

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

Lecture - 47
Dynamic Panel Data Model - Part X

(Video Starts: 00:19)

Now look at what is happening here $nL1$ is 0.62. Now theoretically it should be between 1.04 and 0.73 it is little lower than this why is it happening because my model in this model I have taken only 2 variables as exogenous variable only w and L sorry here w and k . But as you know I have $wL1$ $wL2$ then I have $kL1$ $kL2$ then I have y_{s1} y_{s2} all those variables I have not included and that is why this is coming like this in xt in Arellano Bonds model.

But one thing what we should understand from here is after estimating this model first thing is that what is your what is the number of instruments that you have used and how are they reported in this data output look at here I have 2 types of instruments one is called GMM type another one is called standard. GMM type means those instruments which are coming from Arellano and Bonds logic that means second year lag and further for the dependent variable.


So, that means $L2$ this is L slash dot means second year onwards lag of the employment what does it mean it means y_{it-2} y_{it-3} y_{it-4} wherever it is available that is what exactly Arellano and Bond set okay that is what exactly Arellano on Bond said and if you do. So, that means if you do. So, then my total number of instrument how many instrument I will be getting that we should understand right total number of instruments are 40.

So, that means out of this how many instruments you are getting for these exogenous variable $D \cdot w$ $L \cdot D \cdot w$ $L2 \cdot D \cdot w$ $D \cdot k$ $L \cdot D \cdot k$ $L2 \cdot D \cdot k$ $D \cdot year$ $a \cdot t$ $D \cdot year$ 81 82 83 84 and here. So, if we count 1 2 3 4 5 6 7 8 9 10 11 12. So, out of this 40 instruments that is reported here look here 40 instruments are used in this estimation procedure suggested by Arellano Bond 1991 when I have only 2 exogenous variable w and k right w and k and these years.

So, if 12 instruments are coming from this side that means I must get 28 instruments from GMM Style you understand what I am saying I will write it down this we should understand total number of instruments is total number of IVs equals to 40 out of that GMM Style sorry standard IV standard IV equals to 12. That means my GMM Style IV must be $48 - 12$ equals to 28 right 28 and that whether it is happening or not we have to check it.

So, once again if you recall the instrument Matrix in Arellano and Bonds approach was this yeah. So, y_{i1} y_{i2} and y_{i3} to y_{i1} . So, this is how this is the instrument Matrix. So, first instrument we got when t equals to 3. So, in third period I got only one instrument when it came to t equals to 4 how many instruments I got y_{i2} and y_{i3} that means second and third year lag this y_{i2} .

(Refer Slide Time: 06:42)



$t=3 \Rightarrow 1979 \rightarrow$	\cdot	\cdot	\cdot	$= 2$
$t=4 \Rightarrow 1980 \rightarrow$	\cdot	\cdot	\cdot	$= 3$
$t=5 \Rightarrow 1981 \rightarrow$	\cdot	\cdot	\cdot	$= 4$

$t=7 \Rightarrow 1982 \rightarrow$	\cdot	\cdot	\cdot	\cdot	$= 5$
$t=8 \Rightarrow 1983 \rightarrow$	\cdot	\cdot	\cdot	\cdot	$= 6$
$t=9 \Rightarrow 1984 \rightarrow$	\cdot	\cdot	\cdot	\cdot	$= 7$

Total number of instruments for GMM style
 $= 1 + 2 + 3 + 4 + 5 + 6 + 7$
 $= 28$



Please keep in mind this y_{i2} basically second year lag is coming from Arellano bones logic that y_{it-2} . and then he also mentioned that this y_{i1} that means additional lag if it is available Arellano Bond suggested to use that. While Hole Jackin et al he said only one instrument for each period y_{it-2} to improve the efficiency of the estimates what Arellano and Bond improved over Hole Jackin et al he said you use y_{it-2} absolutely fine but you check if additional lag is also available or not.

