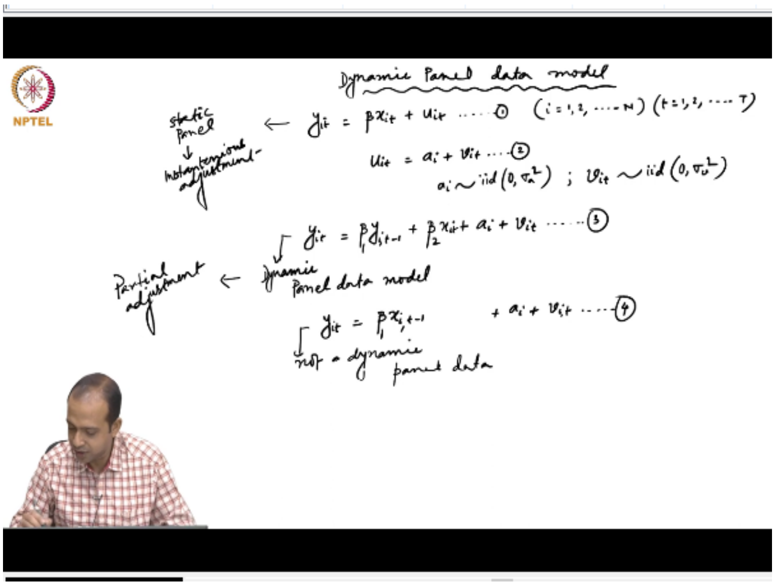
Applied Econometrics

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

Lecture-38
Dynamic Panel Data Model-Part I

Dynamic panel data model is an advanced econometric model in the class of panel data. Even though we are taking several years data and that itself gives a sense in our mind that panel data is dynamic while the cross section is static in nature, but that notion is actually not true even though we take several years data while modelling a panel, constructing a panel the panel data itself is not dynamic in nature. So, to start with let us talk about the simple static and fixed effect panel data model what we have discussed earlier.

(Refer Slide Time: 01:59)



So, this is $y_{it} = \beta x_{it} + u_{it} - - - - (1)$, Equation (1) which is assumed to be a simple panel data model and we assume $u_{it} = a_i + \vartheta_{it} - - - - (2)$ where a_i is basically the individual specific time constant unobserved effect and ϑ_{it} is the standard idiosyncratic error and we assume that i = 1, 2, n and t runs from 1 to T and further we assume that a_i follows iid with 0 mean and σ^2 variance. And similarly ϑ_{it} also follows iid with 0 mean and σ^2 variance. It means both $a_i \; and \; \vartheta_{it}$ are independently and identically distributed that means there is no relationship correlation between $a_i \; and \; \vartheta_{it}$ and they are not serially related as well.

Why the simple panel data model with equation 1 is not a dynamic one?

Here y_{it} means t[th] period y depends on the value of the x on the same period. If any changes happen to x then y will also change and adjust itself in the same time period t. That means the changes in x will definitely result in a change in y but the change is instantaneous in nature. That is why this the equation 1 indicates the panel data model. It is a simple fixed effect panel data model and is not dynamic in nature because there is no dynamism, that is there is no persistence over a period of time. Any changes in x will result in instantaneous changes in y. That means this model is called instantaneous or instantaneous adjustment and that is why this framework is called static panel data model. Thus, this type of model is called static panel data model because the adjustment happens in the same period that means the adjustment happens instantaneously.

So, when we need to construct a dynamic panel data model that means we need to introduce dynamism in the system and how do you introduce dynamism in the system?

So, let us assume another class of model where,

$y_{it} = \beta_1 y_{it-1} + \beta_2 x_i + \ a_i + \vartheta_{it} ----(3)$

What is the difference between model 1 and model 3? In model 3 we have introduced a lag dependent variable and by introducing lag dependent variable we are actually introducing dynamism in the system. So, that means if anything happens to y_{it-1} that will carry forward in the next period. So, that means y_{it} depends on its previous value. So, by including a lag dependent variable we are actually introducing dynamism in the system. This is called a dynamic panel data model. So, this is a dynamic panel data model. So, this model is basically not an example of an instantaneous adjustment rather this is called partial adjustment model. So, $y_i$cannot adjust itself in the same period rather it also depends on its previous value. That is why this model becomes a partial adjustment model or dynamic panel data model. So, this is called partial adjustment contrasting to the previous one which was instantaneous adjustment.

Now there is another class of model where, $y_{it} = \beta_1 x_{i,t-1} + \beta_2 x_{it} + \ a_i + \vartheta_{it} ----(4)$

