity better to run with the robust option.

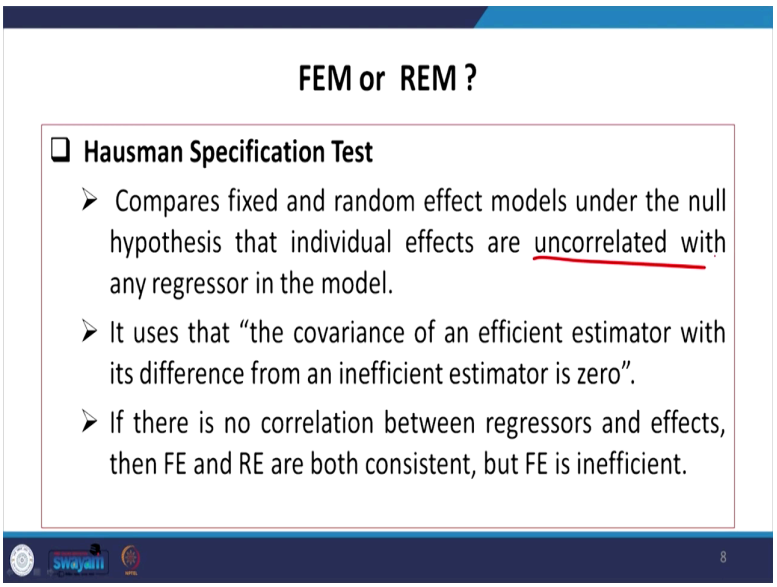
(Refer Slide Time: 09:53)



This is the one we have already derived on your screen and I have already explained this and similarly all those variables, and interpretations also, we have attached here for your reference; similar approaches we did it in our previous lecture.

Another aspect is that the difference across units is actually uncorrelated with the regress so, rs. That is also, important. so, the this is mentioned here on the screen, you can see that this is assumed to be 0 as per the random effect model,

(Refer Slide Time: 10:31)

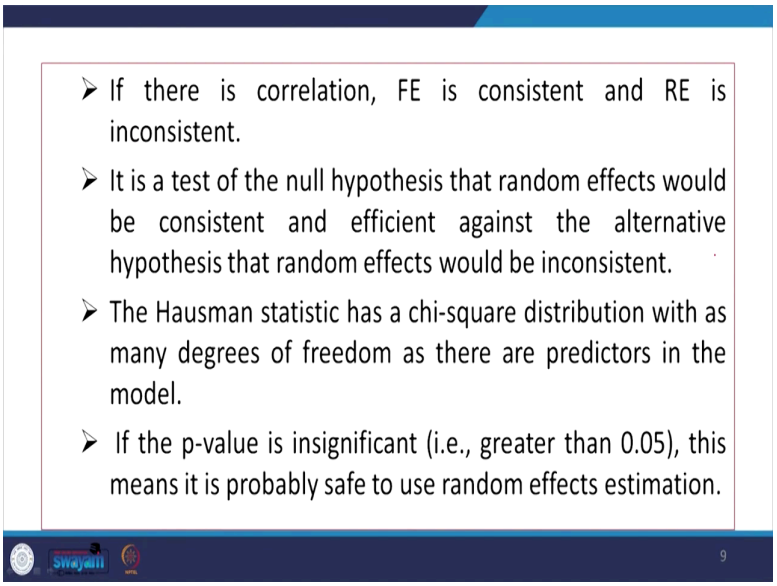


so, the rest interpretations are perfectly fine and you can easily do it. Now we are going to clarify whether our data is going to be interpreted with a fixed-effect model or a random effect model, or we cannot just take randomly with any of the models.

so, we need to specify with a test called Hausman; Hausman Specification Test. This compares fix effect and random effect models under the null hypothesis that individual effects are actually uncorrelated with any regression in the model, so, that is it it's very clearly spelled and we already said, they know individual effects are uncorrelated with the regress so, r. It uses that the covariance of an efficient estimator with its difference from an inefficient estimator should be 0.

If there is no correlation between regress so, rs and effects then fixed effect and random effect are both consistent, but the fixed effect is in fact inefficient.

(Refer Slide Time: 11:37)

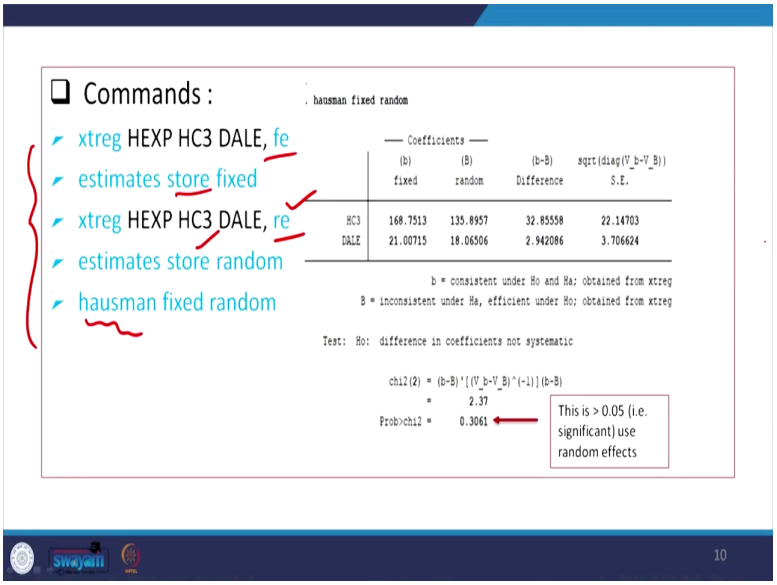


If there is a correlation, then FE is consistent. Therefore, RE is not suggested to be applied. It is a test of the null hypothesis the random effects would be consistent and efficient against the alternative hypothesis, that random effects would be inconsistent the Hausman Test has a specific chi-square distribution with as many degrees of freedom, and as there are predictors in the model.

If the p-value is insignificant; that means, the assumption is not rejected. If insignificant; means, if it is greater than 0.5 as per the standard practice of 0.05 this means it is probably

safe to use the random-effects model. so, since the assumption is that coefficients are not systematically distributed differences in coefficients are not systematic.

