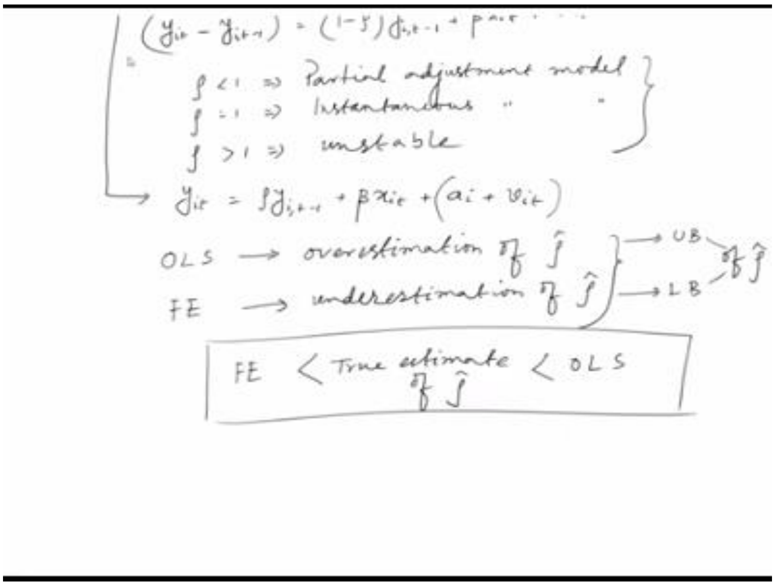
Applied Econometrics
Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology-Madras

Lecture - 56
Dynamic Panel Data Model - Part XIX

Welcome to our discussion on the dynamic panel data model. Today what we will do since we have already completed all the discussions about the theory and I have also demonstrated using the statistical software how to estimate a dynamic panel data model and the different issues involved while you estimate, in today's class I will just summarize from the beginning so that you will be able to know what are the key elements that you should keep in mind while estimating a dynamic panel data model.

This is the summary of the entire discussion that we had in the dynamic panel data model.

(Refer Slide Time: 01:34)



Our model is $y_{it} = \rho y_{i,t-1} + \beta x_{it} + u_{it}$, where $u_{it} = a_i + v_{it}$ which is actually summation of the term observe individual specific effect plus the idiosyncratic error. Firstly, we need to keep in mind that before applying a dynamic panel data model, we need to first justify theoretically, why you think that you need a dynamic panel data model in your context.

Therefore, your situation should be of partial adjustment type. That means, for any changes in your explanatory x_{it}variable, y_{it} should not be able to adjust itself in the same period. If it does, then it would be a case of instantaneous adjustment. When it fails to adjust in the same period, then only we call it a partial adjustment model.

That partial adjustment model is also called the dynamic panel data because by partially adjusting, dynamism is introduced in the model. So, a simple transformation that we have discussed is, if we subtract $y_{i,t-1}$ from both the sides of this equation 1, then what will we get is $y_{it} - y_{i,t-1} = (1-\rho)y_{i,t-1} + \beta x_{it} + u_{it}$.

So, this is the model and by partial adjustment. Here, what we mean is that rho should be less than 1 ($\rho < 1$). This means, that if $\rho < 1$ then only we get a partial adjustment model. If $\rho = 1$, $y_{i,t-1}$ will be 0. That means, the lag-dependent variable will disappear from the model and that will imply an instantaneous adjustment model which is basically the static panel data model and ($\rho > 1$) is unstable.

This is a situation of partial adjustment. There might be many reasons for getting partial adjustment or a dynamic panel data model. We have discussed only one case wherein we were discussing mainly the employment data, employment of a firm or employment by any sector that depends on what is the employment you are already having because what we discussed is the hiring of a new employer and firing the existing one is costly.

That is why employment depends on your previous year's employment, $y_{it} = f(y_{i,t-1})$. There might be other cases. For example, if you want to model let us say, a firm’s output that is also kind of dynamic in nature, because your t^{th} period output may depend on the previous period’s output how much you have already produced, how much you could sell and how much was there in your inventory.

