Applied Econometrics

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

Lecture - 27
Course outline for Applied Econometrics

Panel Data Model Estimation - Part VII

So, welcome once again to our discussion on panel data econometrics and yesterday we were learning about how to estimate a fixed effect model and how to interpret the coefficients, what are the different interpretations of 3 R square measure what we get after estimating a fixed effect model. And then we have also compared the fixed effect coefficients with LSDV and then we said what we observed the coefficients are actually same in magnitude as well as in their standard error. But LSDV model is practically very difficult to implement when you have too many observations right because it includes too many dummies in the model right. But one thing we have not discussed about the R square measure in LSDV and fixed effect model and that we would like to see today in our estimates. So, we are going to use the same data set which is basically the impact of labour or job market training on employees performance. So, this is Xt set once again Xt set F code and then F code and then ER.

So, what we will do we will first estimate the fixed effect model which is Xt reg then we have we have performance measured by L_scrap and then we have grant and we have grant previous year and then we will say that this is basically the fixed effect model that we are going to estimate. So, this is our fixed effect model we have estimated and from this what you can observe that within R square basically which is most relevant for the fixed effect model because that explains what is the total percentage variation time variation in Y we can actually explain by time variation in X that is mostly 17.34 percent. Now the same model if we estimate by LSDV sorry this is once again.

So, instead of Xt reg now we have to put reg command. So, this is basically reg regres then L scrap then grant and 1 and i dot F code that is the command when you are using LSDV model and here we are considering one way error component model. So, there is no time fixed effect. So, this is the LSDV model we have estimated and now if you look at the R square of this LSDV R square is 92.51 percent whereas in fixed effect model here it is only 17.

34 percent. Now if you look at the R square measured that means LSDV gives almost 5 times higher R square than the fixed effect model does that mean an LSDV model is always preferred to fixed effect first of all why is the R square so high in LSDV model can you think of why the R square is so high in LSDV. So, this is the question I am asking why the R square. So, this is Fe versus re LSDV and my question is why is R square LSDV is so higher than as compared to R square Fe. Now the answer is very simple because in LSDV we are actually including too many explanatory variable in the form of too many dummies.

So, obviously when you include too many dummies that means you are almost explaining the unobserved heterogeneity quite well by including too many individual specific dummies that is the reason R square is too high in LSDV. So, there are too many explanatory variable that is the reason there are too many explanatory variables in LSDV in the form of individual specific dummies. As a result we are able to explain a significant portion of YIT that is the reason. But that does not mean that LSDV is always preferred to Fe because if you include too many explanatory variables obviously you will be able to explain a larger variation in the dependent variable but based on the R square we will never select Fe and LSDV there should be a proper justification. When there are too many observations we cannot actually apply LSDV because of the degrees of freedom problem and it would be practically impossible to estimate sometimes.

That is one thing I would like to mention upfront. Now in the fixed effect model when we write our equation is $Yit = \beta 1xit + Ai + Uit$. So, this is the model and the assumption what we made that covariance between XIT and Ai is actually not equals to 0 for Fe to be implemented. Now suppose covariance between XIT and Ai is actually 0. There is no correlation between the unobserved heterogeneity individual specific time constant effect and the explanatory variable.

None of the explanatory variable is actually correlated with the unobserved effect. So, in that case that fixed effect transformation is not applicable because if we remove that then what will happen we are actually calling for some kind of inefficiency in our model. So, when that is not correlated we need a specific form. So, when the next question is then if our model is $Yit = \beta 1Xit + Ai + Uit$, covariance between Ai and UIT is actually 0. Then the first question that comes to our mind should we work with a single cross section? We can actually work with a single cross section when unobserved effect is actually not sorry here I made a mistake here this is XIT.

When the unobserved effect is actually not correlated with the explanatory variable we can always work with a single cross section because the major main reason for using panel data if you recall I said that if you use more than one period's data then actually we can eliminate Ai by first difference or fixed effect transformation. That was our objective. So, we included Ai in the model and we have also accommodated the fact that this Ai is actually correlated with your XIT. Now if they are not correlated then there is no need of panel data we can actually work with a single cross section. But if we work with a single cross section then what will happen the other advantage of using panel data we cannot enjoy and what is the other advantage? Other advantage is capturing multiple distributions for the dependent and independent variables.

When you work with a cross section you will be able to capture only a single cross section for both YIT and XIT. So, that is the reason. So, we can work with a single cross section in this context but we will not be able to capture multiple distributions of YIT and XIT in cross sectional data. So, we will not be able to enjoy the other benefits of panel data. So, because of that we still need to work with panel data but fixed effect and first difference transformation will not work we need to have a different model known as random effect model.

So, that means in random effect model please keep in mind the main assumption is actually this that unobserved effect individual specific time constant unobserved heterogeneity is not correlated with the explanatory variable that is the assumption. Now random effect model we can get starting from a fixed effect model only. So, in fixed effect model our equation is YIT equals to beta 1 XIT plus ai plus uit. Now what I will do I will decompose this unobserved effect ai as β0 plus εi. What is this β0? β0 is actually the common unobserved effect what I can say that this is average and epsilon i is epsilon i.

So, β0 is the average or common unobserved effect and εi is a random deviation from β0. So, if we think of ai is basically ability which is the unobserved factor β0 indicates what is the average ability in that sample and ε i is basically a random deviation from that common ability factor. So, εi is basically the ith individual's random deviation from the sample average ability factor it can be positive or it can be negative. So, if we decompose ai into β0 and εi then our model would become YIT equals to β1 XIT or I will first write the intercept $\beta 0+\beta 1 Xit+\varepsilon i+ui$. And these together sometimes you can write VIT composite error term which is indicated by εi and ui t.

