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

Lecture - 46
Dynamic Panel Data Model - Part IX

Once again welcome to our discussion on dynamic panel data models. And if you recall in our previous session we were discussing about the empirical estimation strategies for the dynamic panel contexts. And we were taking an example where the forms are actually hiring the labour from the UK context this is 140 firms data that we were using. And then we assumed that employment the firms employ a particular ith forms employment in T8 period basically it depends on its own lag value.

That means Tth period employment is a function of T - 1 th periods of employment also and that is how dynamism was there in the picture why this was so? Because we assumed that higher sharing and firing is actually costly activity. If we want to hire new levelers then the form of the firms they have to advertise they have to conduct interview these are additional costs over and above the extra salary or wage that the firms are going to pay.

That is why firms mode or less they determine a Tth a particular periods employment based on what is the employment they already have from the previous period right that is how dynamism was introduced in the system. And we also assumed that this particular employment is a function of not only its previous value but also what is the prevailing wage W what was the capital that the firm is having then what is the aggregate industrial output.

Why you have taken industrial output because industrial output is basically acts as a proxy for the demand for industrial sectors product. So, that is how we formulated our model a dynamic panel data model for estimating determinants of forms employment. And then they are also or two factors in the model one is the idiosyncratic error and the other one is the unobserved form specific Factor.

And if we ignore the unobserved form specific time constant factor and simply estimate the model using OLS ignoring the dynamism that is present in the model then we discussed that there would be a dynamic panel bias. If we estimate the model using OLS then we will get

overestimated coefficients and if we estimate the models using fixed effect transformation then it will lead to downward bias of the estimates.

And then we prove that then the true estimates must lie between the upper and lower limit that means the OLS and the fixed effect transformation that is what we discussed. So, before we start our today's discussion that means the new approaches that will overcome these issues of overestimation and underestimation if we follow OLS and Fe we will quickly use the same data try the model once again using OLS and Fe. And we will note down the estimates derived from these two particular models.

(Video Starts: 04:19)

So, we are again using this Arellano Bonds original data set and if you have this data version this data licensed version if you use wave use a b data then as I said you will get the data imported in the software and we can easily request the model. Now to estimate the OLS model in this system what we have to do we have to use regress and then our dependent variable is n then in L1 that means lag up employment 1, nL2.

Sorry we have to give one space nL2 we can actually use the variables from here and then we have w, w is basically the wage rate then we have wl1 then we have capital then capital lag of 1 kL1 then we have kL2 also this is scale two second period lag also we are taking. Then we are taking y s that means industrial output and then lag of first and second lags of industrial outputs and then we will take uh your dummies yeah this is your dummy.

So, this is basically estimating OLS even when there is dynamism in the system ignoring the dynamism in the system we are simply decreasing the model. And then you see that the coefficient of the lagged employment is 1.04 which is upward bias. Why it is opposed bias because in this model we have yit - 1 is actually positively correlated with the unobserved effect AI.

We discussed in our previous class if a shock is happen a particular employment sock is given let us say negative employment shock is given for a form in a particular yea.r Then in the next period employment as well as this unobserved effect both of them will go down and this same Trend in the variable will lead to a positive correlation between yit - 1 and AI than observe effect and that is why that positive correlation will lead to overestimation if we apply OLS.

And then when we go for fixed effect transformation then what happen y i t - y i bar that is the dependent variable then your independent variable our variable of concern is Delta y i t - 1 - y i t - 1 bar and that will be correlated with - v i bar. So, v y i t - 1 we are talking about y i t - 1 and v i bar and the negative sign of v i bar leads to the negative correlation and as a result of which the fixed effect estimation of this dynamic model will lead to underestimation.

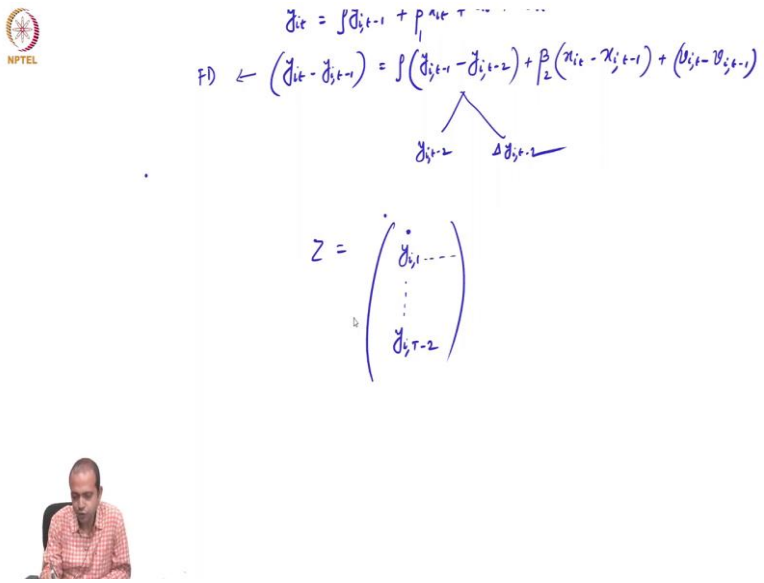
And how do you estimate that we discuss three approaches to estimate the fixed effect model first one was LSDV just to take the AI component out from the error term and we said that we will we will simply use the individual specific time constant error time constant factor and we will use a form specific dummy in place of AI so, that bring the AI component out of the error term and because the AI was earlier creating problem right.