(Refer Slide Time: 12:37)

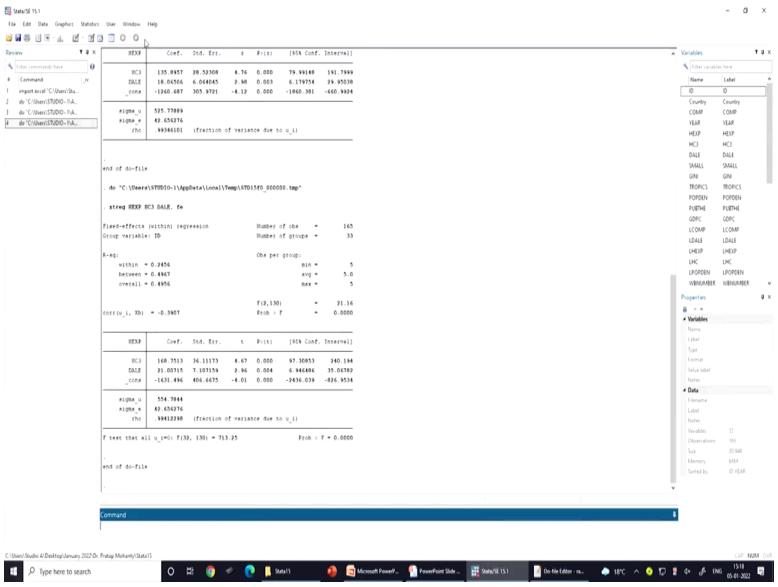


so, now it is not rejected. if it is since the assumption is that it is not systematic; that means, it may follow randomness in the relationship with correlation covariance is going to be 0.

Now, this is not rejecting; that means, it is it in this case in our result also, we can check these are the command you can easily see I will also, operate with it, we will first go by the xtreg with fixed effect, then we will estimate and store that fixed effect result. We will store it then we will go for the random effect model, then also, we will store then we will check Hausman Test.

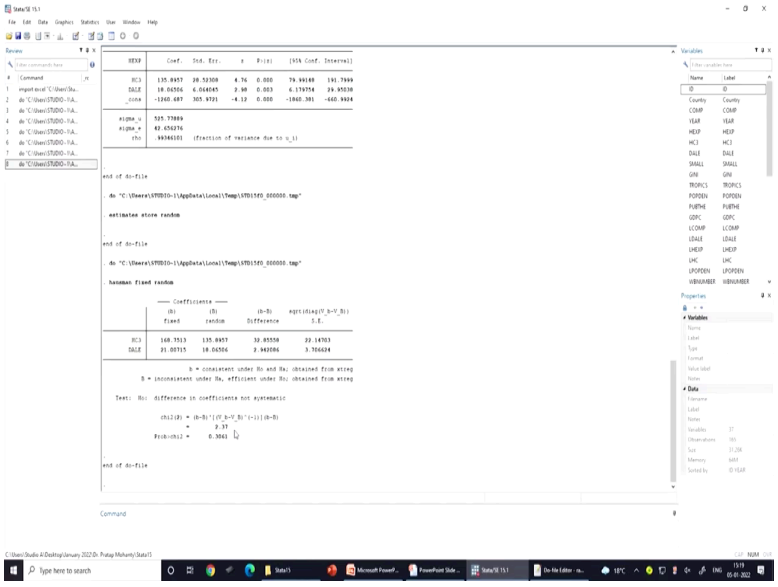
this is how it is followed you can have a check it is here. Now we have this first we will run with once again with a fixed effect with the same data set.

(Refer Slide Time: 13:35)



Then we will we can this is the fixed effect result and we have already interpreted this, and earlier and you can check you can store this fixed effect result estimate store fixed with the name fix we have given. Then we will run the random effect model and we also, store it store with the name random and we have stored it then we have run the Hausman Test.

(Refer Slide Time: 14:07)



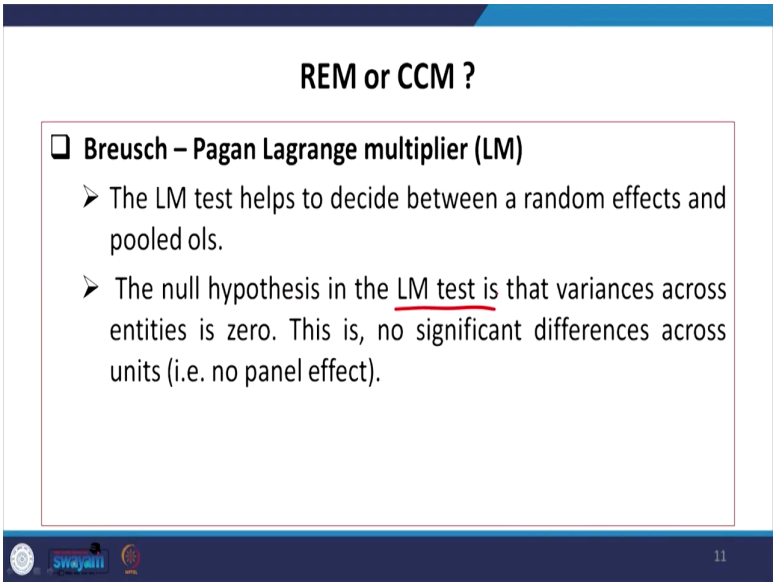
Hausman Test is the most suggested instrument to check the difference,

Now, you can see that your hypothesis is that it should not be systematic, the difference in coefficients is not systematic, so, systematic means it follows a certain order, not systematic means it is having certain randomness, alright.

Now it is saying that the randomness is this not in fact violate rejected. Because of the p values, you can see the p-value is of value 0.3061; that means, it is greater than that of 0.05. it is not rejecting this model is not rejecting the null hypothesis; that means, it says that it is in fact nonsystematic.