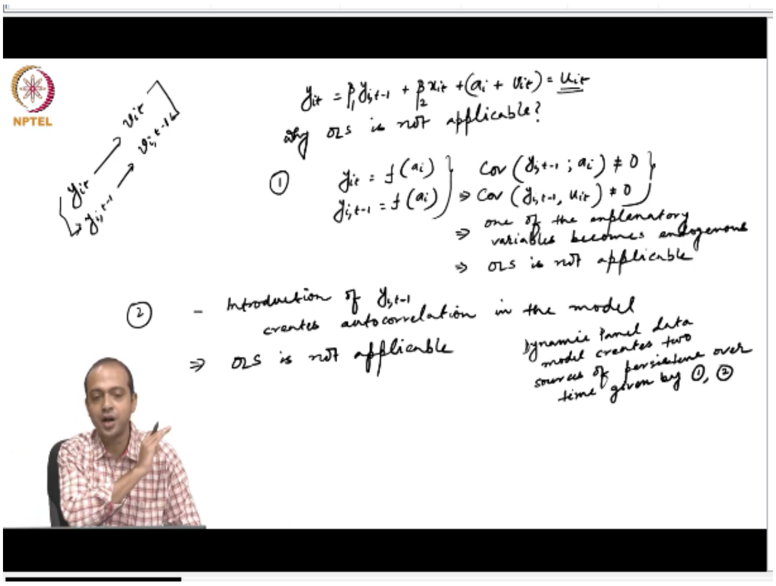
So, instead of introducing lag value of the dependent variable now I have introduced lag value of the independent one. Now by introducing lag value of the independent variable do you think I have introduced dynamism in the system? The answer is no. Because whatever happens to x, that means in the previous periods x_i value y_{it} will adjust in the same period. So, just that instead of x_{it} , you can remove this x_{it} also from the model and you can simply say that this is basically

$y_{it} = \beta_1 x_{i,t-1} + \quad a_i + \vartheta_{it} ----(4)$

So, instead of y_{it} depending on x_{it} it just depends on the previous value of x_{it}, but that does not mean that we have introduced dynamism in the system. So, this is not a dynamic panel data model. So, model will become dynamic only when we introduce lag value of the dependent one.

So, this dynamic panel data model given by equation 1 and 2 actually it introduces couple of problems regarding as far as estimation is concerned. Let us try to understand what are the problems of applying ordinary least square method in estimating this type of model given by a dynamic panel data.That means the moment you introduce lag dependent variable it calls for certain problems that makes OLS to be inapplicable.

(Refer Slide Time: 12:49)



So, I will write the dynamic panel data model once again which is,

$y_{it} = \beta_1 y_{it-1} + \beta_2 x_i + \ a_i + \vartheta_{it}$

Now first of all the question is why OLS is not applicable here? So, you can easily understand here y_{it} is basically a function of a_i. There is no ‘t’ subscript here. So, that means I can say that even y_{it-1} is also a function of a_i.

That is, if y_{it} depends on a_i and a_i has no ‘t’ subscript, that means y_{it} depends on the unknown observed individual specific and time constant since this factor is time constant it remains unchanged. If y_{it} is a function of a_i, y_{it-1} is also a function of a_i. So, that means the covariance or correlation between y_{it-1} and a_i is actually not equal to 0 and if you can recall one of our 10 assumptions of classical linear regression model was that none of the regressors should be correlated with the error term.

When u_{it} is introduced as $u_{it} = a_i + \vartheta_{it}$, it basically indicates u_{it} is one way error component model because I have introduced only individual specific time constant effect. I have not introduced anything which accounts for the time effect. That means I have not assumed that even the times are also different among each other. So, a two-way error component means, $u_{it} = a_i + \mu_t + \vartheta_{it}$, where μ_t does not change across individual but changes only over a period of time. That factor is not there. So, since I have introduced only a_i this is called one way error component model.

So, as you all know there are two types of panel data model; one way and two-way error component model. For simplicity sake, for the time being we are just assuming that this model is one way error component model that is all. Now here, since y_{it-1} is correlated with a_i, so covariance between y_{it-1} and u_{it} is not equal to 0. So, that means one of your explanatory variable has now become endogenous. That basically indicates OLS is not applicable. This is one problem. So, introduction of the lag dependent variable creates this problem - one of my explanatory variable becomes correlated with the error term.

Secondly what happens here? Introduction of the lag dependent variable creates autocorrelation in the model. That is, introduction of y_{it-1} creates autocorrelation in the model. So, we are basically trying to understand how introduction of lag dependent variable creates autocorrelation in the model. y_{it} is related with ϑ_{it} and y_{it-1} is similarly related to ϑ_{it-1}.

And since y_{it} is a function of y_{it-1}, I can say that then ϑ_{it} is actually a function of ϑ_{it-1}. So, that means since y_{it} is related with y_{it-1} and $y_{it-1}$is related to ϑ_{it-1}, y_{it} is also related to ϑ_{it}. So, that means this is a correlation between ϑ_{it} and ϑ_{it-1}. That is, ϑ_{it} is getting correlated with ϑ_{it-1} through this y_{it-1} because y_{it-1} is correlated with ϑ_{it-1} and y_{it-1} is also correlated with y_{it} .

That is why the moment we introduce lag dependent variable in the model it will lead to auto correlation and as we all know that if we have autocorrelation in the model then we cannot apply OLS to estimate the parameters because that will lead to inconsistent and biased estimates. So, introduction of y_{it-1} creates autocorrelation in the model and that implies again OLS is basically not applicable. So, this is one type of persistence.

That means these two are two types of persistence over a period of time in a dynamic panel data model. First one is the autocorrelation created by the lag dependent variable; second one is this

a_i is also getting carry forward over a period of time because a_i component is there in the composite error structure. So, at any point of time your error component will always be capturing a_i. So, there are two types of persistence that is why again and again we are discussing that OLS is not applicable. So, then naturally one question comes to our mind why not removing this a_i component so that at least we will get rid of one type of persistence like the way we you are doing in the context of fixed effect and first differencing of the panel data model. So, since there is a problem with this individual specific time constant effect which is creating one sort of persistence in the model, we are trying to eliminate this individual specific time constant unobserved effect by fixed effect transformation and let us see what will happen.