So, these are the three ways this is the first approach and then we have also discussed about removing the influence of AI from each of these variables and then running the final regression on the residuals. So, that is one approach and second is one we can straight away use actually the fixed effect transformation right fixed effect transformation. And how do you use fixed effect transformation it is basically x t reg in place of in place of regress what I will do x t a reg.

And then same variable we are going to use in and then nL1 that we are using nL2 second lag and then w then w lag one capital of lag one capital of lag 2 and then ws1 sorry w y s before that we have to take ys and then ysL1 and then we have yr star and then we will put Fe. And if we do so, then we see the coefficient is 0.73. So, these two values we have to actually note it down if we apply OLS in this dynamic model we are getting 1.04 which is upward biased.

If we apply OLS then we are getting downward bias of the estimate which is 0.73. So, two estimates then must lie between 1.03, 1.04. So, I will note it down these two values.

(Refer Slide Time: 11:09)



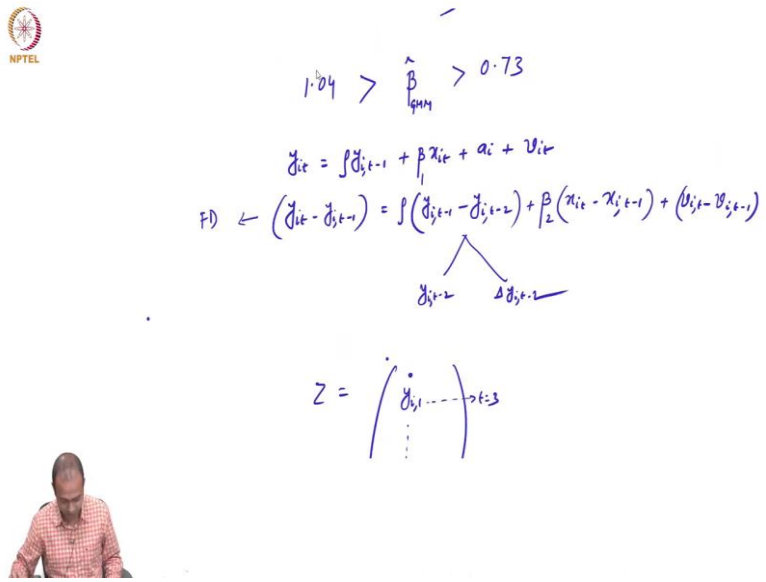
So, one point OLS is giving 1.04 and fixed effect transformation is giving 0.73. So, these two values we have to keep in mind. So, then my True Value beta hat beta hat in case of GMM must be lower than. So, lower than 1.04 but greater than 0.73 so, it should be it should be between these two values all right. So, now what we will do we will. Now go back and when fixed effect transformation and well s is not applicable they notice the solution.

Our solution is then to go for the Anderson and Hsaio approach the first model of dynamic panel data right if you recall the first model what we discussed was Anderson and Hsaio where he said that second leg of the dependent variable that means y i t - 2 should be used as an instrument after first referencing the model. So, for your better understanding I will write the model once again y i t equals to rho y i t - 1 plus beta 1.

Let us say x i t plus a i Plus v i t and when we take the first reference to remove the unobserved effect this will become y i t - y i t - 1 equals to rho y it - 1 - y i t - 2 plus beta 1 x i t - x i t - 1 plus v i t - v it - 1.. So, this is the first difference FD this model is FD and Anderson and Hsaio he suggested that either you use y i t - 2 or Delta y i t - 2 as instrument . So, lag of the level or lag of the difference that we have to use say instrument for this particular variable.

