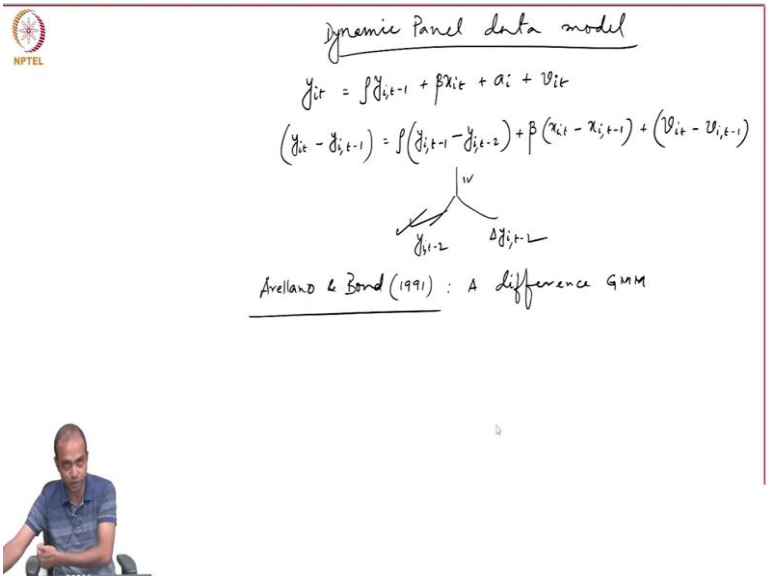
Applied Econometrics
Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology – Madras

Lecture - 48
Dynamic Panel Data Model - Part XI

(Refer Slide Time: 00:19)



So welcome once again to our discussion on Dynamic panel data model and in our last couple of sessions. We were basically discussing the empirical estimation of several Dynamic panel data models taking the data from original Arellano and Bonds data which was in the context of UK. And the context was the dependent variable was forms employment that means we were basically trying to understand the determinants of firm employment.

Considering the fact that employment depends on its previous AR's employment because of the reasons we have already explained that the cost of hiring and firing. So, that means State period employment bought by the ith form depends on the t - one is employment by the same ith form. So, mathematically if we write then the model will look like y i t equals to rho y i t - 1 + beta x i t + the unobserved individual specific time constraint effect + the standard idiosyncratic error.

And then we started with Anderson and Hsaio approach that was basically y i t - y i t - 1 taking the first difference equals to rho y i t - 1 - y i t - 2 + beta x i t - x i t - 1 + v i t - v i t - 1 and then we say that basically in Anderson and Hsaio approach we need to use IV for this variable y i t

- 1 and there are 2 IVs that we are discussing either lag of level y i t - 2 or Delta y i t - 2. So, if we take y i t - 2 that is called the difference GMM.

And when we consider Delta y i 2 that is called system GMM we have already explained in our previous sessions what exactly we mean by difference and system GMM. So, quickly what you will do will estimate these 2 model once again why you are doing this Arellano Bond or difference and system GMM. So, Arellano and Bonds model basically we are discussing Arellano and Bond 1991 model.

We are discussing which is basically a difference GMM model yeah different GMM based on this idea right based on this idea of Anderson and Hsaio will difference the equation and then you use lag of level for the instrument for the difference equation that is why it is called difference the GMM and if we take a lag of the difference that means Delta y i 2 and consider both the level equation and the difference equation as a system of equation that is called a system GMM equation.

So, one advantage is with this difference GMM when we are taking lag of level we will get little more observation because Delta y i t means at least your observation will start from fourth period whereas here in y i t - 2 in your instrument Matrix your first observation for instrument you will get for the third period. So, one period also matters when your panel size is really small.

So, that was the context and I will once again estimate these Arellano Bonds model before we go to our next model that we are going to discuss today.

(Video Starts: 05:00)

So as you know that when you use Arellano Bonds original data of this UK forms 140 forms data we have and to access the data we need to use where we use AV data and then once you put that the command is xtabond. So, xtabond then our dependent variable which is in employment and then we have after this what we need to do we need to take lag of several variables and we are taking a wage and capital.

So, these 2 are my exogenous variable we are taking dot. So, we consider sorry we consider only lag of the dependent variable that means lag of employment is the only endogenous variable and wage and capital they both are exogenous variables in the setup and then we are

taking AR 1979 to 1984 right and then we are taking ER also as a separate variable of explanation and then we are putting vcerobust command.

So, this is basically Arellano of Bonds 1991 the difference GMM model that we are estimating. So, this command basically says that the moment we put xtabond automatically it will consider that I am going to cons I am going to treat only in as my endogenous variable no need of putting its lag variable. So, it will automatically take one lag right if you want more lag then only we need to specify. So, we are considering our model as y i t equals to y i t - 1.

