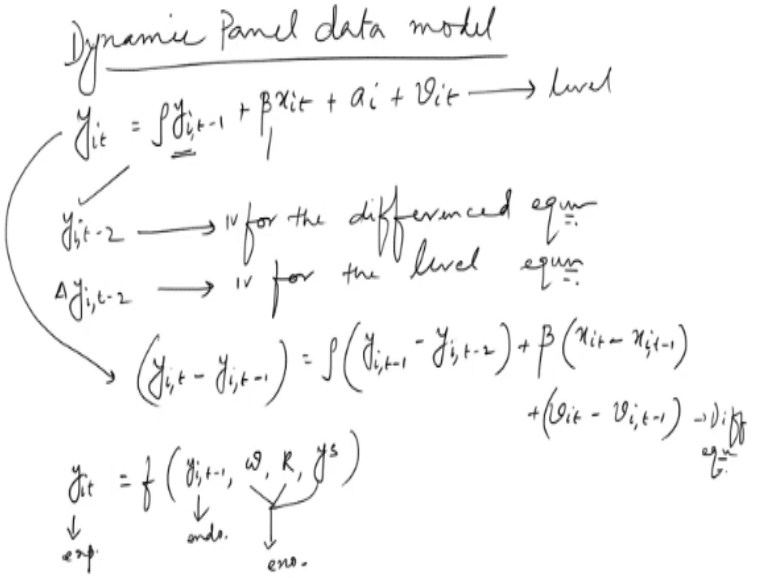
Applied Econometrics
Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology – Madras

Lecture - 52
Dynamic Panel Data Model – Part 15

Welcome to our discussion on dynamic panel data model and if you recall we were discussing a model wherein we are interested to estimate the determinants of firm's employment from the UK context. We have 140 firms in data and the situation is dynamic in the sense i[th] firm's employment in t[th] period depends on what the i[th] form has already employed in the previous period. Therefore, $y_{it} = f(y_{i,t-1})$. So basically, this was the situation we were discussing.

(Refer Slide Time: 01:09)



Dynamic panel data model and our relationship is $y_{it} = \rho y_{i,t-1} + \beta_1 \chi_{it} + a_i + v_{it}$ and this equation is called level equation and the presence of $y_{i,t-1}$ introduces dynamism in the system and in our previous class, we were discussing how to implement *xtabond2* command to estimate the system GMM. Before that, we estimated a difference GMM using $y_{i,t-2}$ as instrument for the difference equation and when we said that lag of level performs poorly as an instrument for the difference variable then we say that following.

Arellano and Bover (1995) and Blundell and Bonds (1998) were estimating a system GMM, using the status that *xtabound2* command and we said that differentiating system GMM from the difference GMM in terms of this command is very simple. In system GMM we are removing

the known level equation from the command which means I need a level equation also, so a level equation and the differenced equation both of them constitute a system of equations that we are going to estimate.

When we have a system of the equation then the lag of level is used as an instrument for the differenced equation and the lag of differenced variable that means $\Delta y_{i,t-2}$ is used as the instrument for the level equation. So once again, I will repeat that $y_{i,t-2}$ this is *iv* for the differenced equation and $\Delta y_{i,t-2}$ *iv* for the level equation and the difference equation is written in this way $y_{it} - y_{it-1} = \rho(y_{i,t-1} - y_{i,t-2}) + \beta(\chi_{it} - \chi_{i,t-1)} + (v_{it} - v_{i,t-1})$ this is called differenced equation. When we have both level and differenced equations to be considered as a system of equations then what we are doing we are using $y_{i,t-2}$ as well as $\Delta y_{i,t-2}$ as an instrument. This is for the difference equation and this is for the level equation so that the quality of our estimates improves drastically, and how do you measure the quality?

While difference GMM cannot always guarantee that my estimates lie between the theoretically determined upper and lower bound, we have ensured that in a same model system GMM can ensure that estimates are within the interval that means within the upper and lower limit of the interval determine by OLS and FE (fixed effect). Then what we were discussing yesterday that these estimates are quite sensitive towards the fact that what factor, what variable you are considering endogenous.