So, first we will use y i t - 2 that as an instrument in Anderson and Hsaio approach. So, what happens then the instrument Matrix that means if you recall what we say that instrument Matrix once again fast observation would be lost then y i 1 and Dot y i t - 2 this is the instrument Matrix and first information is available when t equals to 3 right. So, this is the structure of the instrument Matrix in Anderson and Hsaio approach.

(Refer Slide Time: 15:31)



So, then what we will do to estimate Anderson and Hsaio model here what we will do we will put a simple iv reg command just like we have discussed the instrumental variable estimation previously simply I will a command then our dependent variable is d dot n what is d d dot n is d for difference. So, that means y i t - y i t - 1. So, if you go back and check d dot means this particular variable this is my dependent variable that is how I am specifying.

And now after this what we have to do we have to say that d dot d dot nL1 that means what is nl1 nL1 is y i t - y i t - 1 and if you take a difference of that that means y i t - 1 - y i t - 2. So, d dot nL1 actually equals to n l two that is n l 2 is the instrument all right. So, y i t - 1 nL1 is y i t – 1 and we are taking the difference that means y i t - 1 - y i t - two this is my dependent variable if you go back dnl1 is this d nL1 is this nL1 is this. So, nL1 that means y i t - 1. And then when you take difference where d means the difference of that right y i t - 1 - y i t - 2 that is the difference and then what we are using instrument as this y i t - 2.

So, that is nL2 y i t - 2 that is what we are doing this is nL2. So, we have used this instrument we are assuming that this is the only endogenous variable and other variables in my model over exogenous. So, all my other variables are actually exogenous we in this case I have denoted only one exogenous variable x i t but in our model we have apart from employment and its lag first leg.

We have second leg then we have wage we have Capital we have aggregate output ys lag of first lag of ys a second lag of ys everything we are considering as exogenous variable. So, only

endogeneity that we are assuming so, we have we will write over here that in this approach in this Anderson and Hsaio approach indigenous variable what we are considering is y i t - 1 that is the indigenous variable that we are considering.

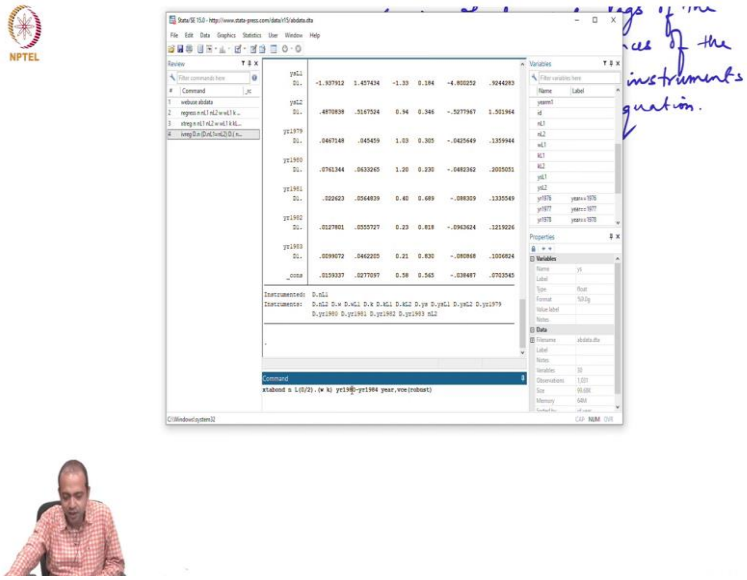
That is why we are using instrument for this but we are not using any specific instrument for that but as you know that exogenous variable or predetermined variables they themselves act as instrument right. So, then what we will do this is our instrument and then we will be using after the d dot all my variable. So, nL2 I have nL2 then I have w I have w then wL1 then I have capital then capital of lag one then lag 2.

And then we have where Ys then y s of lag one y a sub lag 2. So, these are the variables we have considered and then we have year 79 then 80, 81, 82 and 83. This many years we are taking. Now you see when you are including y here we are including from 1979 even though the data starts from 1976. Why this is so, because when you take the first difference in your Z Matrix if you go back your observations is available only from t equals to 3.

So, 76, 78 they are gone. So, my instrument my observation in the instrument Matrix observations will appear only from t equals to 3 that is the reason we are considering the year from 1979 which is the third year of this sample period 76, 77 and 77 and 78 they are gone and we are taking the Year from the third. So, after this what we will do we will put this and this is my estimates.

So, now our variable of interest is nL1 that means 2.30 right. This is 2.30 that is what we were interested in. Now this 2.30 is it the question is then is this coefficient lies within the theoretically defined bound or not right.

