Applied Econometrics

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

Lecture - 57
Course outline for Applied Econometrics
Dynamic Panel data Model - Part XIX

Welcome to our discussion on Dynamic Panel Data Model. And today what we will do, since we have already completed all the discussions about the theory and I have also demonstrated using the statistical software how to estimate a dynamic panel data model and different issues involved while you estimate. In today's class I will just summarize from the beginning so that you will be able to know what are the key elements that you should keep in mind while estimating a dynamic panel data model. So, this is a summary of entire discussion that we had in dynamic panel data model. So, this is dynamic panel data model that we are going to discuss. So, our model looks like

$$y_{it} = \rho\, y_{it-1} + \beta\, x_{it} + u_{it}$$

where ui t is actually summation of the turn offs of individual specific effect plus the idiosyncratic error.

Firstly, we need to keep in mind before applying dynamic panel data model we need to first justify theoretically why do you think that you need a dynamic panel data model in your context. So, your situation should be of partial adjustment type that means for any changes in your explanatory variable xi t, yi t should not be able to adjust itself in the same period. If it does then it would be a case of instantaneous adjustment. When it fails to adjust in the same period then only we call it a partial adjustment model and that partial adjustment model the other name of is dynamic panel data because by partially adjusting dynamism is introduced in the model.

So, a simple transformation that we have discussed is if we subtract y_{it-1} from both the sides of this equation 1 then what we will get?

$$y_{it} - y_{it-1} = 1 - \rho y_{it-1} + \beta x_{it} + u_{it}$$

So, this is the model and by partial adjustment what we mean that this ρ should be less than 1. ρ should be less than 1 and that means if ρ is less than 1 then only we get a partial adjustment model. ρ equals to 1 means this term will be 0 that means the lag dependent variable will disappear from the model and that will imply an instantaneous adjustment. And model which is basically the static panel data model and ρ greater than 1 is unstable.

So, this is the situation of partial adjustment. There might be many reasons for getting partial adjustment or dynamic panel data model. We have discussed only one case where in we were discussing merely the employment data. Employment of a firm or employment by any sector that depends on what is the employment you are already having because what we discussed is the hiring of new employer and firing the existing one is costly. That is why employment depends on your previous year employment

$$y_{it} = f(y_{it-1})$$

There might be other cases. So if you want to model let us say a firm's output even that is also kind of dynamic in nature because your this period output may depend on previous period's output how much you have already produced how much you could sell and how much was there in your inventory. There might be many cases for that. So we need to carefully justify our need for a dynamic panel data model. Otherwise when we just because we have some data and we know the code about how to estimate it in the statistical software we should not use when it is not required.

We need to carefully understand the situation and justify. Now in case we have a dynamic panel data model then what happens here that the main feature of this model I am writing this model once again

$$y_{it} = \rho\, y_{it-1} + \beta\, x_{it} + a_i + v_{it}$$

I am writing in this form. Since, $y_{it} = f(a_i)$ and there is no t subscript over here what we discussed that $y_{it-1} = f(a_i)$ is also correlated with ai that means it leads to endogeneity because ai and vit constitute the composite error term. So there is endogeneity in this model and OLS is not applicable. Then when OLS is not applicable if we do still estimate the model using OLS basically we will end up with having a dynamic panel bias and we discussed that OLS actually leads to in this situation when you have dynamic panel data model if you apply OLS that will lead to overestimation, of $\hat{\rho}$.

To remove this ai if we apply the fixed effect transformation like what we used to do in the context of static panel data model then what we discussed that a fixed effect model will lead to underestimation of $\hat{\rho}$. So that means OLS will give you an upper bound and Fe will also give a lower bound of $\hat{\rho}$. Is it clear? OLS is giving upper bound because we are ignoring that there is dynamism in the system and what is happening here that a positive correlation between this y_{it-1} and a_i and we have also given an example why y_{it-1} and yi is positively correlated because you can understand you can take a situation where the form is experiencing a negative employment shock then in the next period both ai the unobserved effect and y_{it-1} would be lower than what was happening earlier. So in all the consecutive periods rather your ai and y_{it-1} would be lower and as a result of which they will have a positive correlation that we have already discussed and in fixed effect transformation it will have a negative bias because the transformation when you make y_{it} - $\overline{Yi}$ then you will get v_{it} - $\overline{vi}$ a negative sign before that $\overline{vi}$ will lead to negative correlation between y_{it-1} and your v_{it} and as a result of which fixed effect transformation will give you underestimation of rho hat and these two will give you the upper and lower bound of your estimate. So true estimate of $\hat{\rho}$ then should be greater than your fixed effect but sorry it should be greater than fixed effect but less than your OLS.

So this is something whenever we are applying and estimating a dynamic panel data model we need to first check the upper and lower limit of your estimates and then applying dynamic panel data model estimation technique when you get a $\hat{\rho}$ then you will see whether your $\hat{\rho}$ is actually lying within the interval of Fe and OLS. Then we have also discussed that when neither OLS nor Fe is applicable the other transformation is first difference. So Fd transformation we can apply and that will make it

y_{it} - y_{it-1} = ρ y_{it-1} - y_{it-2} =βx_{it} - x_{it-1} + v_{it} -v_{it-1}.

So here what is happening even after first difference also we can see that y_{it-1} is here and v_{it-1} is here so that is why y_{it-1} that means this variable.

