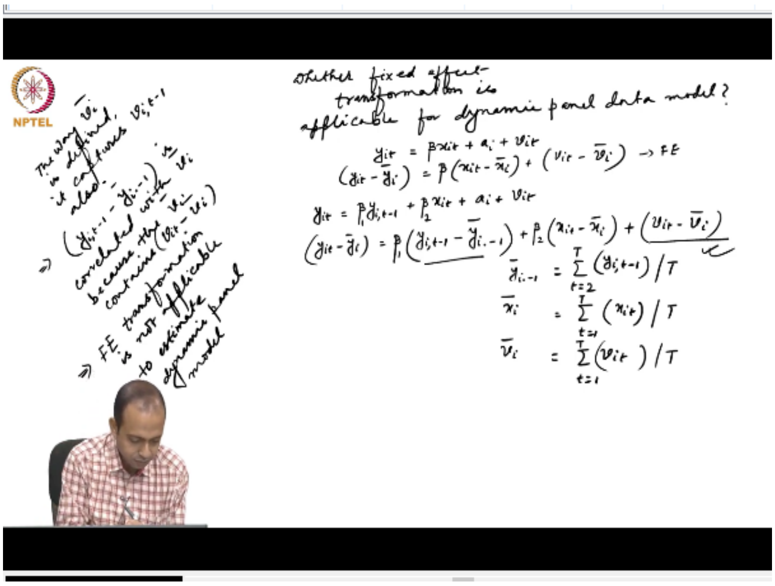
Applied Econometrics
Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology-Madras

Lecture-39
Dynamic Panel Data Model-Part II

(Refer Slide Time: 00:22)



So, we are trying to understand whether fixed effect transformation is applicable for dynamic panel data model. Our model was : $y_{it} = \beta x_{it} + a_i + \vartheta_{it}$. Fixed effect transformation means time demeaning. So, : $y_{it} - \bar{y_\iota} = \beta(x_{it} - \bar{x_\iota} + (\vartheta_{it} - \overline{\vartheta_\iota})$

That is called fixed effect transformation. This is FE. Same thing we will try to apply in the context of a dynamic panel data model. So, dynamic panel data model is:

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

Same fixed effects transformation we are trying to apply here. So, this will become

$y_{it-1} - \bar{y_\iota} = \beta_1(y_{it-1} - \overline{y_{\iota.-1})} + \beta_2(x_{it} - \overline{x_\iota)} + \ (\vartheta_{it} - \overline{\vartheta_\iota)}$

$\overline{y_{\iota.-1})} = \sum_{t=2}^{T} (\ y_{i,t-1})/T$, because we have taken the difference. So, that is why one observation is lost. So, while summing over we are taking the summation from t equals 2 to T

$\bar{x_\iota} = \sum_{t=1}^{T} (\ x_{i,t})/T$

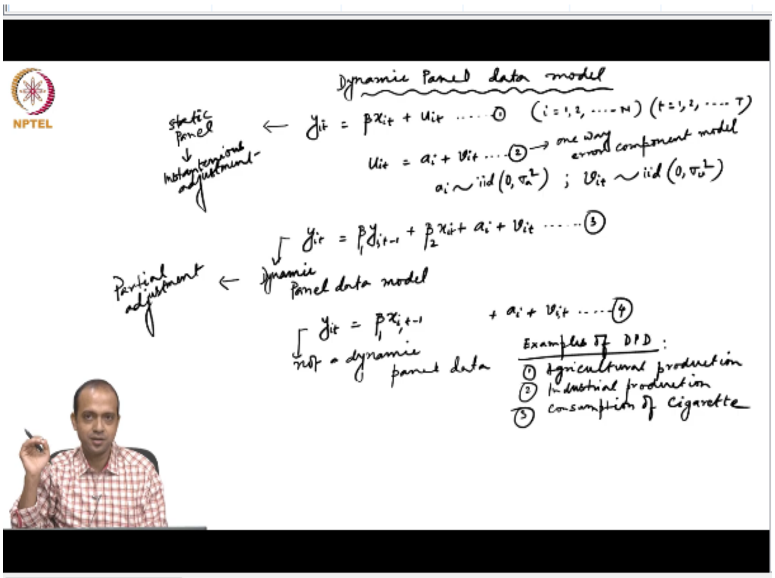
$\bar{\vartheta_\iota} = \sum_{t=1}^{T} (\ \vartheta_{it})/T$

Now if we look at carefully since $\overline{\vartheta_\iota}$ which is the sum of $\vartheta_{it})/T$, t equals 1 to T. So, obviously within this (ϑ_{it-1} is also there. So, that means I can write by the way I have defined the way $\overline{\vartheta_\iota}$ is defined it captures ϑ_{it-1} also. That is why simply by construction of $\overline{\vartheta_\iota}$ and you see one of your explanatory variables is y_{it-1}. since $\overline{\vartheta_\iota}$ contains ϑ_{it-1}, so that means what I can say that $y_{it-1} - \overline{y_{\iota.-1}}$ will be is correlated with $\overline{\vartheta_\iota}$ because $\overline{\vartheta_\iota}$ contains ϑ_{it-1}.

Obviously, y_{it-1} would be correlated with its own error term ϑ_{it-1}. So, that is simply by construction. So, even though ϑ_{it}they themselves are not serially correlated the way we have modeled this dynamic panel data you see one of the explanatory variables which is $y_{it-1} - \overline{y_{\iota.-1}}$ is actually correlated with your error term which is basically $(\vartheta_{it} - \overline{\vartheta_\iota)}$ because $\overline{\vartheta_\iota}$ actually contains ϑ_{it-1}.