The covariances are expected to be 0 therefore, it is suggested to go for a random effect model, instead of a fixed effect. Had it been the fact that it is this p-value is significant it is always suggested to go for your, always suggested to go for a fixed-effect model, alright. Now we can see here and these are all presented on your screen even though I have clarified using this screen,

(Refer Slide Time: 15:35)



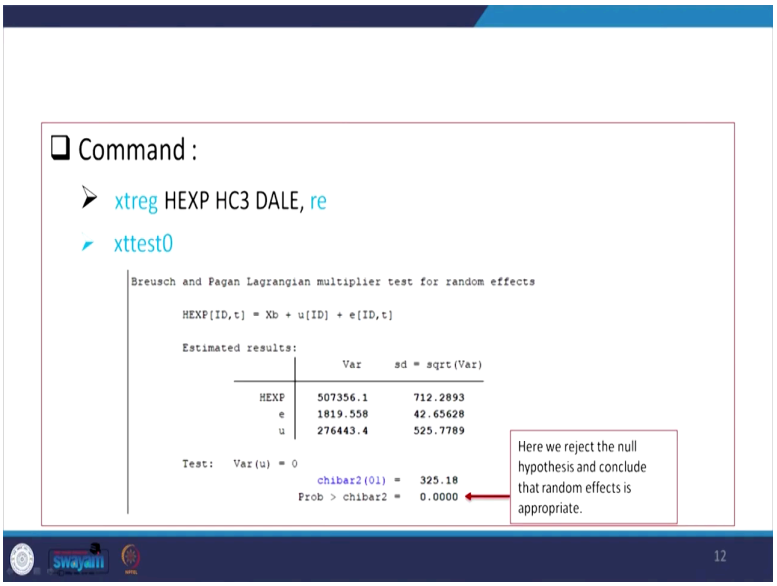
so, let us move to another comparison, the constant common model as compared to the random effect model.

one of the tests that are important in this case which is frequently used by the researcher is called the Breusch-Pagan Lagrange multiplier, in short, it is called the LM test, LM test helps to decide between a random effect and a pooled OLS. so, pooled ols which we have already said for CCM Constant Common Multiplayer or model. In that case, we have a comparison in

this particular segment that is the null hypothesis, in the LM test is that likewise, we did for the Hausman test here also, certain there are certain null hypotheses.

This suggests that variance across entities is 0. Variances across entities are 0, this is a no significant differences across units, so, across units, there are no significant differences; that means, there is no panel effect, alright. That is the assumption if it is violated then accordingly, we can take the decision.

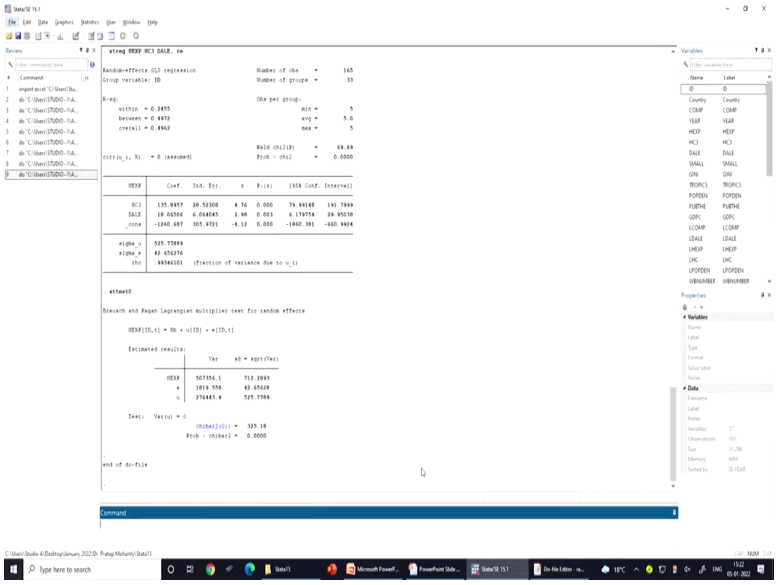
(Refer Slide Time: 16:54)



Here is the command like you will you have to go with the xtreg, then random effect then the next text is xttest0. Here this gives you the idea that here the level is significant p-value is significant, this suggests that we reject the null hypothesis and conclude that random effects are in fact appropriate.

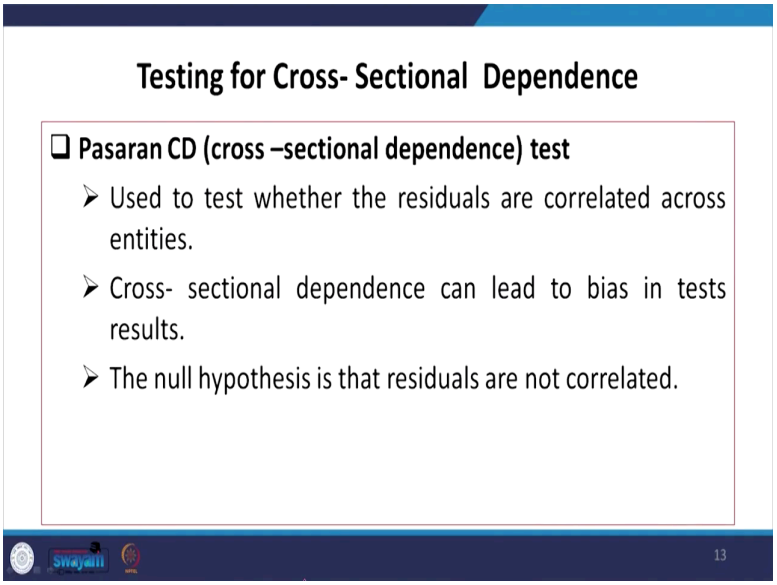
Based on a likewise assumption it suggests that the variance across entities is 0 and no significant difference across units, but here you are saying it is rejecting that one, and accordingly we take the decision. This is what you can also, check it we have kept everything over here, so, this is the one we have run it on your screen and you can just have a lot, alright and here is the level of significance and since it is significant.