Now we cannot estimate this model so that means this becomes β0 plus β1 XIT plus VIT. Now when this ai this ε i and the assumption what we make here covariance between ε i and XIT equals to 0. Now in this model can we apply pooled OLS can we apply pooled OLS. Since XIT is not correlated with ε i can we apply pooled OLS. Now pooled OLS is not applicable why because this ε i we cannot apply pooled OLS as ε i will always be there in VIT for all t and as a result of which VIT will be serially correlated leading to autocorrelation.

So, pooled OLS is not applicable this is the justification. Then what we will do we will make a different type of transformation when our model is this. So, YIT equals to β0 plus β1 XIT plus VIT equals to ε 0 plus UIT so that is basically VIT. So, what we will do instead of removing the entire ε this is sorry this is ε i. Now what we will do we will use this transformation $Yit - \lambda\overline{Y\iota} = 1 - \lambda\beta 0 + \beta 1 Xit - \lambda\overline{X\iota} + vit$ - $\lambda\overline{V\iota t}$ and what is this λ how is this λ defined λ equals to (1- σ ^2 U) / σ^2 U + t into σ^2ε to this.

Where σ^2 U is basically variance of U and σ^2ε is basically variance of ε. Now, the question is then to get this transformation we need to have $\hat{\sigma}$ sorry this is λ. λ is always unknown but can be estimated using $\sigma^2\hat{u}$ and $\sigma^2\ \hat{\varepsilon}$ from FE or LSDV. So, this $\sigma^2 u$ $\sigma^2 e$ they are all population unknown population parameter but once we estimate a model either using fixed effect or LSDV model we can always get their sample counterpart which is$\sigma^2\hat{u}$ and $\sigma^2\ \hat{\varepsilon}$ and using this we will get $\hat{\lambda}$. So, this is the random effect transformation this is called random effect transformation RE transformation.

That means instead of removing the entire yi bar what we are doing this is called quasi time demeaning. Fixed effect transformation is called time demeaning because that was Yit - $\overline{Y\iota}$ here we are removing only a fraction of $\overline{Y\iota}$ λ($\overline{X\iota}$) that is why instead of calling time demeaning random effect transformation is called quasi time demeaning model. So, with this what we are doing we are solving the auto correlation problem at the same time we are solving the auto correlation problem that we were having in the model of pooled OLS. So, this transformation is known as when we apply OLS in this transform model that is called generalized least square. So, application of OLS this is called GLS which is nothing but application of OLS in the transform model.

And the transform model is basically this model which is the quasi time demeaning because this Vit -λ this is vi bar lambda vi bar is basically free from the auto correlation problem. Free from the auto correlation problem that is how when the unobserved effect is actually uncorrelated with the explanatory variable instead of using pooled OLS or fixed effect transformation what we do actually we use the lambda transformation that means a lambda fraction of yi t and xi t sorry yi bar and xi bar is removed from yi t and xi t. Now this random effect model if we look at very carefully we can understand a random effect model is the more generalized version of pooled OLS and fixed effect model. We can prove that for example let us say two alternative cases. Case one let us assume that sigma square epsilon is sigma square epsilon tends to zero that means which implies insignificant unobserved effect.

So if sigma square e sigma square epsilon tends to zero from this lambda expression of lambda you can understand this becomes 0 then this become σ^2u divided by sigma square u square root that also become one so lambda will become equals to 1 – 1 equals to 0. So if that is the case if λ is 0 then our model becomes yi = β1xit +vit. When lambda becomes zero and this is nothing but pooled OLS which is quite reasonable that means what does this transformation what does this particular case indicate if at all the unobserved heterogeneity is insignificant in our model why should I go for fixed effect or random effect transformation? Because the entire idea behind using the fixed effect and random effect transformation was a significant presence of the unobserved effect ai if that itself is insignificant then we can very well apply the pooled OLS because that was only creating problem. If you look at why pooled OLS was not applicable because in this model this epsilon i will always be common in this v i t so if we apply pooled OLS then v i t will be auto correlated serially correlated so if that is not significant then instead of using fixed effect or random effect transformation what actually we should do we can very well use a pooled OLS model. So that means we can understand that pooled OLS is a specific case of this random effect transformation when lambda equals to zero we get pooled OLS

Similarly case two when t total number of time period tends to let us say infinity you have a large number of time periods then what will happen if that is the case then lambda will become 1 minus when t tends to infinity these denominator t multiplied by sigma square e also become infinity and sigma square u when it is added with infinite that is also become infinite and σ^2u/ ∞ = zero so 1 minus 0 equals to 1. When lambda equals to 1 this transformation becomesYit - $\overline{Yi}$ = β1xit - $\overline{Xi}$ + v i t - $\overline{Vi}$ which is nothing but fixed effect model. So that means with these two cases what we can say that pooled OLS and fixed effect pooled OLS and fixed effect models are special cases of the random effect model because both the models can be derived from the random effect model assuming a specific value of λ. So what we can say that both pooled OLS and Fe are special cases of random effect model.

That we have to understand. So when λ actually tends to 1 then random effect estimates will also tend to fixed effect estimates when lambda equals to 0 we can understand that random effect estimates will become will tend to the pooled OLS. Now what we will do we will take one data set and we will try to understand how to estimate a random effect model and their interpretation.