δ y_{it-1} is actually correlated with δv_{it} and as a result of which we are again ending up with endogeneity. So even first difference transformation also cannot rule out the possibility of endogeneity because this is correlated.

So what is the solution then we offered we said that when Fd is not workable because of this endogeneity we will use IV either δ y_{it-2} or δ yi t. These are the two instruments we can use and the method we said that this is the suggestion given by Anderson and Hesio he said we should take the first difference so that means this is the suggestion given by Anderson and Hesio. We should take the first difference and then use the instruments. Now one thing we should mention over here about the autocorrelation problem. In this model when we are writing

y_{it} = ρ y_{it-1} + β x_{it} + a_i + v_{it}

what we are saying that y_{it-1} is correlated with ai and what we can write that y_{it-1} is actually so I can write y_{it} = f(y_{it-1}).

Now the way I have written the equation I can write the same equation for y_{it-1} and that in that equation the error term would be v_{it-1}. So that means y_{it-1} = f(v_{it-1}). So that means from this what I can write then y_{it} = f(v_{it-1}). So that means v it is correlated with yi t, yi t is correlated with v_{it-1} then ultimately v_{it} $= f(v_{it-1})$.

So that means the presence of lag dependent variable is leading to autocorrelation and that autocorrelation is of order 1. So the structure of the model itself gives error 1. then error 2 and higher let us say error p should not be there. Now what will happen actually if error 2 is there? What happens when? What happens? So if you look at what is the instrument we are using? We are using y_{it-2} assuming that y_{it-2} would be correlated with δy_{it-1} but they will not be correlated with v_{it} - $v_i \delta v_{it}$. But in case there is higher order autocorrelation that means v_{it} is also correlated with in presence of error 2 v_{it} $= f(v_{it-2})$ also second order correlation is also there.

That means earlier what we were assuming that this y_{it-2} will not be correlated with this transform error term that will no longer valid because v_{it} itself is correlated with v_{it-1} and that means obviously $y_{it-2} = f(v_{it-2})$. So that means I am saying that $y_{it-2} = f(v_{it})$. So y_{it-2} can no longer be an instrument if there exist second or higher order autocorrelation. That is why when we estimate the dynamic panel data model we must check whether error 1 is there and whether error 2 is not there. Error 1 must be there because of the construction of the model.

If error 1 is not there that means equivalently we are saying that there is actually no need of estimating a dynamic panel data model because there is no importance of y_{it-1}. That is why by construction error 1 must be there but error 2 should not be there. If error 2 is there that means as if we are saying in presence of we should write it down in presence of error 2 neither y_{it-2} nor δy_{it-1} are valid instrument. Okay. That is one thing we need to keep in mind.

Then another important point we talked about dynamic panel bias and we also talked about nickel bias that means application of Fe and Fd transformation how that leads to nickel bias. Because both in case of both Fe transformation and Fd transformation we saw that we could not rule out the possibility of endogeneity and this nickel bias is of order 1 by T. So, that means as T tends to infinity this bias tends to 0. That is why whenever we will be estimating dynamic panel data model or DPD, DPD should be 0 and in that means the number of observations in the panel is large. That is also one thing we have to keep in mind.

If T becomes large then the bias will automatically get vanished. We do not have to estimate the model using this complicated dynamic panel data estimation technique. So, that is one thing we have to keep in mind that this is a situation where in my T is actually small n is large that means we are talking about a micro panel. So we will when there is this type of problem T is small n is large and we have lag dependent variable in the right hand side and then we are assuming that there is no AR2 we can go ahead with this estimation suggested by Anderson and Hesio wherein we will transform the model. We will take the first difference and we will use y_{it-2} or δy_{it-2} as instrument.

So what was happening in case of Anderson and Hesio's model? So Anderson and Hesio's main problem was that in their model since they said only y_{it-2} as iv. So this implies not all available movement conditions are utilized. Therefore we can increase the efficiency of $\hat{\rho}$ to a larger extent following this Hohl-Jakin technique. What they say that we will use y_{it-2} as instrument for all the periods and replace the missing observations by 0 because in Anderson and Hesio's approach the moment we use other lags also as instrument then we saw then your sample length was going down. So there was a tradeoff between lag length and sample length that we discussed.

So following Anderson and Hesio's approach the more lag you include the more observations we are going to lose. Then Hohl-Jakin et al they came up and say that we can actually use one single instrument y_{it-2} for all the periods y_{it-2} for all the period and replace the missing observations by 0. That was Hohl-Jakin et al's suggestion. But then again the original idea is to use more movement conditions whatever lags are available we need to use all of them to increase efficiency of $\hat{\rho}$. That is the reason next model is the most popular one came up by Arillano and Bond 1991.

So what they say that use y_{it-2} as well as all other lags as and when available that is what Arillano and Bond say. So for example in third period what will happen you will have only y_{it-2} that means yi 1 is available. But when you go to fourth period then we will have y_{it-2} that means yi 2 as well as yi 1 because y_{it-2} and y_{it-3} is also available. When you go to fifth period then y_{it-2}, y_{it-3}, y_{it-4} all these three are available. So as we move on to the higher periods we will get more and more lags and we will include whatever is available and missing observations as whole geogonetal suggested will replace by 0.

In that way Arillano and Bond they increase the efficiency of their estimates.