So far, we assumed that endogeneity is there only in $y_{i,t-1}$ that means lag of the dependent variable is only endogenous and all other remaining variables are strictly exogenous. In our context what we were discussing is that employment is a function $y_{i,t-1}$, then wage and capital, we assume that endogeneity is there, this is endogenous and these two are actually exogenous.

Rather we have another variable industrial output also we can take so this is also exogenous. Now if we assume that if labor and if wage and capital also become endogenous in this relationship, then if we model that our estimates they change drastically and there are enough reasons to believe that wage and capital are also endogenous in this relationship because as we all know employment determines capital.

At the same time, capital determines how many laborers you will employ, and at the same time how many labourers you employ that also determine how much capital to be employed in a particular firm. At the same time wage and employment, they also sometimes are determined simultaneously while wage we can say that is a labor demand function and employment is basically a labor supply function.

This labor demand and supply determine the equilibrium wage and labor supply that is coming from purely the demand and supply of labor. If that is the case, we can model *w* and *k* also as endogenous variables along with $y_{i,t-1}$. And if we do so then our estimates they change drastically and what we will do? We were discussing the sensitivity in the context of difference GMM and what we will do today we will once again discuss the sensitivity in the context of system GMM model also.

So that means we will re-estimate both difference and system GMM by modifying the specification which means which variable we will consider as endogenous and which variable we will consider as predetermined and exogenous. So once again, we will use the same data. **(Video Starts: 09:37)** This is the same data that we have been using Arellano-Bond's original data set from 140 UK firms. To access the data as I have already mentioned we have used ab data.

So first what we will do? We will first estimate the model considering only $y_{i,t-1}$as endogenous variable so *xtabond2 n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr**. So, this is my first specification. So, since I have been writing this command so many times by now, you must be habituated with this command *xtabond2*.

So, this line basically specifies your econometric model and next part of this command indicates which variable you are considering endogenous and which variable you are considering as exogenous and that exogenous and endogenous variable are differentiated by GMM style and *iv* style instruments. What we will do now? I will put my GMM style and within that I will put only *nL1* and then I have *iv* style and in *iv* style I will put all those variables like *wL1* and all exogenous variables *xtabond2 n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*,gmmstyle (nL1) iv(w wL1 k kL1 kL2 ys ysL1 ysL2 yr*) noleveleq robust small*, this is my model, here, I am considering only $y_{i,t-1}$ as the endogenous variable and look at this in this model I have 0.68 as the coefficient of *nL1*.

Here, the coefficient of my lag-dependent variable comes out to be 0.68 and as you know 0.68 is not lying within the theoretically determined bound by FE and OLS estimates, it should lie between 0.73 to 1.04. Therefore, system difference GMM by considering only that factor is not able to ensure my estimates lying within the interval. What we will do now? We will estimate the same model by considering the fact that my *wL1* and *kL1* they are also to be considered within the GMM style instrument.

That means wage and capital also are considered as endogenous now, what happens in that situation? So, we will put *xtabond2 n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*,gmmstyle (nL1 wL1 kL1) iv(ys ysL1 ysL2 yr*) robust small,* here wL1, kL1 are also now included in the GMM style.

This is my specification. Here, I am removing that no level equation, why I am removing this because, if I remove that means I am considering the level equation also as a system of equations, earlier I was considering only the differenced equation that is why no level, and if we do so then what is happening?

My coefficient is going beyond 1, this mean it is also problematic, if we estimate a system GMM with capital and wage also as endogenous along with the employment then it is again going beyond the limit, we need to modify. Therefore, what we will do we will try to estimate the same system GMM with a different model specification and we will take *xtabond2 n nL1 L(0/1).(w k) yr*, gmm(L.(n w k)) iv(yr*, equation(level)) robust small*

Here in this model, I have made a little change here, I have removed ys and I have considered only one period lag of employment and then estimating the same system GMM considering again n w and k all three are endogenous unlike the previous case where n was considered to be endogenous that means within the GMM it was only nL1. So L dot n w k we are putting within the GMM, then we are estimating the same model.

Now the coefficient is 0.93 which is lying well within the interval. So that means what I am saying whatever model you estimate whether it is a system GMM or difference GMM. Whether you consider only *nL1* as endogenous variable or you consider *w, nLl* and *kL1* also as endogenous at the end of the day my estimate should lie within the theoretically defined bound

and how do you get the bound, your FE and OLS estimates they set the upper and lower bound wherein your estimates from dynamic panel data model must lie, and if it does not then we need to modify our model.