(Refer Slide Time: 22:54)



We will go back and see what is our limit this is the upper limit our upper limit is 1.04 but the variable what we are getting is 2.30. So, that means what we can say that this 2.30 the AH estimate is two point what is the value 2.30. So, 2.30 which is greater than 1.04 the wireless estimate so, which is much higher than the theoretically defined upper bound. So, that means even though in Anderson and Hsaio approach we have gained some amount of consistency which was not there while applying OLS.

The estimates lie beyond the theoretically defined limit and if that is the case if the estimates go beyond this even if you gain little bit of consistency we cannot take this estimates because we know they are not reliable they are not lying within the limit. So, why this is happening probably because in Anderson and Hsaio approach we are using only second lag of level as instrument, so, only second year lag.

So, probably we are not using all the potential movement conditions that means all the information available in the system only one orthogonality condition y i t - 2 and expectation of y i t - 2 and that v i star is 0 that is the only one orthogonality condition using only one moment condition is not probably enough and that is why the Anderson and Hsaio instrumental variable approach which is considered as the first model of the dynamic panel data model is actually not reliable.

We cannot take these estimates since it is much higher than the theoretically defined upper limit right. So, if that is the case then when we do not take these Anderson and Hsaio model what is the next solution next solution if you recall was coming from Arellano and Bond 1991

which is of course the original idea did not come from Arellano and Bond rather it came from Hole Jackin et al.

What they said that since in Anderson and Hsaio approach you are losing information if you want to use more lags in the system if you want to use more relax that means your Z Matrix become this. This is y i one y i t - 2 and then if you use third year lag also then this will become two observations would be lost then y i 1 I do not y i t - 3 this is the Matrix. So, that means this is t equals to 2 this is T equals to 3 right this is Anderson and Hsaio approach.

And then Hole and Jackin et al they said that you use only one instrument for each period like the way Anderson and Hasiou initially define y i t - 2 but you replace all the missing values with zero even in that case also you will have only one year lag y i t - 2 for each period Then Arellano and Bond said for this lagged variable this endogenous variable use y i t - 2 and above. So, second year lag and higher you use for this endogenous variable for other predetermined and exogenous variables you use whatever lag is available.

So, lags of the dependent variable and then you use first differences of the exogenous variable as instrument. So, that means once again if we go back then what will do in this equation what you will do y i t - 2 will be used at this instrument for this variable. And this variables x i t - x i t - 1 like the way we have this first difference of wage, capital aggregate, output everything we have this difference.

And as a result of which all those variables which has been all those variables which are actually coming from the first difference they will be used as additional instruments that is the idea Arellano and bond said. So, using that so, that means Arellano and Bonds idea is using the lag of second year and above for this y variable and first differences of other variables all of them will constitute the Z Matrix all right.

So, Adeline unborn basically 1991 which uses which uses lags of the dependent variable that means lacks of the dependent variables means y i t that is the dependent variable and first differences of the exogenous variable of the exogenous variable first differences of the exogenous variable are used as instrument yeah for the first difference equation . So, once again I go back this is my first difference equation.

So, lag of the dependent variable I am using lag of the dependent variable y i t - 2 and first differences of all these variables x i t - x i t - 1 then I will have employment - employment t - 1 sorry anal d of nL1 and then I have first differences of capital first differences of lag of capital first differences of aggregate output all those things will also appear as instrument for this first difference equation.

What is the idea exogenous variable they themselves are their instruments. So, we do not have to put any specific instruments for exogenous variable that is the idea. So, based on this we will now estimate this Arellano bonds model and the command for these Arellano Bonds model is xtabond this is the command you have to write it down xtabond. So, xt means panel and a bond is coming from Arellano and Bond.

So, we will once again go back to our equation and then we will see the command is xtabond xtabond then simply my dependent variable I do not have to specifically put anything then L that means lag how many lags 2 lags how we are putting 2 lags 0 by 2 dot w then we can put all the other variables that we have taken w and then w L one then capital then capital then we have ys then ys1.

For Simplicity sake let me take only w and wL1 sorry w and k I am taking w and K for simplicity sake. And then we can take the year y i 1982 year 19 84. 1984 and then we have also taken your variable specifically then we will put vcerobust. So, let us first try to understand what we are writing in the command box xtabond is the command to estimate early learner bonds Model n is my dependent variable y i t then I am taking.

Since I have put x t a 1 command I do not have to specifically put in L1 again x t where one command itself will take care the fact that I am trying to estimate a dynamic panel data model. Only thing that we have to specify is how many lags you are taking for the exogenous variable and we are considering exogenous variables as only w and k for Simplicity you can include other variables also.

And then we are taking the year 1980 to 1984 we are taking additionally year as a variable then we are taking the robust standard error of the estimate that is given by vcerobust.