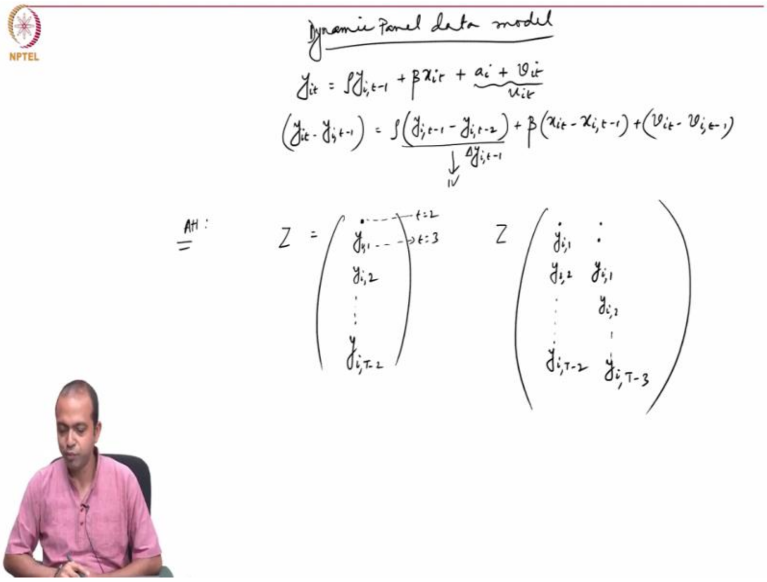
Applied Econometrics
Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology – Madras

Lecture - 42
Dynamic Panel data Model: Part V

Instrumental variable approach can be applied to estimate a dynamic panel data model wherein neither first difference nor the fixed effect transformation was really helpful. So, this instrumental variable approach was basically based on the Anderson and Hsiao suggestion firstly. They only suggested that we can take the first difference of the original dynamic panel equation and since one of the explanatory variables after taking the first difference is getting correlated with the error term, we can still use the first difference model, but we should use an instrument for that explanatory variable which becomes endogenous after first differencing. A brief idea about what exactly is Generalized Method of Moment is discussed before moving to Arellano-Bond's approach so that you can better understand Arellano–Bond's approach. So, I would like to quickly recap what we have already discussed in our previous class. So, it was all about discussing different types of instrument metrics under three approaches. Firstly, we started with Anderson and Hsiao then we discussed about Holtz-Eakin et al and lastly, we saw how Arellano and Bond basically modified the Holtz-Eakin et al idea. So, our dynamic panel equation was :

(Refer Slide Time: 02:25)



$y_{it} = \rho y_{it-1} + \beta \ \ x_{it} + \ a_i + \vartheta_{it}$

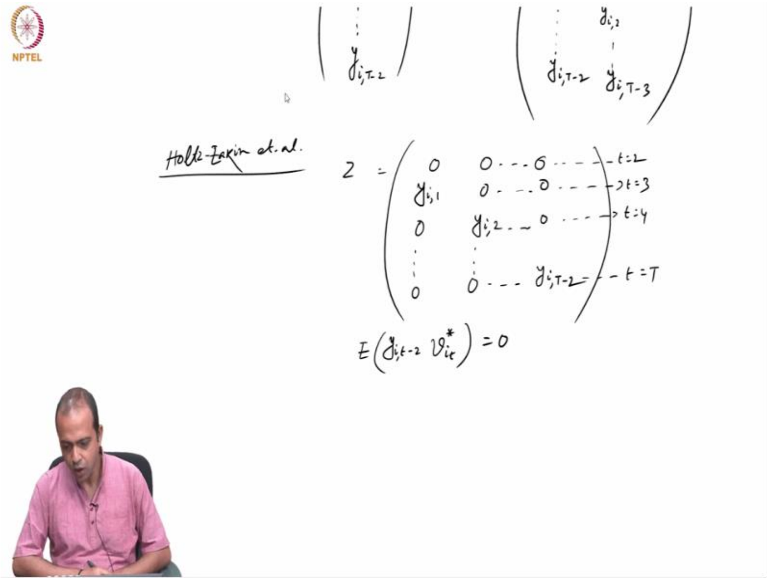
$u_{it} = a_i + \vartheta_{it}$

Taking the First Difference,

$y_{it} - \ y_{it-1} = \rho(y_{it-1} - y_{it-2}) + \beta \ \ (x_{it} - x_{it-1}) + \ (\vartheta_{it} - \vartheta_{it-1})$