So, the inbuilt command of x t upon will consider please keep in mind in as the only indigenous variable and it will consider only one AR lag. That is all then after that we are putting what does this L 0 by 2 means that means I am considering w and k as exogenous variable and I am considering 2 ARs lags of both the very variables that means wL1 and wL2 kL1 kL2 where L indicates lag of 1.

And then these are the my year dummies this is my AR and this is I am taking the robust standard error sorry I have to put here over here I have missed here. So, I will put the command once again. So, this is here 1984 and if you put that then you will see the Arellano bond estimates. So, it is coming 0.62 in the difference GMM model and it should lie between 0.73 to 1.01 but as you know in this model I have included only w and k I am not including the ws.

And then I asked to take it as an assignment to see after inclusion of all the variables whether this nL1 coefficient is lying within the interval or not. Because if you recall previously when we first estimated f e and OLS we have included w s w s L 1 and W sorry y s L 2 that means industrial output we took as a proxy for aggregate demand right in that particular AR. So, if you include those variables whether your coefficient is changing or not.

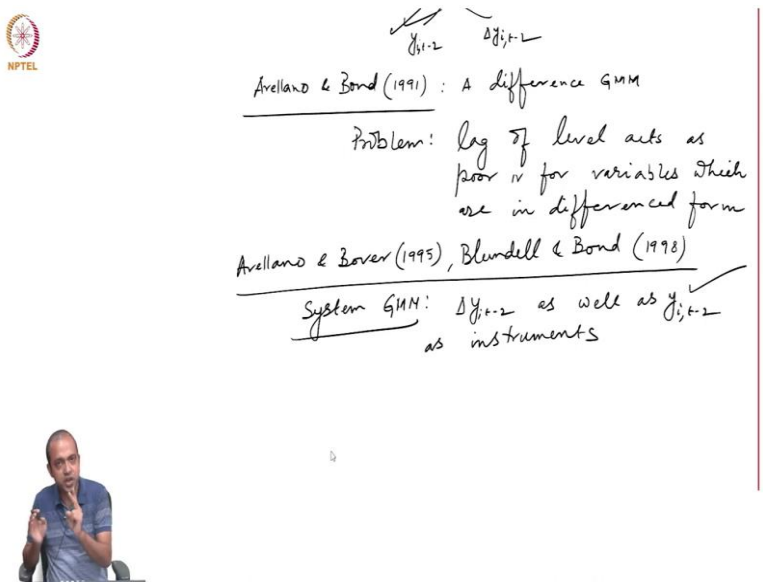
So, this is basically in short Arellano bonds 1991 model we have estimated and we have 40 instruments over here and we have also explained that out of this 40 instruments 12 we are getting from the early learner bonds 12 we are getting from the standard instruments from the exogenous variable and if you count want it 1 2 3 4 5 6 7 8 9 10 11and 12. So, 12 standard type instruments and 28 instruments are coming from that endogenous variable from the third periods onwards.

If you consider if you relook once again the instrument Matrix in Arellano Bonds approach 1991 approach you see that in the third period I have one instrument fourth period 2 fifth period three likewise it will go on and if you wired all these instrument you will get 28 instruments totally. So, 28 instruments are GMM type and 12 instruments are standard type total 40 instruments that we were getting in Arellano Bond difference GMM.

But we explained that in difference GMM it has a shortcoming because we are using lag of level y i t - 2 as an instrument for variables which are in different form. So, lag of level are of poor instruments type for differenced variable right and the severity problem severity of this problem increases when the variables are actually variables they follow a random work. To overcome that problem Arellano and Bover we said that Arellano and Bower 1995 and Blundell and Bond 1998 they basically suggested system GMM.

(Video Ends: 11:37)

(Refer Slide Time: 11:38)



So, this is Arellano Bond difference GMM here the problem is as we have already mentioned that lag of level what is the problem lag of level lag of level acts as poor proxy for poor IV for variables which are in different form. So, to overcome this Arellano and Bower and Bower all right 1995 and Blundell and Bond 1998 they suggested system GMM. So, what is system GMM basically lag of difference variable Delta y i t - 2 as instrument.

So, instead of taking lag of level I am taking lag of sorry the difference variable right. So, this in this approach I am considering level equation as well as different equation as a system of equation lag of level acts as instrument for the differential equation and lag of difference act as

my instrument for the level equation right. So, once again with the same variable if we use system GMM then to use to estimate system GMM the command or to use is xtdpdsys.

Xtdpdsys for dynamic panel data model and CIS for system GM system GMM. So, totally xtdpdsys and then we have the same command the whatever we have used earlier. So, that means I can put this part of my command here. So, xtdppsys is then my dependent variable then twice lag of w and k and then here again I should mention here here and then if you put then you will get this is the estimates of system GMM.

