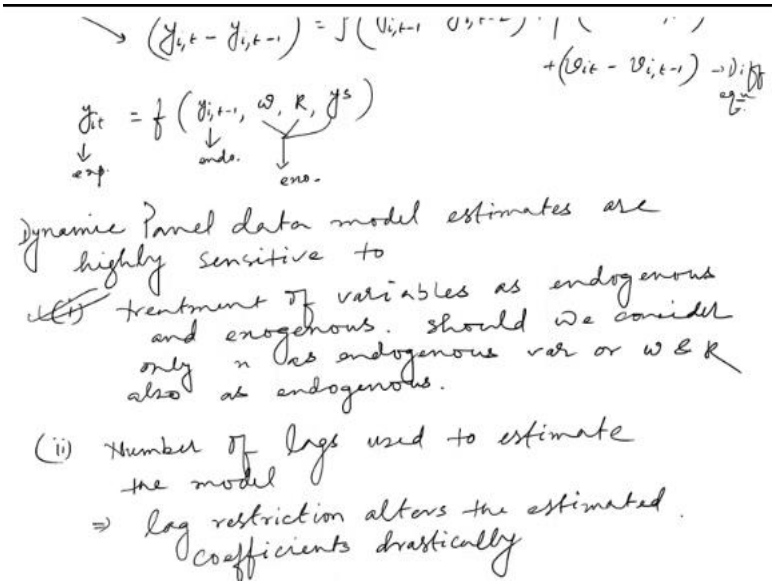
Applied Econometrics
Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology – Madras

Lecture – 53
Dynamic Panel Data Model – Part 16

(Refer Slide Time: 00:15)



Now we will see whether the number of lags can also change your estimates or not. We will first estimate the model without any restriction on the number of lag and with the same model, we will estimate by imposing a restriction on the number of lags to be used. **(Video Starts: 00:46)** So what I will do again I will be using *xtabond2 n L(1/2).n L(0/1).w L(0/2).(k ys) yr*,gmm(L.n) iv(L(0/1).w L(0/2). (k ys) yr*) noleveleq robust small*

Here, what I have specified then *xtabond2* is my dependent variable. I am taking two periods lag of employment as these are all after *n* everything I am specifying as independent variables, two-period lags of *n* and then one period lag of *w* wage, then two-period lags of *k* and *ys* and then year dummy. Then I am considering again *L*.n only as my GMM and within *iv* I have taken the first period lag of wage, two lags of *k ys* and *, I do not need level equation because it is a difference GMM and then I need robust standard error and small sample correction, with all these, now we will estimate the model.

Now without restriction what is the model I got 0.68, which is not lying within the interval. Then what we will do we will now impose restrictions on the lags to be used, how to do that?

The command is *xtabond2 n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, gmm(nL1, lag(2 5)) iv(w wL1 k kL1 kL2 ys ysL1 ysL2 yr*) noleveleq robust* here I am taking only lag 2 to 5, I am restricting the lag up to 5 instead of going up to 8, earlier as I showed it was taking up to 8. Now one question may come to your mind, look here I am specifying the lag for the endogenous variable but I am not specifying any lag for the *iv* style variables, this is because it is not required, if you recall that exogenous or predetermined variables, they themselves are considered their instruments, so there is no need, the moment you put the variables within the bracket of *iv* it is assumed that these variables are either exogenous strictly or predetermined and the appropriate lag.

If it is not lagged then first-order lag, if it is different, then the same variable will act as their instrument. So, while the endogenous variable, is the lag from second to higher whatever is available they are used as an instrument, for the exogenous and predetermined variable we do not have to specify any lag that means, imposition of lag-length is redundant within the *iv* bracket because they themselves are considered as their instrument, so lag length restriction is valid only for GMM.

Now if we estimate *xtabond2 n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, gmm(nL1, lag(2 5)) iv(w wL1 k kL1 kL2 ys ysL1 ysL2 yr*) noleveleq robust* we see that the coefficient *nL1* which was 0.68 earlier not lying within the theoretically determined bound, now you have achieved, by restricting the lag up to 5 now this is coming out to be 0.83 which is perfectly lying within the interval. **(Video Ends: 10:13)** So that means we have proved one thing that restriction of a number of lags.