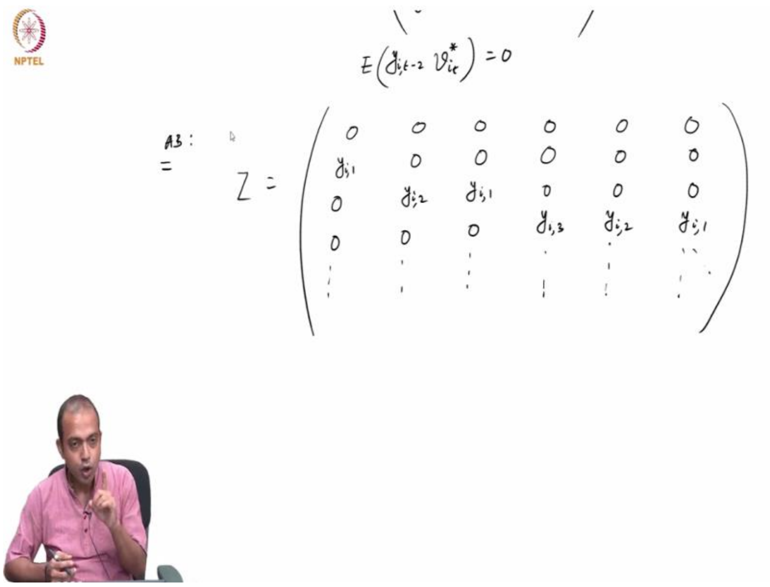
And since this y_{it-1} is correlated with ϑ_{it-1} , so this particular expanded variable is correlated with the transformed error term. So, that is why we say that we need to use IV for this particular variable $y_{it-1} - y_{it-2}$ which is basically this is nothing but Δy_{it-1}. The instruments that can be used are either the twice lag of the level y_{it} that means it will become y_{it-2} or twice lag of the difference that means Δy_{it-1}. So, initially we were discussing about taking y_{it-2} as an instrument. And then in Anderson and Hsiao approach (AH in short) Anderson and Hsiao approach in the instrument matrix first observation would be lost and then we will get y_{i1} and then y_{i2} and then it will become $y_{i,T-2}$. So, this is that means observation is lost for each panel. So, sample size goes down and if we take the thrice lag also as an additional instrument then your Z matrix become basically this is $y_{i1},\ y_{i,t2}$ then $y_{i,T-2}$ and when I take the thrice lag also as an additional instrument I will lose two observations. These will become your instrument matrix. So, that means in Anderson and Hsiao's approach when I add more lag into the instrument matrix, my lag length is increasing. But at the same time my sample length is decreasing and that is why we say that there is a trade-off between lag length and sample length. That was Anderson and Hsiao's approach and then to overcome the sample length what Holtz-Eakin et al they suggested is that instead of taking too many lags as instrument we will take only one instrument for each period and that is second lag of the level variable and we will replace all the missing observations as 0.

(Refer Slide Time: 07:30)



So, in Holtz-Eakin et al they said that instrument matrix should be this way. (as given in the above figure). Only second lag is taken as instrument and all the missing observation will be replaced as 0. So, that means in each period you can see that I have only one instrument per period. So, this is Holtz-Eakin et al approach where one instrument is taken for each period and replacing all the missing observations are 0 and then we also discuss that even though setting 0 in the place of missing observations appears like arbitrary. Interestingly, each column of this instrument matrix satisfies the orthogonality condition that means $E(\ y_{i,t-2}\ \vartheta *_{it})$=0 and then in this approach even though we ensured enough sample our lag length was confined only to 2.

(Refer Slide Time: 09:38)



And based on this Holtz-Eakin et al approach lastly, we discussed about Arellano and Bond's instrument matrix wherein Arellano and Bond says that instead of using only one instrument for each period why do not you do one thing. You take the idea of Holtz-Eakin et al that means second lag of the untransformed variable for each period as the instrument and also additionally whatever lag is available for that particular period for all the variables we will include. And then we say that for endogenous variable it would be lag 2 and higher, and for predetermined variables which includes the endogenous variable as well as lagged dependent variable we will take even first lag also because that would be correlated only the error term with (t-2) and earlier. So, then we said that in Arellano and Bond's approach or instrument matrix was this. (as given in the above figure)

And then this is basically your Z matrix in Arellano and Bond's method that means we can easily see that in Arellano and Bond's method we have too many instruments. So, 1 instrument for second period, 2 instrument for third, 3 instrument for the fourth and so on. So, too many instruments are used in this method and then this method is basically dependent on we say that Arellano and Bond's method is basically a method of generalized method of moment. Why we are using moment? Because we are using all potential orthogonality conditions available that means we are using all potential moment conditions here and that is why it is an approach called as generalized method or moment wherein I will tell you what exactly is method of moment, where your number of instruments are actually much more than number of parameters to be estimated.