So, when I come to $t - t$ equals to 3 I have only one instrument because that is the available instrument available y_{it-2} when it is third period my instrument starts from third period. But when I come to t equals to 4 in the fourth period I have y_{it-2} additionally I have y_{it-3} also

third period lag is also available that is why I included that. So, that means my instrument number of instruments are like this.

It is t equals to three that means when in in this data set implies my data is 1978 okay 76, 77 gone instrument starts from third period 1978. And then number of instrument for this equals to 1 when t equals to 4 that means for 1979 number of instruments equals to 2 right t equals to 5 implies 1980 number of instrument equals to three t equals to 6 implies 1981 number of instruments equals to 4.

Similarly for t equals to 7 implies 1982 number of instruments equals to 5 t equals to 8 implies num 1983 number of instrument equals to 6 and for the final T equals to 9 1984 which is the last period of my data set this is equals to 7. So, if I add all those total number of instrument number of instrument s for GMM Style equals to then $1 + 2 + 3 + 4 + 5 + 6 + 7$ equals to $2 + 1 + 3 + 3, 6 + 14 + 5, 15 + 6, 21 + 7, 28$.

Now if you check that is the number of instruments what we are expecting GMM style instruments would be 28 for this kind of setup. Now in our previous in one of our previous discussions I think while discussing about Arellano Bonds instrument metrics I made a small mistake I said this is t equals to 2 actually the instrument starts from t equals to 3 no instrument for second period right.

This is quite simple third period only instrument starts and then he said wherever it is available you additional lag available we should include that. So, following this logic we can very well check whether my stator is actually reporting proper number of instruments or not this is 40 instruments out of that we are getting 12 from these exogenous variables and 28. So, L in the bracket 2 slash dot means second year and onwards lag of this employment variable is generating GMM type instruments number of which is 28.

Now you can understand for this many periods we are getting 40 instruments are out of that 28 are coming from this lag dependent variable 28 all right. So, this is basically Arellano and Bonds method Arellano Bonds 1991 method that we use to estimate the dynamic panel data model which is called a difference GMM actually because here this model what we are discussing the Arellano bonds 1991 this is called difference GMM.

This is called difference GMM why this is called difference GMM because we are first because we are foreign first difference of the original equation what is my original equation original equation is this. This is my original equation we are taking the first difference. So, that is why difference GMM because we are taking first difference of the original equation and then using legs of the dependent variable I am simply writing y_{it} as $I V$ for the difference equation.

So, this is that is why this is called difference GMM. So, idea is you would take the first difference of the model and use lag second year lag of your y_{it} starting from second year and higher as instruments. So, when you take y_{it-2} as instruments for your endogenous variable then that model is called difference GMM right. So, since we are taking one year lag over here in this approach only 1 lag of if you look at in my in my model what I am say I have specified only x_{t-1} .

So, that means I am taking Only One lag of my employment variable that is why x_{t-1} uses lag 2 and higher as instruments okay and then differences of the exogenous variables are also used as instruments lag 2 of in lag 2 and higher for n and differences of the exogenous variables are used and totally I am using 40 instruments here.

(Refer Slide Time: 15:20)



$$U = 28$$

Limitations of Arellano and Bond (1991) model:

Here we use lag of levels as IVs for variables which are first differenced. These IVs are poor instruments for diff. variable as noted by Arellano & Bover (1995) and Blundell & Bond (1998)



Now what is the limitation of this Arellano Bonds approach? Let us now try to understand limitation limitations of Arellano and Bond 1991 model what are the limitation what is the major limitation of this model? So, major limitation is that we are using we are using if you go back to the equation we are using y_{it-2} that means lag of level as instruments for variable which are actually differenced.

Lag of level lag levels are used as instruments for the first difference and later on Arellano and Bond later on Arellano and Bover in 1995 and Blundell and Bond 1998 they realized. So, here we use here we use lag of levels lag of level okay as IV for variables which are different which are first referenced these IVs are poor instrument that means when you use lag up level as IVs instruments for the first different variables these instruments are poor instruments for different variable as noted by Arellano and Bover and Bover 1995 and Blundell and bond foreign 1998.