There might be many cases for that. Therefore, we need to carefully justify our need for a dynamic panel data model. Otherwise, just because we have some data and we know the code about how to estimate in the statistical software, we should not use when it is not required. We need to carefully understand the situation and justify.

Now in case we have a dynamic panel data model, then what happens here in this model $y_{it} = \rho y_{i,t-1} + \beta x_{it} + v_{it}$. Since y_{it} is a function of a_i and there is no *t* subscript over here what we discussed that y_{it-1} is also correlated with a_i.

That means it leads to endogeneity because a_i and v_{it} constitute the composite error term. Therefore, there is endogeneity in this model and OLS is not applicable. Then when OLS is not applicable, if we do still estimate the model using OLS, we will end up with having a dynamic panel bias and discussing or applying OLS actually leads to this situation of overestimation of $\hat{\rho}$.

To remove this a_i we apply the fixed effect transformation like what we used to do in the context of the static panel data model, then what we discussed that a fixed effect model would lead to underestimation of $\hat{\rho}$, which means OLS will give you an upper bound and if FE will also give a lower bound of $\hat{\rho}$.

OLS is giving upper bound because we are ignoring that there is dynamism in the system and what is happening here, a positive correlation between $y_{i,t-1}$ and a_i . We have also given an example why $y_{i,t-1}$ and $y_{i,t}$ is positively correlated because you can understand you can take a situation where the firm is experiencing a negative employment shock.

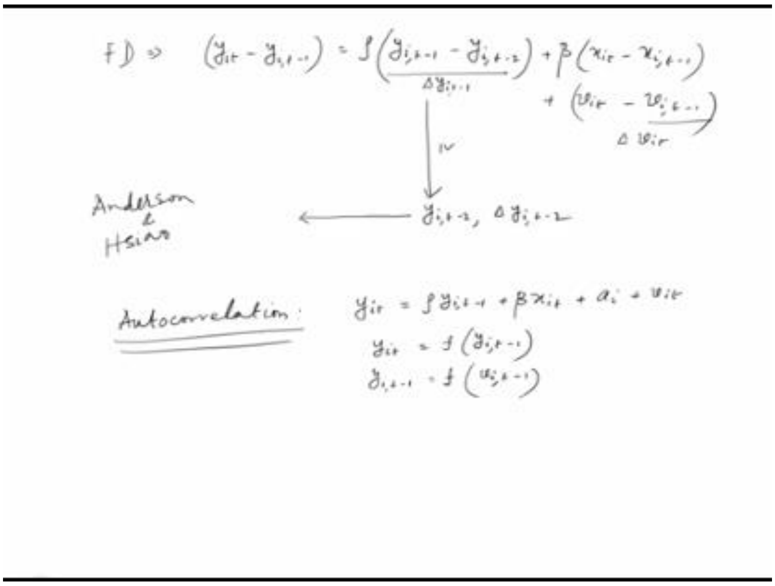
Then in the next period both a_i the unobserved effect and $y_{i,t-1}$ would be lower than what was happening earlier. So, in all the consecutive periods rather your $y_{i,t}$ in a_i and $y_{i,t-1}$ would be lower and as a result of which they will have a positive correlation that we have already discussed.

And in fixed effect transformation it will have a negative bias because the transformation when you make $y_{it} - \bar{y}$ then you will get $v_{it} - \bar{v}$, a negative sign before that v i bar will lead to negative correlation between y_{it-1} and your v_{it} and as a result of which fixed effect transformation will give you under estimation of $\hat{\rho}$. And these two will give you the upper and lower bound of your estimate.

So true estimate must lie true estimate of $\hat{\rho}$ then should be greater than your fixed effect, but less than your OLS. So, this is something whenever we are applying and estimating a dynamic panel data model, we need to first check the upper and lower limit of your estimates.