Major assumptions in the context of Arellano and Bond's model to be applied we need to keep this thing in mind. So, we have only one left hand side variable which is dynamic in nature. So, that means if I go back, you can see that only left-hand side variable which is dynamic in nature. Addition of the lagged variable makes the system dynamic and if you recall we have also discussed that for dynamic stability the value of ρ should be less than 1. So, if ρ equals to 1 then we proved in our last class then the model does not become dynamic it becomes a static one.

And if ρ is greater than 1 that does not define the stability of the dynamic system. You can easily understand by subtracting $y_{i,t-1}$ from both the sides. So, this would become

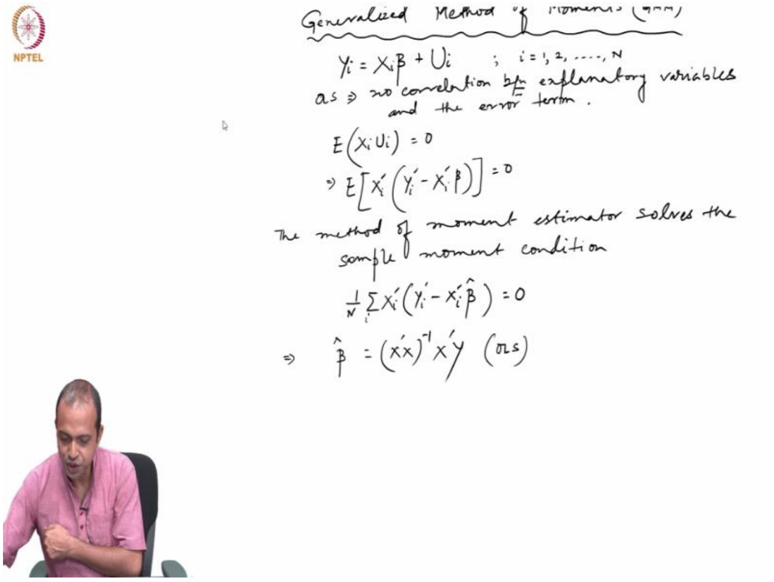
$$y_{i,t} - y_{i,t-1} = (1-\rho)\, y_{i,t-1} + \beta x_{i,t} + a_i + \vartheta_{it}$$

From this equation to get a dynamic panel data model and dynamic stability ρ should be less than 1.

And that is the reason we say this is also partial adjustment model because ρ equals to 1 means complete adjustment in the same period that does not give any dynamic panel image. So, with this now what I will do. So, number one there is only one variable in the left-hand side which is dynamic in nature and then the second assumption is that the relationship is linear. This is a linear relationship between $y_{i,t}$ with $y_{i,t-1}$ and any additional explanatory variable that might be there in the system. Number two and very importantly here as we said that number of time period t should be small compared to number of observations n. So, it is a case of micro panel where t is very small and n is very large. Why this is so? Because when you take the first difference we said that first difference estimate would be biased and inconsistent and the order of bias will be 1/T.

So, that means as time period increases, bias will go down. So, that means if your t is very, very large then actually we do not have to use any other dynamic panel data model to estimate. We do not need any separate technique to estimate this type of dynamic panel data model- simple OLS will do but when t is small then we proved that the bias is quite high and even when t equals to 30 we prove that the bias is 20 percent of the original estimates that is the second assumption. And then we said that we have one unobserved or individual specific unobserved effect in the right-hand side and then we have also mentioned that we have auto correlation of order 1 in the system because that is by the construction of the model says the moment you introduce lag dependent variable it will result in autocorrelation. So, AR(1) should be there in the system, but not AR(2). And we should not be happy if even AR(1) is also not there because the moment AR(1) is not there then you should understand that means this model is not at all dynamic in nature and then we said that autocorrelation and heteroscedasticity will be there within each of this individual, but not across. So, these are the contexts, these are the different assumptions that define Arellano–Bond context for dynamic panel data model.

(Refer Slide Time: 18:18)



To give a brief overview of Method of Moment - GMM or Generalized Method of Moments. OLS is also as a method of moment estimator. OLS estimator what we got that we can derive from the method of moment's approach also.

So, let us assume for the time being this is our model y i equals to X i beta plus U i. okay. and here i equals to 1. why I am writing X i beta instead of beta X i because these are all matrix notation that is why y i equals to X i beta plus U i that is my matrix notation and then if you recall one of our standard assumptions in classical linear regression model that none of my explanatory variable should be correlated with the error term.

