Applied Econometrics

Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology-Madras

Lecture - 28
Course outline for Applied Econometrics

Panel Data Model Estimation - Part IX

So, we will take another data set, here it is, yeah, first we will look at the data. So, this is particularly, the data is, data looks like this. This data is on individual level data, their wage, their colour, whether they are experienced, whether they are Hispanic, non-Hispanic, how much hours they have supplied and then whether they are married or unmarried, so on and so forth, different types of this. So, this is the data. Now what we will do, we will first estimate, we will first estimate the model. So, we have to explain STATA that this is a panel data, xtSet, what is the panel variable here, nr.

So, this data we are using from the Ulrich database and the files name, data files name is wagepan.dta, wagepan.dta, I will be sharing the data with you, that is not a problem. So xtSet, nr, er, we have explained that this is a random effect model and then first we will estimate the random effect model.

So, xtReg, xtReg, lWage equals to education, whether the individual is black or white, then whether Hispanic or non-Hispanic, experience, experience square, and then whether married or not and whether that individual is a part of any labour union or not. So after that we will put re and to get an estimate of that λ transformation, in STATA you need to put a specific command called theta, after random effect, after re you need to put theta. What is theta? STATA's theta is equivalent to the λ factor what we have discussed just now. So this is a random effect estimates. Now STATA is reporting theta equals to

0.6426. Since theta value is 0.6426 what we can understand that means the λ is tending towards 1, it is not exactly 1 but we can say that it is tending towards 1. So what does our theory says? When λ is tending towards 1 then random effect estimate should more or less converge towards fixed effect not pooled values because it will tend towards pooled values when lambda tend to 0. Here it is λ equals to 6426.

So we will see whether that is true or not. So now we will estimate same model using

fixed effect transformation. Now when you estimate fixed effect model see education black, Hispanic, non-Hispanic these qualitative variables are dropped. Why? Because these variables are not changing over a period of time. These are all dummy variable that is why they are dropped which is quite understandable from the theory.

But the time changing variable for example let us say experience what is the coefficient 0.1168. Here it is 0.1012. So it is almost, sorry the experience is 1121.

So 1121 and here it is 1168. So they are almost similar and that is what our theory also said that when λ tend to 1, λ tends to 1 if Re will tend to Fe that is what is happening here. Now what we will do? We will estimate the same model using pooled values. So this is our pooled OLS estimates. And what is the value? Look at the coefficient of experience it is 0.

08. So that means compared to Re which is 0.11 this is 0.08. So quite a lot of difference between Re and pooled OLS and that is understandable because the theory says that a random effect model will converge to pooled OLS only when λ tends to 0. But here the value of λ is 0.

64 which is tending towards 1. So that is what we understood in the theory and what we are getting here that is actually matching. Our results are matching with the theory. And one more thing when you estimate a random effect model see here Stata is reporting random effect GLS regression. What we said? The application of OLS in the transformed model is called generalized least square and that is what Stata is reporting.

Is that fine? So now we understood how to estimate fixed effect model, how to estimate random effect model. But the question is now, the question is given a particular data set which model is then to use? We have learned both fixed effect and random effect model which model to use. So that is our concern. How to select between FE fixed effect or random effect? Now Hausmann, we generally use Hausmann test in this case. So Hausmann test, Hausmann's null hypothesis is basically covariance between X it and ε i equals to 0.

That is the null hypothesis. So that means Hausmann test is basically indirectly assuming that the model is fit for random effect model. And what is the test statistic? Test statistic, the idea in Hausmann test, Hausmann said that we will estimate the model using both fixed effect and random effect model and then we will see is there any significant difference between the fixed effect and random effect model. So Hausmann test statistic is defined like $\hat{\beta}$ FE minus $\hat{\beta}$ RE transpose / by variance of $\hat{\beta}$ FE - variance $\hat{\beta}$ of RE into FE. FE $\hat{\beta}$ minus RE and that will follow a chi-square distribution with k

degrees of freedom.