(Refer Slide Time: 17:52)



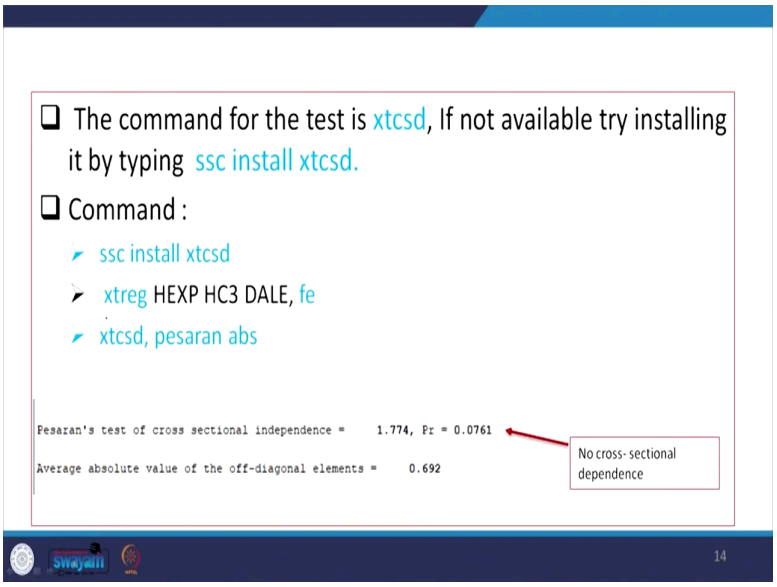
so, we suggest that you should go for a random effect model instead of C M 1, alright.

(Refer Slide Time: 18:03)



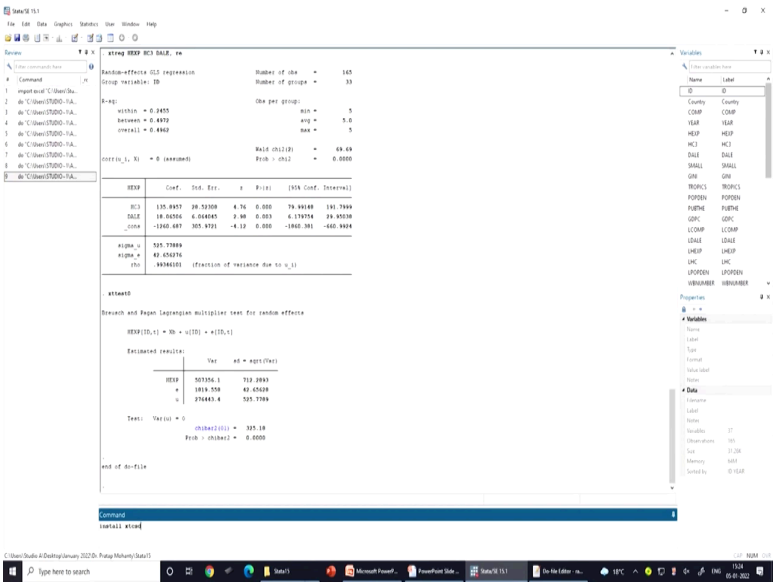
Then it comes to understanding whether they occur in a cross-sectional dependence. It might be the case that the cross-sectional units are also, dependent on each other. so, for that Pasaran CD or cross-sectional dependence test C D in short. This test is used whether the residuals are actually correlated across entities. The residuals we have are actually expected to be correlated, then cross-sectional dependence can lead to bias in test results. so, the null hypothesis suggests, that residuals are not correlated.

(Refer Slide Time: 18:44)

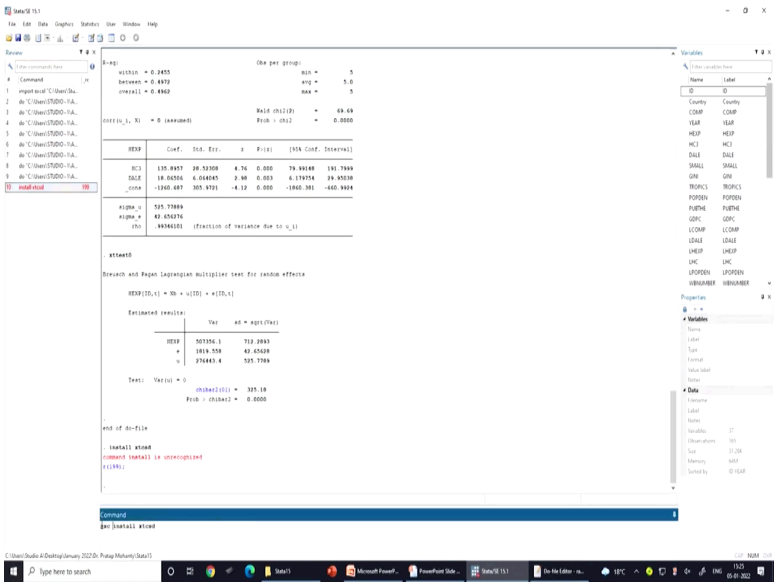


if they are correlated then we have to make the decision accordingly. Like here our, p-value the probability value for it actually, suggests that it is having a value of 0.0761. Now we can see that there is no cross-sectional dependence. Because of its level of significance and the command for this test is xtcsd, alright if not available, then we need to actually install it by typing ssc install xcsd. so, that you can do it on your own it is on the screen xtcsd.

(Refer Slide Time: 19:34)



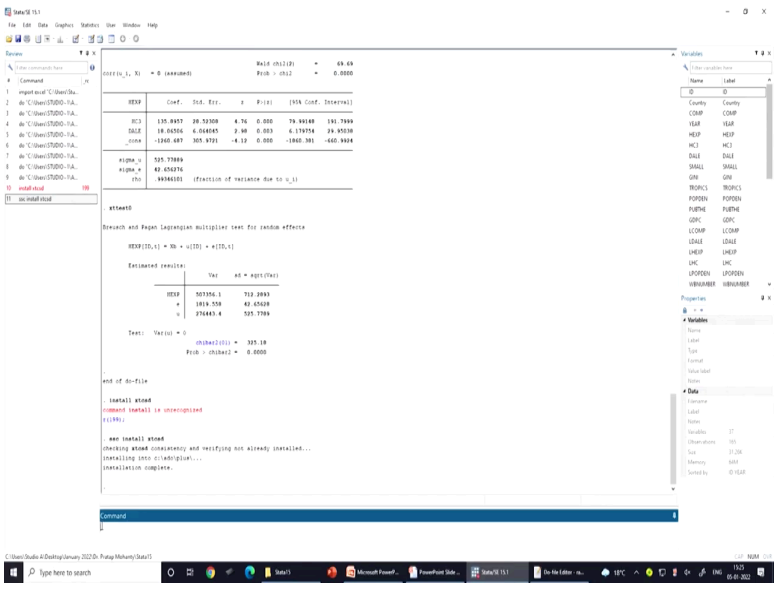
(Refer Slide Time: 19:44)