How will you modify that is up to you as a researcher, you need to contextualize, you need to understand your problem properly, you need to read the theory to understand which variable should be considered as endogenous, which variable to be included in the model, all those things. And then if you see again the post estimation checkup the Arellano and Bond AR 1 is rejected but AR 2 is not rejected.

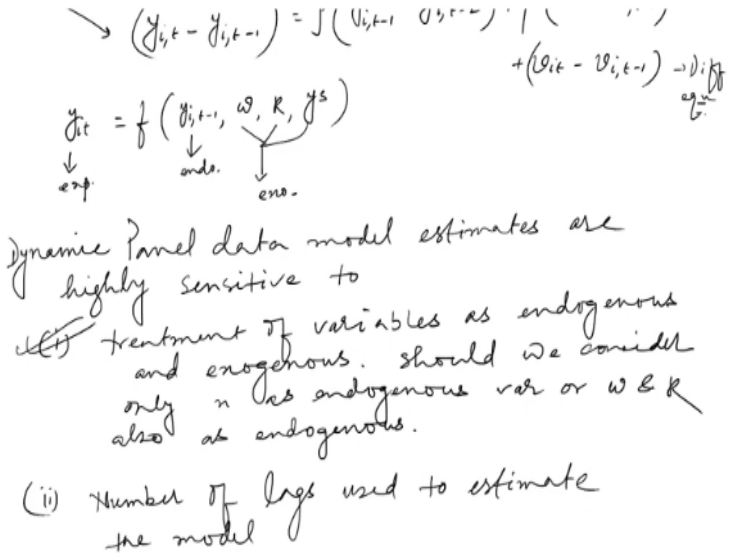
Which indicates there is the presence of first-order autocorrelation but there is no presence of second-order autocorrelation, and again the overidentification test Sargan and Hansen, if you recall yesterday, we discussed that the Sargan test as a test of overidentifying restriction the test statistic of Sargan test is undersized and extremely with low power in a context when you are using too many instruments and that is what is happening here.

How many instruments? There are 113 instruments, therefore, Sargan test becomes weak when you have 113 instruments and if that is the case then what is the alternative? We said that we will go for Hansen test which basically depends on that quadratic loss function or GMM criterion function and then test statistic is constructed by getting that quadratic loss function and total number of observation N into Q.

And that follows again a chi square distribution and that works better in the context when you have too many instruments. And then if you go here that Hansen test is satisfied. So Sargan test is mostly when it becomes weak then what happens, it rejects the null in almost all the cases when you have too many instruments and it never rejects the null in almost all the cases when T is extremely large.

So that is why we are going by generally we report the Hansen test of overidentifying restriction after this post-estimation checkup. And then as I said that this dynamic panel data model estimates are highly sensitive towards your model specification and two types of specifications basically, we are talking about. **(Video Ends: 26:44)**

(Refer Slide Time: 26:47)



Two types of specification, which variable you consider as indigenous, so I will write it down clearly. The dynamic panel data model estimates are highly sensitive to number one treatment of variables as endogenous and exogenous. So that means so do we consider only *n* as endogenous variable or wage and capital also as endogenous? So if your theory says no *w* and *k* can also become endogenous while estimating employment, then we must include *w* and *k* also within the GMM style bracket.

Which specifies what should be your GMM style instruments and what should be your standard *iv* style instruments. So right, so specification of this GMM and *iv* alters your coefficients drastically, so our dynamic panel data model estimates are sensitive to treatment of your variable itself whether you consider a variable as exogenous or endogenous. Secondly, dynamic panel data model estimates are also sensitive to a number of lags used to estimate.

See so far whatever dynamic panel data model we have estimated we have not imposed any restriction on the number of lags. **(Video Starts: 30:04)** If you see here in the GMM type instruments lag of L n, L w and L k it starts from first and it goes up to eighth order. So, when you are using that many instruments it may so happen that the quality of the instruments goes down drastically. So instead of eighth-period lag, I can always impose restrictions on how many lags to be used in my estimation. **(Video Ends: 30:45)**