So, the idea is in 1991's approach of Arellano one Bond they used lag of level as the instruments for the variable which are first difference. Now when you do so, lag of level acts rather a poor instrument for the variable which are first differenced and this is even the problem is very severe when your variables are actually variables are following random walk in nature that means if you look at my difference this is called first order moving average.

So, in this type of system when the error structure is moving average of order 1 in this type of system using $y_{i,t-2}$ that means lag of level as instrument for variables which are la which are actually first differenced they create some kind of problem these are poor instrument and what I am saying the severity of the problem is even more when you have your error structure is this. So, you cannot really use lag of level as instruments for the variables which are differenced.

Then what is the solution the solution actually then given by these 2 models Airline on bovar Blundell and bond they suggested system GMM approach.

(Refer Slide Time: 21:06)

NPTEL

- was diff of the level var as IV for the level eqn...

xtddfsys : LD.n
 $n = y_{it}$
 $D.n = (y_{it} - y_{i,t-1})$
 $LD.n = (y_{i,t-1} - y_{i,t-2})$
 IV

78 79 80 81 82 83 84
 $1 + 1 + 1 + 1 + 1 + 1 + 1 = 7$

IV = 47



So, in system GMM approach I will write the equation once again y_{it} equals to Δy_{it-1} as $\beta x_{it} + a_i$ Plus b_i this equation is called level equation and $y_{it-1} - y_{it-2}$ equals to $\Delta y_{it-1} - y_{it-2} + \beta x_{it} - x_{it-1} + v_{it} - v_{it-1}$ this equation is called FD first difference equation. So, the system GMM approach they say that you consider both the equation and consider them as a system of equation.

So, earlier we were discussing only about the first difference equation and we are saying lag of level as instrument for this $y_{it-1} - y_{it-2}$ we are using. Now when you are using this level equation also then Arellano Bover 1995 and Blundell and Bond 1998 or they suggested for the level equation you use Δy_{it} that means difference of your dependent variable as instrument.

So, for the first difference equation instruments where as it is additionally I am considering additional instruments for my y_{it-1} that means for my level equation y_{it-1} I am considering instrument as Δy_{it} . So, while level of the dependent variable lag of level lag of level is considered as instruments for the difference equation we are considering difference of the level difference of the level variable as instrument for the level equation right that is the idea given by the Arellano Bond, Arellano Bover 1995 and Blundell and Bond 1998 right.

So, if you do so, consider them as a system of equation and then use this is the first step and then use difference of the dependent variable as instrument as IV for the level equation. So, what we have to then we have to use the command called `xtdpdsys` x t means panel d p d means dynamic panel data and sys system that is why the command is `xt dpd sys`. This is the command if we want to use a system GMM method.

So, we will be using `xtdpdsys` this is the command and then my dependent variable then again lag of 0 by 2 dot dot w k same type and then again `ar(19, 82, 1984)` and one more year and then `vce(robust)` oh sorry we have to write year. So, this data is easy to get if you have a license data version just you have to use the command `vs` sorry wave use `AV` data early learner bonds original data. So, this is my system GMM method.

So, in system GMM if you estimate the knot is happening look at my coefficient is 0.82, 0.82 significantly improved even when I have used only 2 exogenous variable is it lying within the bound let me see this is one point zero four this is point seven three and how much I got 0.82.

So, it is lying well within the interval when you use a system GMM method. So, that means when I use difference GMM still my coefficient was 0.62 lower than the lower limit.

But of course this model and the earlier model is not directly comparable because I have used y_t and all these things probably you can check how much this value this is this you can take as an assignment that when I use same set of variables and estimate a difference GMM whether my estimates are lying within the OLS and Fe interval or not. But here when I am using when I am using system GMM fantastic Improvement on my lag dependent variable it is $L y_{t-1}$ it is.

Now 0.12, 0.82 exactly between the interval of 1.04 and 0.72 but what is the number of instruments. Now the number of instrument is 47. So, that means number of instruments it has increased by seven. Why it is increasing where from where are this seven additional instruments coming if you look at then earlier I was getting instrument only for the difference equation and lag of level that means see L in the bracket 2 slash dot lag of the level variable n was used as instrument.