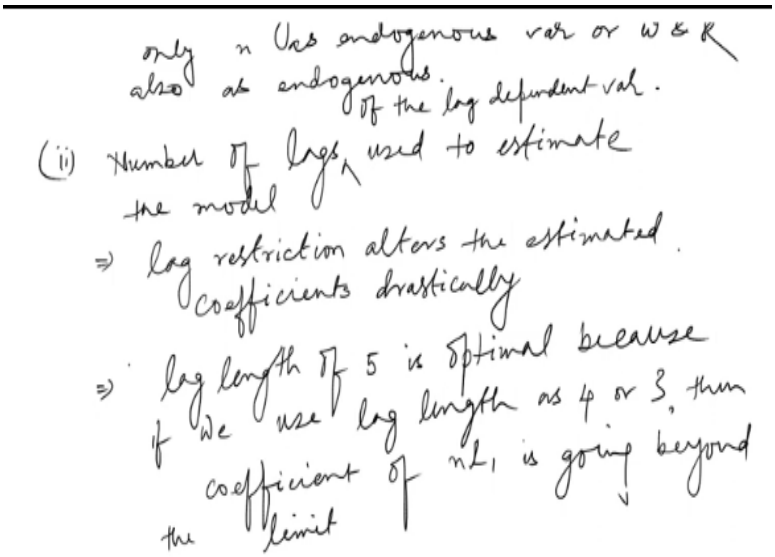
The lag restriction alters the estimated coefficient drastically, with no restriction. While the estimated coefficient was lying beyond the interval, with the restriction imposed it is lying within the interval. Then you might be thinking instead of restricting the lag why not reduce the lag to 4, what will happen? **(Video Starts: 11:27)** So let us see if we reduce the lag to 4 instead of 5 what will happen?

Here, instead of 2 and 5 let us see what happens if we put 4 *xtabond2 n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, gmm(nL1, lag(2 4)) iv(w wL1 k kL1 kL2 ys ysL1 ysL2 yr*) noleveleq robust*, so that means we are keeping everything else the same, we are just reducing the lag from 5 to 4. See, the coefficient of *nL1* again goes beyond the interval. What happens if we reduce

the lag to 3 *xtabond2 n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, gmm(nL1, lag(2 3)) iv(w wL1 k kL1 kL2 ys ysL1 ysL2 yr*) noleveleq robust small* So that means again it is going 1.10 and here also it is 1.10, so that means instead of 5 the moment we use 4 and 3, then it is not working well.

So, with this you can understand that in this particular situation maybe lag length 5 is optimum, we have to go with 5-year lag so that we are able to estimate. **(Video Ends: 13:08)**

(Refer Slide Time: 13:10)



Therefore, the number of lags of the lag-dependent variable, implies that a lag length of 5 is optimal because if we use lag length as 4 or 3 then coefficient of *nL1* is going beyond the limit. **(Video Starts: 14:39)** Now instead of estimating a difference GMM what will happen if we put a system GMM? A two-step system GMM *xtabond2 n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, gmm(nL1, lag(2 5)) iv(w wL1 k kL1 kL2 ys ysL1 ysL2 yr*) two robust small*, here we are restricting the lag, and again it is coming down to 0.63. This means that when you are restricting the lag to 5 while the difference GMM ensures the coefficient to lie within the interval, the system GMM could not.

Similarly, if we do not do all these things and do not put any restrictions then what is happening? Let us remove the restriction *xtabond2 n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, gmm(nL1) iv(w wL1 k kL1 kL2 ys ysL1 ysL2 yr*) two robust small*, again it is going beyond, so that means if we put *nocons*, that is *xtabond2 nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, gmm(nL1)) iv(w wL1 k kL1 kL2 ys ysL1 ysL2 yr*) two robust small nocons*, none of these system GMM model could ensure my coefficient to lie within the interval. You may think that I will change the lag

length to 4 what will happen? *xtabond2 n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, gmm(nL1, lag(2 4)) iv(w wL1 k kL1 kL2 ys ysL1 ysL2 yr*) two robust small* Here, when I put 4, the coefficient is lying again beyond the interval.

Here, what I need is 0.73 but it is coming out to be 0.55. Therefore, we get the same problem, probably the only solution here what we have to do is if I remove this ys, now I am considering only this, there is no industrial output and I am again going back to the same command which means I am restricting the *lag* and then let us see *xtabond2 n nL1 w wL1 k kL1 kL2 yr*, gmm(nL1, lag(2 5)) iv(w wL1 k kL1 kL2 ys ysL1 ysL2 yr*) two robust small*. Therefore, this model is basically saying OLS is the same and then I have removed only the industrial output as a proxy for demand as an explanatory variable.

It is again coming down 0.37, so even that is also not working. If I remove *nL2 I* will consider only one period lag, then what happens? *xtabond2 n nL1 w wL1 k kL1 kL2 yr*, gmm(nL1, lag(2 5)) iv(w wL1 k kL1 kL2 ys ysL1 ysL2 yr*) two robust small*. Now this model is working fine, here, what is the specification that I have used? Look at the specification, I have taken only *nL1* as the one period lag of employment in my model, I have taken *w* and then I have taken a lag of *w*, I have taken two periods lag of *k*.