you need to install it and we can take the help of this,

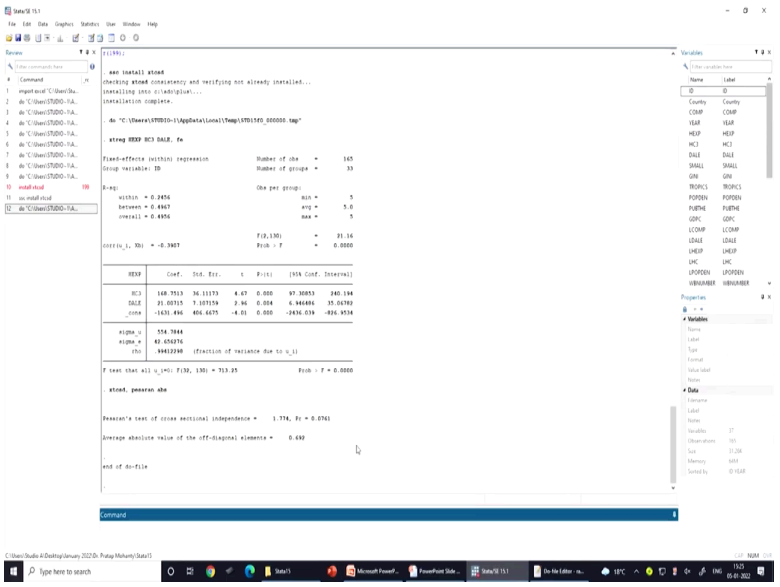
ssc install xtcsd ssc, alright.

(Refer Slide Time: 20:02)



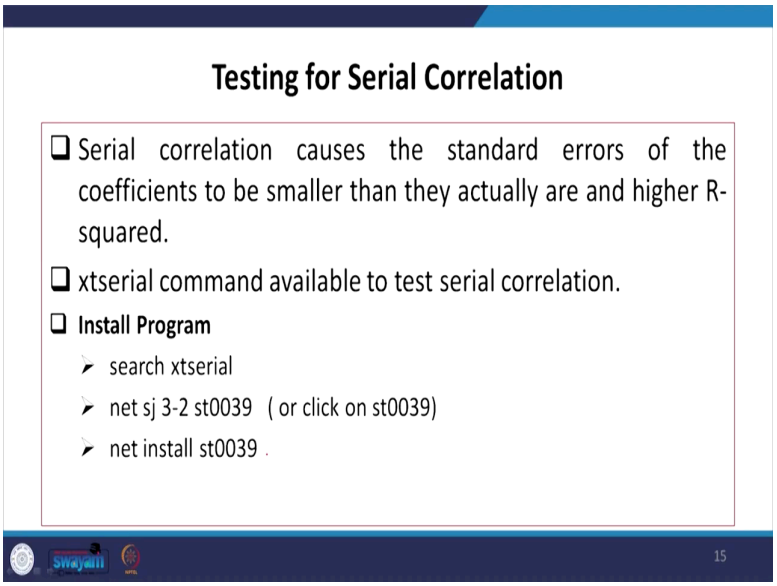
it is, verifying not already installed so, now installation is complete. Now you can take the command and accordingly we can take the decision of running it.

(Refer Slide Time: 20:28)



this is the both of these we are actually running it and we found that this is actually significant and there is no cross-sectional dependence, alright.

(Refer Slide Time: 20:41)

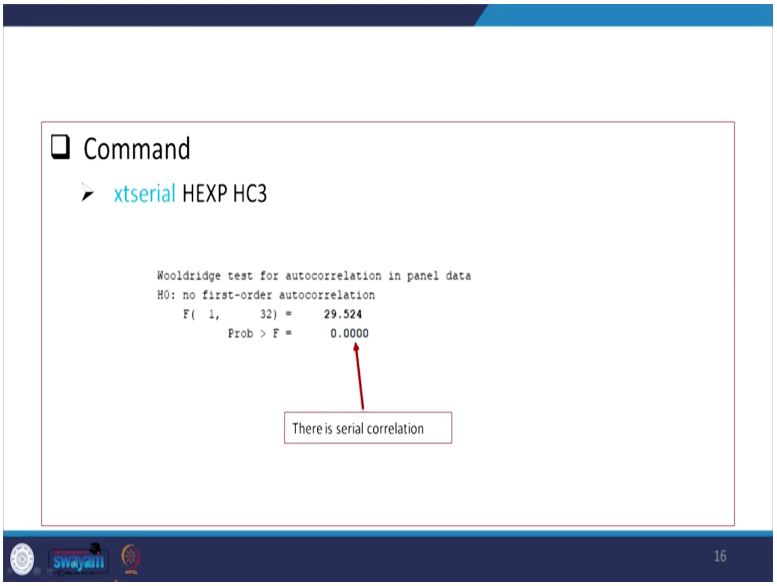


And now this is done. Then another possible aspect is to also, check the serial correlation, because of the time component.

so, serial correlation causes the standard errors of the coefficients to be smaller, than they actually are and a higher R square is possible, we need to check. so, xtserial command is

available to test serial correlation. so, xtserial we need to search this xtserial then net sj 3-2 1and its that based on this string value of it and accordingly string since it is the variable correlation string. so, accordingly, we can do it.