And then applying dynamic panel data model estimation technique when you get a $\hat{\rho}$, then you will see whether your $\hat{\rho}$ is actually lying within the interval of FE and OLS. Then we have also discussed that when neither OLS nor FE is applicable, the other transformation is first difference.

(Refer Slide Time: 11:35)



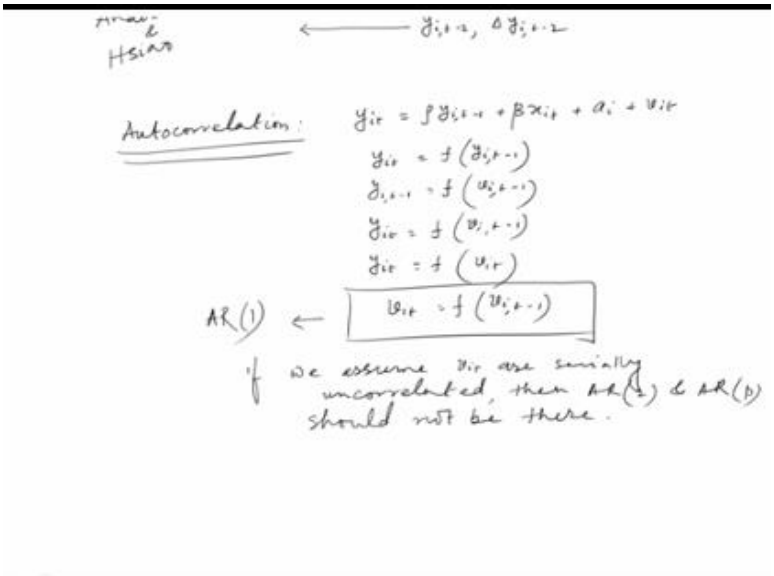
So, FD transformation we can apply and that will make it $(y_{it} - y_{i,t-1}) = \rho(y_{i,t-1} - y_{i,t-2}) + (\beta x_{it} - x_{i,t-1}) + (v_{it} - v_{it-1})$. Here what is happening even after first difference we can see that $y_{i,t-1}$ and v_{it-1}, means the variable $\Delta y_{i,t-1}$ is actually correlated with $\Delta v_{i,t}$ and as a result of which we are again ending up with endogeneity.

So, even first difference transformation also cannot rule out the possibility of endogeneity because this is correlated. So, what is the solution then we offered? We said that when FD is not workable, because of this endogeneity we will use IV either $\Delta y_{i,t-2}$. These are the two instruments we can use. And that method we said that, this is the suggestion given by Anderson and Hsiao.

He said we should take the first difference. We should take the first difference and then use the instruments. Now one thing we should mention over here about the autocorrelation problem. In this model when we are writing $y_{it} = \rho y_{i,t-1} + \beta x_{it} + a_i + v_{it}$, what we are seeing that y_{it-1} is correlated with a_i.

And what we can write is that y_{it} is a function y_{it-1}. Now the way I have written the equation, I can write the same equation for y_{it-1}. And in that equation the error term would be v_{it-1}. So that means y_{it-1} function of v_{it-1}. So, this means, y_{it} is also a function of v_{it-1} and y_{it} is already a function of v_{it}.

(Refer Slide Time: 16:00)



So that means v_{it} is correlated with y_{it}, and y_{it} is correlated with v_{it-1}. Then ultimately v_{it} is a function of v_{it-1}. So that means, the presence of lag dependent variable is leading to autocorrelation and that autocorrelation is of order 1. Therefore, the structure of the model itself gives AR 1. But if we assume that v_{it} are serially unrelated, uncorrelated then AR 2 and higher, let us say AR P should not be there.

Now what will happen actually if AR 2 is there? So, if you look at what is the instrument we are using.

(Refer Slide Time: 17:55)

$$y_{it} = f(v_{i,t-1})$$

$$y_{it} = f(v_{it})$$

$$AR(1) \leftarrow \boxed{v_{it} = f(v_{i,t-1})}$$

if we assume v_{it} are serially uncorrelated, then AR(2) & AR(p) should not be there.