What is k? k is the number of explanatory variable that change over time. So again if the calculated, these are all in matrix notation because you can understand sometimes we are using transpose, sometimes we are using inverse. So basically these are all in matrix notation. This $\hat{\beta}$ FE and $\hat{\beta}$ RE is basically a difference between two matrices and then we will take transpose of that. Here again var($\hat{B}$) FE, variance of var($\hat{B}$) RE and then we will take inverse of that and multiply it by var($\hat{B}$) FE - var($\hat{B}$) RE.

That is basically the Hausmann test statistics which is purely based on the difference. Here it is difference, here it is difference in variance, here again it is difference. So we are trying to understand is there any systematic difference in the estimates between FE and RE and based on that we will decide. So what we have to do now, we have to estimate both the models using, we have to estimate both FE and RE and what we will do now, this is basically, now we will estimate this model. So first we will use once again the FE model.

This is the FE model. FE XT reg we are using and then after FE. Then we have to store this result. So EST store fixed. So I am asking Stata to store my fixed effect coefficients. What is the command I am using? EST store fixed.

Next what we will do, we will estimate the random effect model and then we have to ask Stata to store the random effect EST store random. So random effect coefficients are also saved. Then what we will do? We will put the Hausman command, Hausman fixed random. Now what is the test statistic? Here as we said test statistic is chi-square and that will follow chi-square what would be the degrees of freedom number of variables which are changing over a period of time. Here you see experience, experience square, merit status, marital status and union.

These are the variables changing over a period of time that is why it is degrees of freedom is 4. And what is the p-value corresponding to chi-square is 0 0 0. So that means we can actually reject the null. In Stata the null hypothesis they have written difference in coefficient not systematic. So once it is rejected from Stata's output it is now you make it confused whether to go for Fe or Re.

Now when you get confused you look at the way we have made our null hypothesis. What is the null hypothesis? Covariance between X it and the unobserved effect is 0. So if that is rejected that means actually unobserved effects are correlated with the explanatory variable. So when H naught is rejected, H naught is rejected you can very well understand what would be our conclusion that means we are actually looking for

fixed effect model. Fixed effect is applicable as covariance between X it and epsilon i is actually that means not equals to 0.

And that is the assumption what we made in the context of fixed effect model. So do not get confused with Stata's this thing. So Stata is saying the null hypothesis the difference in coefficient not systematic. Now why Stata is putting the hypothesis in this way? So when this is rejected that means there is actually a systematic difference between Fe and Re. So if there is systematic difference in Fe and Re we are going for Fe.

If it is not then we should go for random effect because in fixed effect model if you look at the other version of Fe is Lsdv. So that means if there is no systematic difference why unnecessarily including too many dummies in the model? Why unnecessarily going for fixed effect model when there is no systematic difference between fixed effect and random effect model? In this case it is rejected. So there is systematic difference between Fe and Re. So we should go for fixed effect. If not if it is not rejected we have to keep in mind that we must go for a random effect model.

So once we estimate random effect and fixed effect model then we can actually get a more clearer picture about step by step what we should actually follow. Now when we do our econometric analysis first of all we need to see whether our sample is a random sample or not. Yes or no? If it is no if the sample itself is not randomly drawn from the population then please keep in mind we will never go for random effect model fixed effect is only solution. Now when this type of cases arise when your unit of analysis is very large for example let us say we are doing an analysis with Indian states obviously you cannot draw randomly certain states to do the analysis. If I am doing a panel data analysis of all major states in India that is not a random sample.

If I am doing an analysis with OECD countries that is not a random sample. We are not drawing randomly. If I am doing an analysis with BRICS countries that is not a random sample. So whenever the sample is non random not randomly selected from a given population the solution is Fe only. If yes if the sample is yes then we can either go for Fe or Re depending on the Hausmann test.