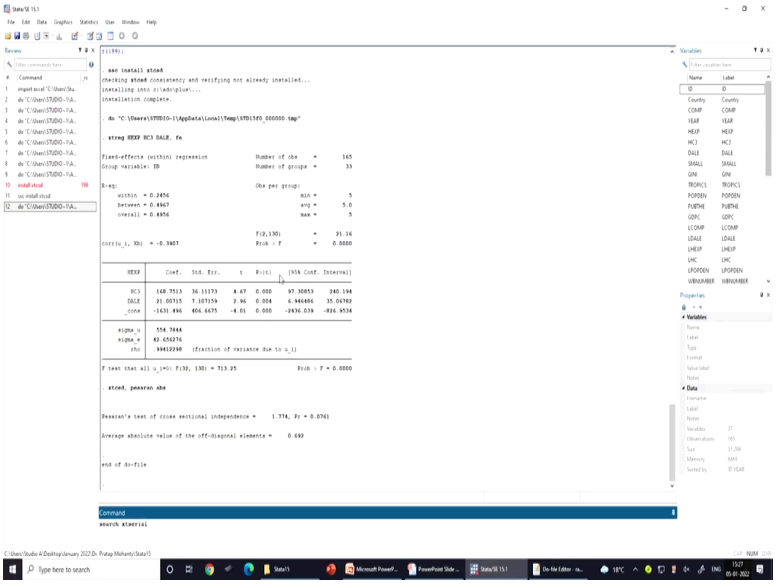
(Refer to Slide Time: 21:24)



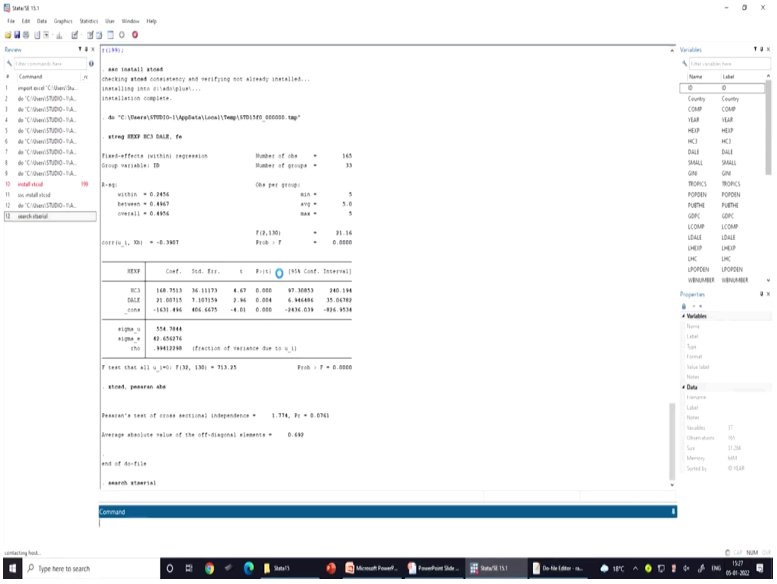
I am just going to show it over here like this. These are the command,

basically, what I am saying we need to install the program. We need to search for xtserial then net sj 3 hyphen 2 st 0039, then we need to install that particular st 0039 on our system. then only it will work.

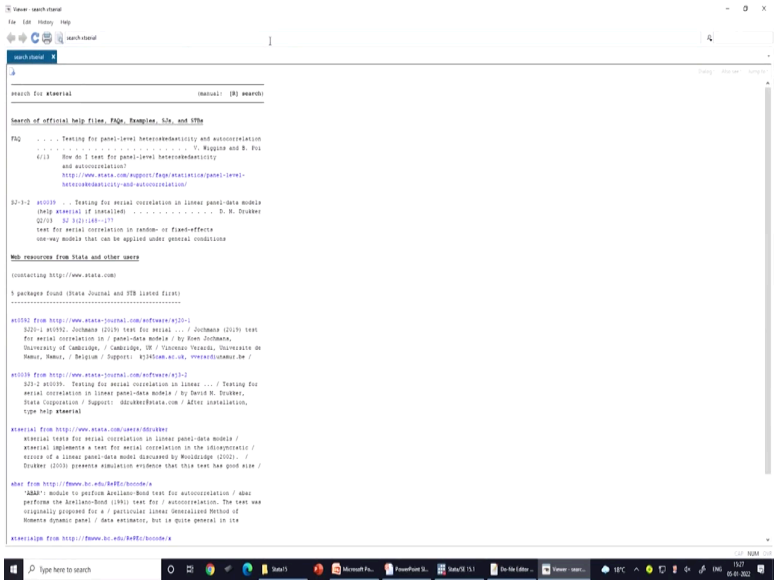
(Refer to Slide Time: 21:51)



(Refer Slide Time: 21:58)

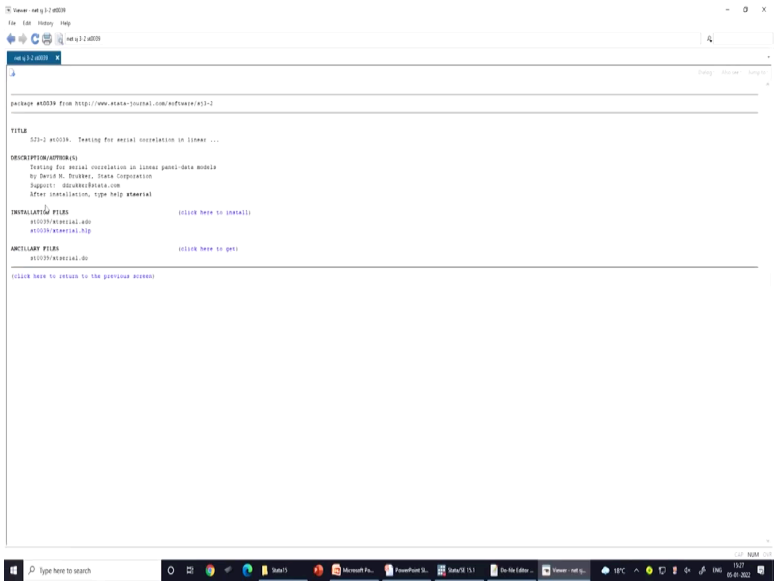


(Refer Slide Time: 22:00)