None of my explanatory variables should be correlated with my error term and in terms of expectation what you can write that means OLS assumes no correlation between explanatory variables and the error term. This is the assumption and if this assumption is violated then we say that OLS estimates will no longer be consistent. So, basically this means expectation of X i and U i equals to 0.

No correlation between explanatory variable and the error term indicates expectation X i U i equals to 0 this is called a population moment condition also writing this expectation X i U i equals to 0 is a population moment condition and then this actually implies expectation of X i prime instead of U i now I will replace. What is U i? U i is nothing, but y i minus X i beta so this will become y i transpose minus X i transpose into beta that is equals to 0.

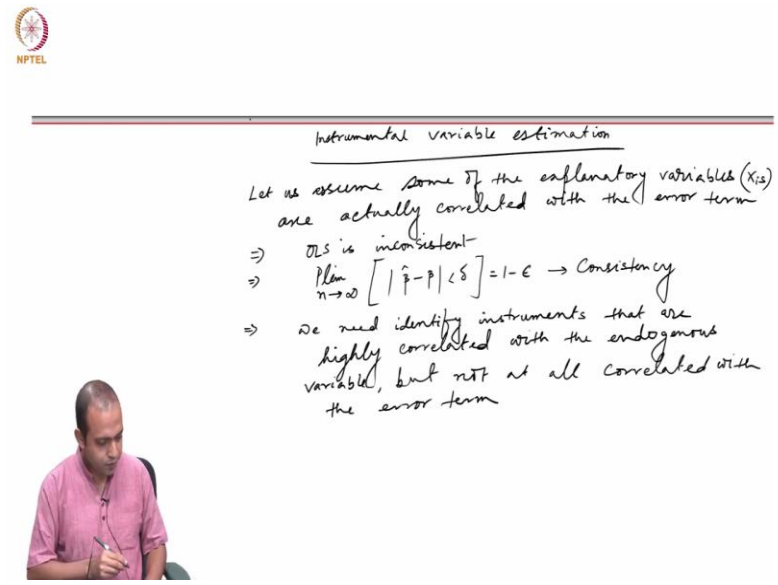
Very simple. in place of U i I am just putting y i prime minus X i prime beta this is all transpose basically. Now the method of moment condition, the method of moment estimators basically solves estimator the sample moment counter part of this population moment condition that is why I am saying the method of moment estimator solves the sample moment condition and what is the sample moment condition?

Basically, it is 1 by N where N is the sample size 1 by N summation you take the summation of these X i prime and then y i prime minus x i prime into beta hat right in this is equals to 0. I have just taken the sample moment counter part of this population moment condition which is 1 by N summation x i prime beta x i prime y i minus x i beta and if you solve this equation then you will get your OLS estimator as beta hat which is nothing, but x prime x inverse x prime y which is the OLS estimator.

So, that means what I have proved that OLS estimator is nothing, but a sample moment condition derived from the population moment, derived from the important assumption that no explanatory variable should be correlated with the error term which basically gives a moment condition which is population moment condition and expectation X i U i equals to 0. So, from that equation I have derived the sample moment condition.

And after solving that sample moment condition I got my OLS estimator as the moment estimator method of moment estimator also.

(Refer Slide Time: 25:23)



Now to come to instrumental variable estimation, basically the assumption was that in OLS, no explanatory variables should be correlated with the error term. Now I assume that some of the X_i's, some of the explanatory variables are actually correlated with the error term because that will lead to endogeneity condition. And endogeneity means we need to apply instrumental variable approach, we need to find down an instrument let us say Z which will be highly correlated with the endogenous variable, but not at all correlated with the error term. So, that means let us assume some of the explanatory variables or X_i's are actually correlated with the error term. So, if that is the case then when error terms are correlated this explanatory variable under OLS is inconsistent. Consistency means the condition where the difference between $(\widehat{\beta -} \beta)$ should be very, very minimum less than the minimum quantity delta and probability of getting such beta when n tends to infinity is equal to 1 – ε.

So, that means as n tends to infinity in a large sample my $\hat{\beta}\ will\ converge\ to\ \beta.$ So, that means the difference between beta hat and the true population parameter will be very, very small and probability of getting such beta is also almost 1 because ε is also very small quantity. So, this is called consistency property of OLS and this consistency property will no longer be valid the moment we assume that some of the explanatory variables are actually correlated with the error term. So, if that is the case, OLS become inconsistent and we need to identify instruments that are highly correlated with the endogenous variable but not at all correlated with the error term.