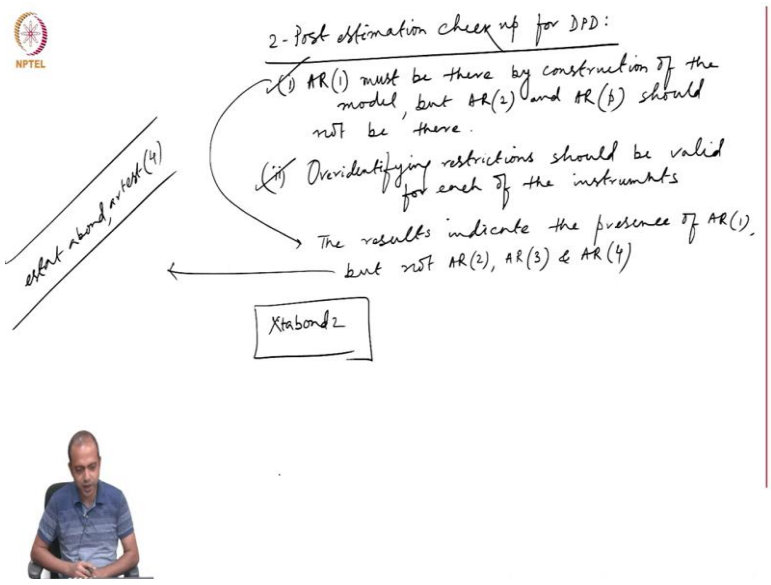
And this estimates of system GMM as you can see this coefficient is 0.82 which is fantastic because it is lying within the FE and OLS estimates as we have discussed previously but we have also mentioned that there is a warning or cost for this system GMM because it puts additional restrictions additional restriction about the initial condition of the data generating process for y i t why this is.

So, earlier we were putting del y i t - 2 as instrument that means the orthogonality condition is that expectation of this as an instrument when you are saying this as an instrument that means expectation of y i t - 2 and the covert covariance or the error term is zero. So, this is basically the moment you put this orthogonality condition it is it is something like I am specifying a specific data generating process for y i t please keep in mind.

So, this when I am putting only one in one such instrument con which is coming from the y i t - 2 that means I am putting only one type of restriction about the data initial condition about the data generating process of y i t now in system GMM what is happening I am considering y i t - 2 as well as Delta y i t - 2. So, I am taking y i t - 2 as well as instrument. So, earlier it was only y i t – 2.

Now it is this + this. So, orthogonality conditions about this and this that means I am putting additional restrictions about the initial process of the data generating initial condition about the data generating process of y i t that is all. So, putting additional restrictions is comes with some command of cost and that is the cost of system GMM we have mentioned in our previous class and we indeed our class saying that once we estimate dynamic panel data model basically we need to go for 2 type of post estimation checkup.

(Refer Slide Time: 18:32)



2 post estimation checkup for dpt dynamic panel data model what are those first one is there should be autocorrelation of order 1 because the structure of the dynamic panel data model itself says since we have introduced lag of the dependent variable in the model AR1 that means autocorrelation of degree 1 is inevitable we must get that if we do not get it that means the presence of y i t - 1 is unnecessary.

So, we are undermining the dynamism in the system. So, ar1 must be there must be there by construction of the model but AR2 and higher should not be there and higher that means let us say ARp right. So, this is basically this is basically the first post estimation checkup regarding the presence of higher ordered autocorrelation is there or not that we have to test and second one is since there are too many instruments used in the estimation of a dynamic panel at a model.

We must check what identifying restrictions what identifying restrictions should be should be valid for each of the instruments. So, these are the 2 post estimation checkup we will start with the first one whether auto correlation of higher order is maintained or not. So, the command for this once we estimate these we can use estat command estat estateabond and then we have to put a comma AR test here I am putting up to fourth order autocorrelation okay AR test 4.

Now look at what is happening look at the Z coefficient - 4.64 and the probability is 0.0000 the P value. So, that means first order Auto correlation is there but none of the higher order correlation coefficients are significant here that is what this particular day is all Source the my

null hypothesis is no autocorrelation. So, I can reject my null hypothesis only for AR1 which indicates the presence of first order autocorrelation.

But null hypothesis of no autocorrelation cannot be rejected for second third and fourth other autocorrelation. So, that means what we can say that the result indicate the results indicate the presence of AR1 but not here to AR3 and AR4 okay that is the result. Now we should check for the second first estimation checkup which is whether over identifying restrictions are valid for each of these instruments.

Now to test that condition post estimation to go for that second post estimation checkup we need to estimate the dynamic panel data model using another command xtabond 2 developed by David Rudman. So, what I am saying we can estimate the same difference and system GMM models by using xtabond 2 xtabond2 and what was the command used for this autocorrelation I will mention that also estatabond, AR test 4 we were taking up to 4th order this was the command.