When I go for system GMM additionally I am considering instruments for the level equation what is the level equations instrument I am using $L \Delta y_{t-1}$ so, if your n is y_{it} what is $L \Delta y_{t-1}$ I will use $L \Delta y_{t-1}$. So, in system GMM approach $L \Delta y_{t-1}$. So, if n is y_{it} that means $L \Delta y_{t-1}$ is $y_{it} - y_{it-1}$ and $L \Delta y_{t-2}$ is then lag upon period lag so, $y_{it-1} - y_{it-2}$. So, this is used as IV for the level equation.

So, that means additional 1 1 instruments in the form of this difference $L \Delta y_{t-1}$ you will get additionally earlier we are getting 40. Now you check whether additional seven instruments are coming from this structure or not so. Now what will happen you will get that means this y_{it-2} and y_{it-1} this $L \Delta y_{t-1}$ this is this is your, now instruments. So, again for the first period that means third period it is one for third period means you will have 76, 78, 77, 78 this is 78 is 1 then + 2 for 79 + 3 sorry these are all 1 1 1 1.

So, from third period onwards 78 79 80 81 82 83 and 84, so, 1 1 1 1 1 1 1 equals to 7. So, that means when you use system GMM apart from those 40 instruments which we already got additional seven instruments we are getting because we are using $L \Delta y_{t-1}$ lag of the difference of your dependent variable as instruments for the level equation that is why additional seven instruments and total number of IV is now equals to 47.

This is the system GMM model number of instruments additional 7 we are getting from this and if you do. So, in system GMM model then we saw that the coefficients of the lag dependent variable is lying exactly within the theoretically defined bound given by AP and OLS within the interval of this which you are not able to get in Anderson and has used standard IV approach.

Now this system GMM you got the first property is that the estimate should lie between the interval that is satisfied. But if you recall we said that the moment you introduce a lagged variable in the model that will generate additional problem. What is that problem you will have autocorrelation that autocorrelation is of order 1 must be there but there should not be higher order autocorrelation.

That means after estimating this dynamic panel data model we must check whether ar1 is there or not if it is not there that means your model is of no use why because ar1 is not there means it is actually negating the importance of the dynamism in the system that means the moment you introduce $y_{i,t-1}$ as additional variable it should produce ar1 in the model. If it is not there that means the there is no need of going for a dynamic panel data model itself.

Because the structure of the model says ar1 must be there auto correlation of order one must be there but there should not be higher order auto correlation so, whether though this models after estimation as a post estimation checkup. Now we should verify that ar1 should be there and ar2 and higher should not be there and that will discuss in our next class. So, then basically in today's discussion we first once again discussed about the Fe and OLS estimates and we showed the upward and downward bias.

And then we first discussed about Anderson and Hsiao approach and then we showed that in Anderson and Hsiao approached the instruments or estimates are not lying within the limit but if you go for Arellano and Bond's model 1991 then it is improved it is approaching towards the limit but Arellano Bond's 1991 model basically the difference GMM it has a problem because it uses lag of level as instrument when the equation is in difference form.

Severity of the problem arises when your error terms actually follows ma1 process that is why Arellano and Bond 1995 and Blundell and Bond 1998 they say that apart from using lag of

level what you can do lag of difference also as an instrument for the level equation that means we are now considering the level equation and the difference equation as a systemic equation. In difference GMM we are considering only the difference equation and using instrument for the differenced variable.

Now we are using level equation as well as the difference equation as a system and then we are using the lag of level for the difference equation instrument and different lag of difference that means $L D \cdot n$ as instrument for the level equation and thereby improving the quality of the estimates drastically. With this we are closing our discussion today and next class we will be discussing those post estimation checkups particularly the autocorrelation.

And we have also discussed that there should not be any uh all my over identifying restrictions should be satisfied. You see the number of instruments are 47 here for t earlier whether all those instruments are actually satisfying the over identifying restrictions or not that we will discuss in our next class, thank you.

(Video Ends: 38:06)