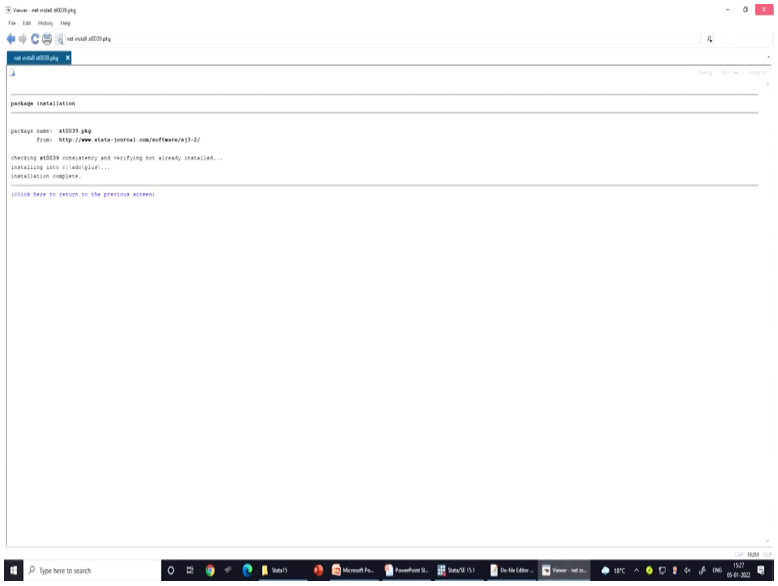
now this is being searched and we will xt and next one xt 0039.

(Refer Slide Time: 22:12)



this is to we need to install it.

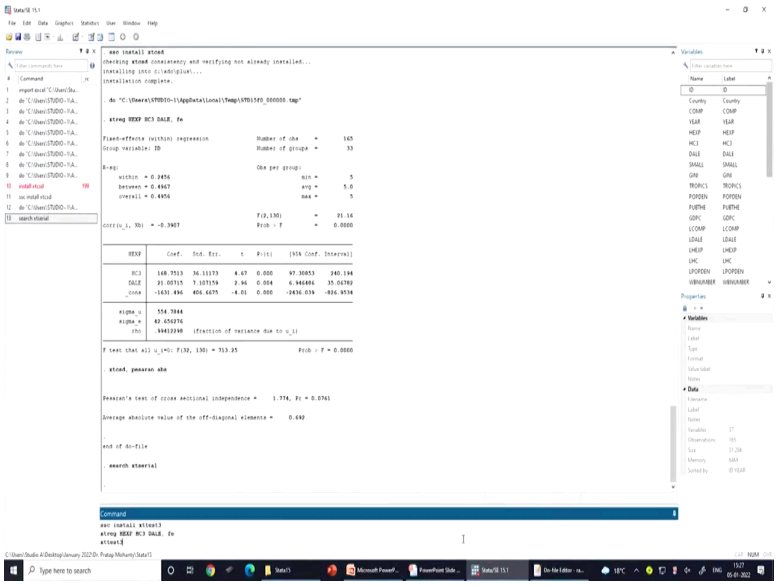
(Refer Slide Time: 22:16)



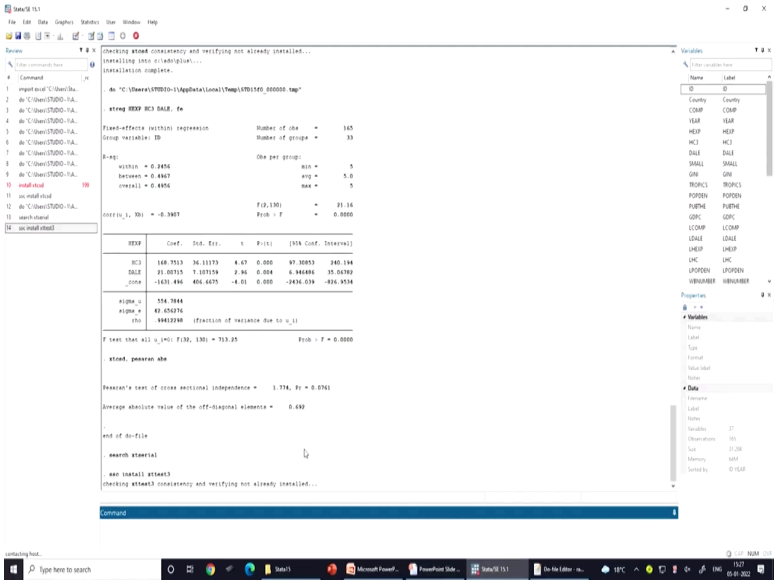
And once it is installed, yes now it is complete we can run that command.

The command is to understand the serial correlation. so, that is basically, ssc install we did it then xtrag.

(Refer Slide Time: 22:36)

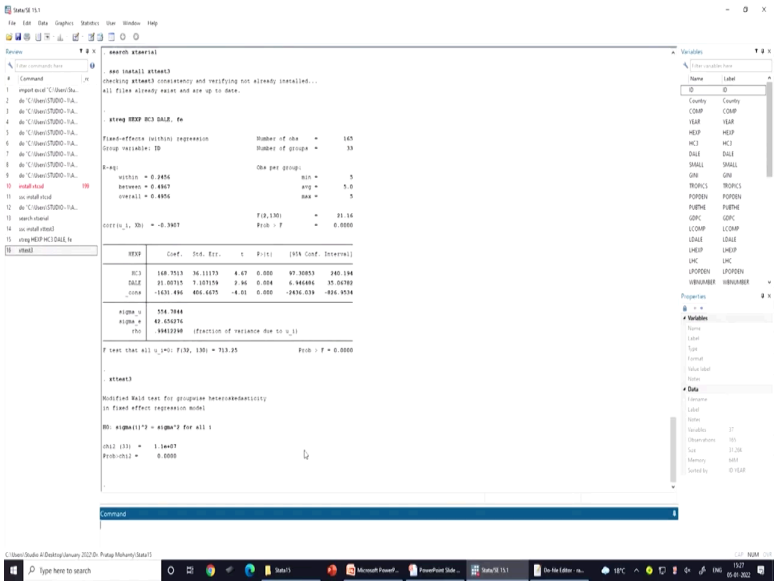


(Refer Slide Time: 22:38)



we need to compare this three.