What happens when AR(2) or AR(p) exist?

$$AR(2) \Rightarrow v_{it} = f(v_{i,t-2})$$

$$y_{i,t-2} = f(v_{i,t-3})$$

$$\Rightarrow y_{i,t-2} = f(v_{i,t})$$

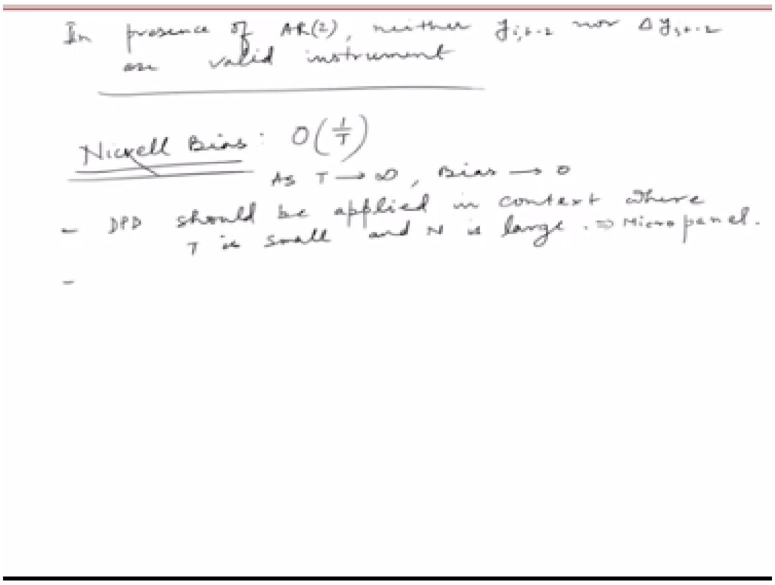
We are using $y_{i,t-2}$ assuming that $y_{i,t-2}$ would be correlated with $\Delta y_{i,t-1}$, but they will not be correlated with v_{it}-Δv_{it}. But in case there is higher order autocorrelation, that means v_{it} is also correlated with in presence of AR 2 v_{it} is a function of $v_{i,t-2}$ also second order correlation is also there.

That means earlier what you were assuming that this $y_{i,t-2}$ will not be correlated with this transform error term that will no longer valid. Because v_{it} itself is correlated with v_{it-1} and that means $y_{i,t-2}$ is a function of $v_{i,t-2}$. So that means, $y_{i,t-2}$ is also a function of actually v_{it}. So, $v_{i,t-2}$ can no longer be an instrument if there exist second or higher order autocorrelation.

That is why when we estimate the dynamic panel data model, we must check whether AR 1 is there and whether AR 2 is not there. AR 1 must be there because of the construction of the model. If AR 1 is not there that means equivalently, we are saying that there is actually no need of estimating a dynamic panel data model. Because there is no importance of y it minus 1.

That is why by construction AR 1 must be there, but AR 2 should not be there. If AR 2 is there that means as if we are seeing in presence of AR 2 neither $y_{i,t-2}$ nor $\Delta y_{i,t-2}$ are valid instrument.

(Refer Slide Time: 20:57)