Only these many explanatory variables and then the restriction of lag also to 2 5 and then my coefficient is lying within the interval. So that is the modification I have to make, but in this model when I remove *ys* and lag length what will happen? *xtabond2 n nL1 w wL1 k kL1 kL2 yr*, gmm(nL1) iv(w wL1 k kL1 kL2 ys ysL1 ysL2 yr*)two robust small* This means in this model I have neither considered *ys* which is the proxy for industrial output nor I have given any lag restriction on this, what happens in that context?

This means that even without restriction it comes out to be 0.91, with restriction it is 0.9112, and this is also 0.9141, With little change in the coefficient, coefficients are all significant and then I have other post-estimation checkups like AR1 is there, AR2 is not there that is satisfied here. In this case, without any restriction on lag length and with restriction on the lag length, both are giving almost the same result.

When my model is only considering *nL1* that means one period lag of the employment, but I have removed the *w ys* from the model industrial output. Without any restriction, I have

achieved this *xtabond2 n nL1 w wL1 k kL1 kL2 yr*, gmm(nL1, lag(2 5)) iv(w wL1 k kL1 kL2 yr*) two robust small,* now when I remove ys and put restriction on lag length then I also get *nL1* which is 0.93. So, this is how to learn the models you need to play around with different model specifications, sometimes you impose restrictions on the lags, and estimate difference GMM.

Then you estimate the same difference in GMM without restriction, estimate system GMM with restriction, system GMM without restriction, use only one period lag of employment, use two periods lag of employment and see with different alternative structures of your model specification what type of results you are getting. So here what is the model I have estimated? I have given a restriction lag length up to 5 and I have removed this *ys*.

Now in this model if I include *ys* also what happens *xtabond2 n nL1 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, gmm(nL1) iv(w wL1 k kL1 kL2 ys ysL1 ysL2 yr*) two robust small*, ys, ysL1 and ysL2, now, I am considering once again industrial output but I am just removing the restriction on lag length and I am considering *nL1* only as my endogenous variable and I have removed level, this is a system GMM two-step robust small, again that is 0.93, so what is the difference between these two?

Here I have considered only one period lag of employment, wage, one period lag, capital two periods lag, *ys* and its two periods lag, everything I have considered but I have made only one change, I have not put any restriction on my lag, so that is why if you come down you see again it is lag of 1 to 8 is considered as that optimum lag length. And if you do so then you are again once again able to achieve the coefficients to lie within the interval.

And this specification lastly one again what we will do? We will put all these and then we will put only *ys* now we are considering this also *xtabond2 n nL1 w wL1 k kL1 kL2 yr* ys ysL1 ysL2, gmm(nL1, lag(2 5)) iv(w wL1 k kL1 kL2 ys ysL1 ysL2 yr*) two robust small*, again, it is lying within the interval. So, I have considered everything here nL1, w, wL1, k, kL1, kL2, ys, ysL1, ysL2 everything I have considered. I have imposed lag length also but I have considered only nL1 as endogenous variable.

And I have not considered the second-period lag of employment that is not required. **(Video Ends: 27:56)** So your model specification means when you play around several model

specifications you must ensure one thing your estimates of the lag-dependent variable nL1 is lying within the interval. If that is, achieved then we have to see other aspects of the model which means whether the post-estimation checkups like AR 1 is there or not, AR 2 is there or not so on and so forth.

With this, we are going to close our today's discussion and before we close we would like to summarize what we have learned so far today. We have learned one important thing model specification plays an important role in the quality of the estimates. And when you are particularly interested in dynamic panels, please keep in mind that when I say the quality of the estimates, we need to check whether my estimate is lying within the interval or not.

How will you get those intervals? You estimate the model by FE, you estimate the model with OLS. While OLS will give you an upward upper limit, FE will give you a lower limit and then you come and estimate your dynamic model and see whether that is lying within the interval. These estimates are quite sensitive to model specification. What is the model specification we are talking about?

Which variable are you going to consider as endogenous, how many lags of the endogenous variable you are considering and how many lags of the instruments while determining the GMM style instrument how many lags are you using that is all? So, by changing the specification of my GMM style and iv style variables what we showed today is that the estimates of your coefficients change drastically.

So, which specification you will use first of all you need to read your theory, you need to understand the context, which variable is endogenous and which is exogenous. If your theory says that wage and capital also should be considered endogenous, there is enough reason to believe that wage and capital cannot be considered as exogenous as in the case here. There is a joint determination between wage and labor supply.

There is a reverse causality from employment to capital which means it ensures these variables are also to be considered as endogenous which means they must enter within the GMM bracket. If you enter that, then you need to specify how many lags you are going to use. First, try your model without any specification on the lag and see whether you are able to get your estimates lying within the interval. If it lies well and good.

If it does not then you play around with changing the lag, you impose a lag restriction. So, these are several ways by which you can get alternative specifications to ensure that ultimately the estimates of your dynamic panel data model, particularly the estimated coefficient of the lag-dependent variable is lying within the interval.