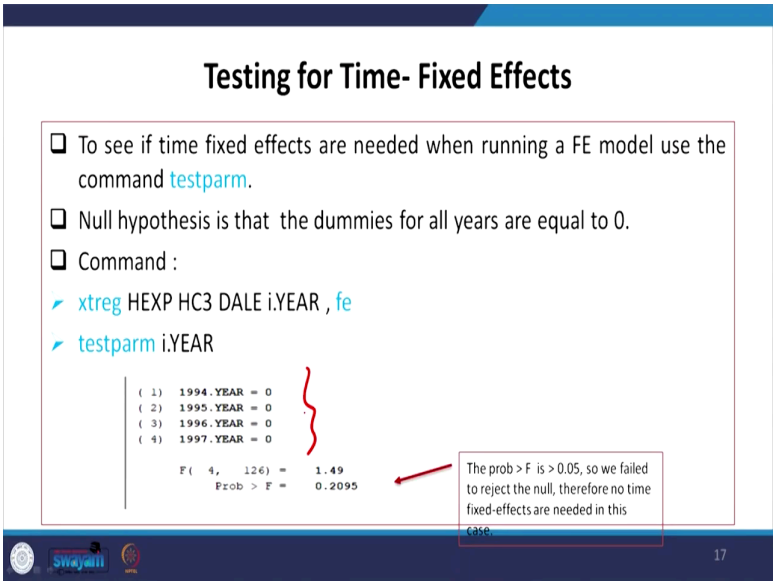
(Refer Slide Time: 22:42)



now we have got the result on your screen you can see that these 3 steps you have to follow and everything is on your screen. Even we have also, kept all those things systematically now, xtserial the 2 variable we have taken, and now basically, this is also, called as Wooldridge test for autocorrelation in panel data, autocorrelation in panel data

And the hypothesis is that there is no first-order autocorrelation and since it is significant; that means, there is a serial correlation, and a first-order serial correlation exists in the data.

(Refer Slide Time: 23:34)

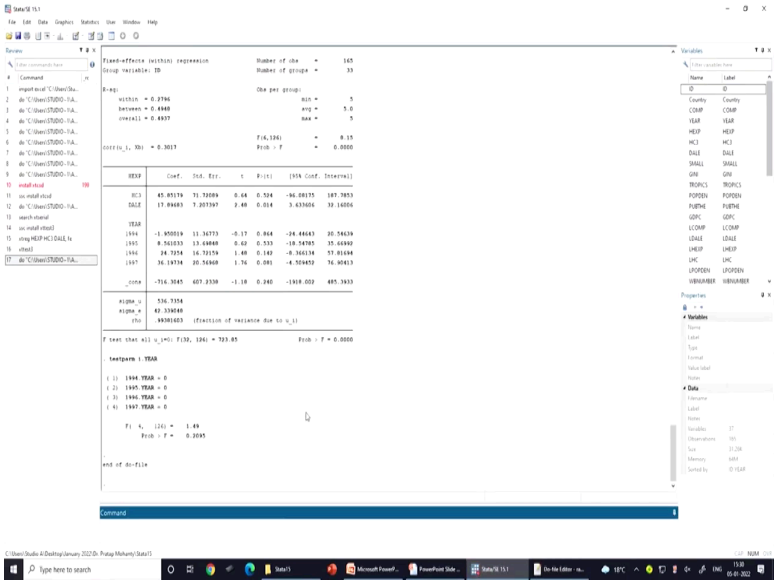


so, now this is the one to see if time-fixed effects are needed when running a, fixed effect model, we need to use the command testparm. based on that serial correlation we can identify whether there exists serial correlation or not. The next one is to understand whether it has a certain time fixed effect or not. the command standard command we take is called testparm. the null hypothesis in this context is that the dummies for all years are equal to 0, the dummies that are taken for all the years of the dummy are actually equal to 0. so, the command is in fact we have taken here as i dot is the dummy. For a year all the year's dummies are equal to 0.

so, testparm i dot year we have taken and based on that we can see that the probability value p-value is greater than point 0.05, we will also, see that. so, in this case, we know we fail to reject the null hypothesis. Therefore, no time fixed effect is needed in this case, you can also, check it with the data this is presented in our do file as well.

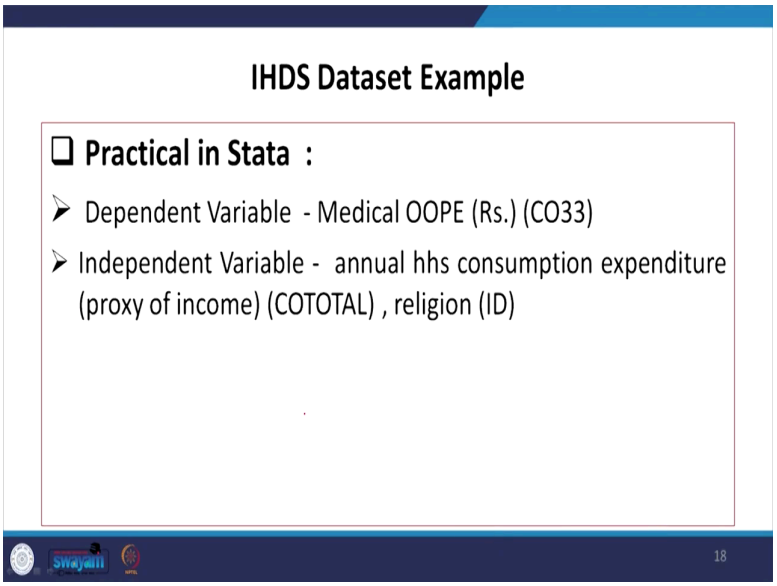
here is the do-file you can draw the do file, to understand the time fixed effect.

(Refer Slide Time: 25:16)



And so, we find that we can see in the result that it is not significant. so, there is no time-fixed effect, in the model, alright. And time-fixed effects are needed in this case.

(Refer Slide Time: 25:34)



Therefore, we can ignore that and it is not necessary to run further.

After saying all those basic or important steps in the random effect model. we have guided the, with guided with the help of the W H O balanced data example data set. You can al so,

run on your own through the IHDS dataset example data set, we have already uploaded to your screen.

And that is given with the name practical in Stata. The dependent variable is medical out-of-pocket expenditure and which is in rupees and the independent variables are given as annual household consumption expenditure, which is a proxy of income and then religion etcetera, religion is the ID variable.

You can just check this I am not experimenting we have already given everything on your folder and for your EG operation, we are keeping this data set for you to understand random effect health care random effect panel data in health care.

so, these are all here for you we lo forward to your participation in the next class.

Thank you.