So here I am applying Hausmann test. And in Hausmann test again we will get two cases H0 rejected H0 not rejected. And what is H0 I will mention once again here. If it is rejected then what will happen if H0 rejected then we will go for fixed effect. If it is not rejected then what we will do we will use random effect model but provisionally. Why provisionally? See when it is not rejected that means we are saying that the unobserved effect e i is not correlated with X it.

But we have not checked whether e i is significant or e i not significant. Now how do you select e i is significant when the variance of e i that means σ^2 e in the total variation which is given by some row in status output is significant. So that means what I am saying the variance in of the unobserved effect in total variance total variance is given by σ^2 e $+\sigma^2$ u that is the total variance of the error term. So what is the out of that total variance what is the contribution of the unobserved effect? If the contribution of the unobserved effect is significant then only we will go for a random effect model. If it is not significant then we will go for pooled OLS.

And what test we will use to see whether e i is significant or not significant that means at this stage Brouche-Pagan LM test B R E U S C H Brouche-Pagan Lagrangian multiplier test. I am not going into the detail of the test but we need to apply to see whether the contribution of the unobserved effect is significant or not. If it is significant then obviously we will go for random effect model. If it is not significant that means when the unobserved effect model is not significant why unnecessarily going for random effect transformation because we have selected R E only provisionally. So now we have to apply pooled OLS because I told you earlier that OLS is the most powerful econometric model unless it is required we should never deviate from OLS.

So when there is no presence of significant unobserved effect we will always go for random effect model. And how do you apply this Brouche-Pagan test? Brouche-Pagan test actually you can apply after estimating a random effect model. For example we will again estimate the random effect model. This is the random effect model after this we will put X t test 0 this command. This is Brouche-Pagan Lagrangian multiplier test and that test shows here you do not have to see all this.

Look at that is variance of U equals to 0. Stata's U is actually in our theoretical model what we discussed is actually the unobserved effect e or A you can think of. So Stata is assuming what is whether variance of U that means whether there is presence of significant unobserved effect or not and that is rejected. So that means actually unobserved effect has significant contribution in the total variance. That is why fixed effect model is actually suitable sorry random effect model is suitable in this case. You can get a sense of this you can get a sense of this row which is nothing but look at here.

Earlier we have not discussed. Look at the σU σE and ρ. ρ is actually defined as (σ^2 U/ σ^2 U + σ^2 E and that is this. So here it is 46 percent so that means the unobserved effect has 46 percent contribution in the total variance. If it is too low if the ρ value is too low then we need to test go for this Lagrangian multiplier test suggested by Bruges and Pang.

So this should be your entire flowchart. So first of all random sample or not if yes then only there is a question of checking fixed or random effect if no then only fixed effect. If Fe or Re we have to apply Hausmann test if H naught rejected what is H naught this is our H naught covariance between epsilon i and xi t is 0. If that is rejected we will go for fixed effect. If it is not rejected we will select random effect provisionally. At this stage we have to test whether e i is significant e i is not significant.

Instead of e i you can even think of variance of e i in stata's language. If that is not significant then we will go full wireless if significant that is suggesting then random effect and this test we will do by Bruges-Pang and Lagrangian multiplier test. And in stata's output, Stata will actually give you everything what is the contribution of this ρ. And by xt test 0 what is the meaning of this command? The meaning is we are actually testing for significant xt.

What is xt? xt is basically panel. Is there any significant panel effect? Panel means the moment we go for panel either Re or Fe actually we are assuming significant presence of Ai. Please keep in mind in the entire panel data discussion the major player the hero of this entire play is actually Ai or e i. So when I am testing xt test when I am putting that command xt test 0 I am actually testing for significant panel effect. That is the meaning of xt test 0. That is why Stata's hypothesis is variance of u equals to 0.

So with this you can actually check whether we should go for random effect model or pooled OLS. Since it is rejected that means there is significant presence of random effect model, Ai and we should go for random effect model. So with this our discussion on pooled OLS fixed effect model random effect model is over. We learned the theory as well as we learned how to estimate. I will share the data with you and now you can actually play around with the data estimate all three types of model and then you please read the textbook also and understanding very much clear to you.