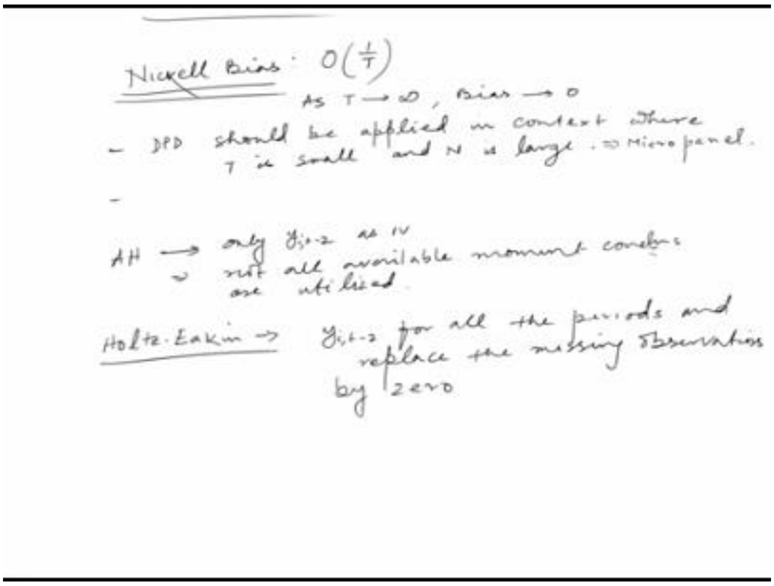
That is one thing we need to keep in mind. Then another important point, we talked about dynamic panel bias and we also talked about nickel bias. That means, the application of FE and FD transformation, how that leads to nickel bias. Because both, in case of both FE transformation and FD transformation we saw that we could not rule out the possibility of endogeneity.

And this nickel bias is of order 1 by T. So that means, as T tends to infinity this bias tends to 0. That is why whenever we will be estimating dynamic panel data model or DPD, DPD should be applied in context where T is small and n that means the number of observations in the panel is large. That is also one thing we have to keep in mind. If T becomes large then the bias will automatically get vanished.

We do not have to estimate the model using this complicated dynamic panel data estimation technique. So that is one thing we have to keep in mind. That this is a situation wherein my T is actually small, n is large, that means we are talking about a micro panel.

So we will when there is this type of problem, T is small n is large, and we have lag dependent variable in the right hand side and then we are assuming that there is no AR 2 we can go ahead with this estimation suggested by Anderson and Hsiao wherein we will transform the model. We will take the first difference and we will use $y_{i,t-2}$ or $\Delta y_{i,t-2}$ as instrument. So, what was happening in case of Anderson and Hsiao's model?

(Refer Slide Time: 24:28)



Anderson and Hsiao's main problem was that in their model since they said only $y_{i,t-2}$ as IV, so this implies not all available moment conditions are utilized. Therefore, we can increase the efficiency of $\hat{\rho}$ to a larger extent following this Holtz-Eakin technique. What they say that we will use $y_{i,t-2}$ as instrument for all the periods and replace the missing observations by 0.

Because in Anderson and Hsiao's approach the moment we use other lags also as instrument then we saw then your sample length was going down. So, there was a tradeoff between lag length and sample length that we discussed. Following Anderson and Hsiao's approach, the more lag you include the more observations we are going to lose.

Then Holtz-Eakin et al., they came up and say that we can actually use one single instrument $y_{i,t-2}$ for all the periods and replace the missing observations by 0. That was Holtz-Eakin et al., suggestion. But then again, the original idea is to use more moment conditions, whatever lags are available we need to use all of them to increase efficiency or rho hat.

That is the reason next model is the most popular one came up by Arellano and Bond 1991.

(Refer Slide Time: 27:22)

Holtz-Eakin → $y_{i,t-2}$ for all the periods and replace the missing observations by zero

Arellano & Bond (1991) : Use $y_{i,t-2}$ as well as all other lags as and when available.

What they say that use $y_{i,t-2}$ as well as all other lags as and when available. That is what Arellano and Bond said. For example, in third period what will happen? You will have only $y_{i,t-2}$. That means y_{i1} is available. But when you go to fourth period, then we will have $y_{i,t-2}$ that means y_{i2} as well as y_{i1}. Because $y_{i,t-2}$ and $y_{i,t-3}$ is also available.

When you go to fifth period, then $y_{i,t-2}$, $y_{i,t-3}$, $y_{i,t-4}$, all these three are available. As we move on to the higher periods, we will get more and more lags and we will include whatever is available and missing observations, as Holtz-Eakin et al suggested we will replace by 0. In that way, Arellano and Bond they increased the efficiency of their estimates.