So, this shows that fixed effect transformation is also not a helpful to estimate a dynamic panel data model because what is happening even though we were able to eliminate the a_i component we are not able to rule out the possibility that one of your explanatory variables is getting correlated with the error term. The transformed error term we are talking about is $(\vartheta_{it} - \overline{\vartheta_\iota)}$ and the explanatory variable here is $y_{it-1} - \overline{y_{\iota.-1}}$. So, your explanatory variable is this which is correlated with the error term. So, OLS is not applicable that means fixed effect transformation which shows that FE transformation is not applicable to estimate a dynamic panel data because explanatory variable is getting correlated with the error term. Now we have learned one more transformation which is called first difference. Let us see whether we can go for the first differencing of the dynamic panel data model and apply OLS.

(Refer Slide Time: 10:50)



So, this is our model: $y_{it} = \beta x_{it} + a_i + \vartheta_{it}$

We will see whether first differencing is applicable. So, if we do the differencing, it implies, $y_{it} - y_{it-1} = \beta_1(y_{it-1} - y_{it-2}) + \beta_2(x_{it} - x_{it-1}) + (\vartheta_{it} - \vartheta_{it-1})$

Once again, we are able to eliminate this a_i component. What is happening? So, in this case we can again clearly see that the error term contains ϑ_{it-1} and one of the explanatory variables contains y_{it-1}. So, once again one of the explanatory variables is getting correlated with the error term. So, this implies $y_{it-1} - y_{it-2}$is correlated with $(\vartheta_{it} - \vartheta_{it-1})$because the latter contains ϑ_{it-1} and when they are correlated. That means it leads to endogeneity problem. That means once again OLS is not applicable. If we apply OLS in this context that will lead to biased estimates of β_1 and β_2 This bias is of order 1 by t.

So, in a dynamic panel data model neither fixed effect transformation nor FD is applicable because one of your explanatory variables is getting correlated with the error term which prevents the application of OLS in this transformed model which we were doing earlier in the context of a simple one-way fixed effect model. If we do estimate a dynamic panel data model the way we learned in the context of FE and FD that will lead to a bias and that bias is of order 1 by t.

This bias is also known as Nickell bias because Nickell, the famous econometrician was the one who introduced this type of bias in the literature. So, in a panel where t is small and n is large that means basically it is a micro panel. So, you have there are two types of panel - micro panel and a macro panel. In micro panel, basically you have large number of n but small t and macro panel has basically large T and small n. If your panel is of this type where n is large but T is small you can understand that bias will be huge. But for a macro panel where T is large and n is small, for example when you are constructing a panel with let us say only 4 or 5 countries and you have large number of time period let us say 50 years data then this bias tends to 0 as t tends to infinite. So, famous econometrician Nickell first introduced the problem- this type of a bias that is why it is called Nickell bias of order 1 by T which occurs if we apply fixed effect transformation and FD to estimate the dynamic panel data model. So, neither fixed effect nor FD is applicable. You can apply FD only when you have a large number of T. But even when T equals 30, the bias is 20 percent. So, that means the estimates $\widehat{\beta_1}$ and $\widehat{\beta_2}$ will deviate 20 percent from their true value. So, neither FE nor FD is basically applicable because they are leading to a bias which is called Nickell bias.

So, in a dynamic panel data model where we introduced a lag dependent variable in the system basically calls for this type of two types of persistence over a period of time. One is due to the unobserved effect and another is due to the autocorrelation. So, lag dependent variable creates autocorrelation in the problem. Naturally the question comes to our mind can we eliminate a i by fixed effect transformation and get rid of at least one type of persistence? We tried that and we showed that no it is not neither FE nor FD is basically applicable to estimate the dynamic panel data model. But one thing we have not discussed is what exactly is the context of a dynamic panel? So, that means what are some examples of dynamic panel data model. So, this is examples of DPD- dynamic panel data model.

First one is we can say that agricultural production. What happens here a farmer decides which crop to cultivate in period t depending on what was the return from that a particular crop last year. Let us assume that last year the farmer earned huge profit from paddy cultivation. So, if paddy cultivation helps the farmer earning huge profit in t – 1 then in tth period also the farmer will go for paddy cultivation and the how much acre of land the farmer will allocate for paddy cultivation that also depends on what was the relative return of paddy over other competing crops.

If paddy gave him or her huge profit last year even this year also farmer will go for paddy cultivation relatively more compared to other commodities. Even in the context of industrial production let us say last year a firm produces the amount y_{it} or y_{it-1}but the firm could not sell that much and lot of the produced item went to inventory. When a large portion of the output or production went to inventory in the previous period obviously the firm will try to reduce the production of tth period because that much inventory the firm has to sell fast. So, that means a higher y_{it-1}basically indicates a relatively lower amount of y_{it}. That is why

$y_{it}\ is\ related\ to\ y_{it-1}$. So, both agricultural production industrial production shows how basically y_{it} is depending on y_{it-1}.

Another interesting example is consumption of cigarette let us say. So, this type of products cigarette, alcohol if you consume then it builds some kind of habit within the customer. So, at the initial period an individual will start smoking 1, 2, 3 and each successive period because of this habit the consumer will try to maintain that type of that level of smoking. So, how many, cigarette oil smoke in period t that means that very well determined by how much you have consumed in t – 1th period, because t – 1th period consumption shows what is the habit that I have created within myself? How much alcohol I will consume in t^{th} period that again is determined by t – 1th period consumption of alcohol. So, this habit formation will go on like this. Then at one point of time if the consumer thinks that he or she will leave the smoking habit they cannot do it overnight. So, if in my t^{th} period observation t^{th} period consumption is let us say 5 packets of cigarette probably in t plus 1th period I will reduce by half packet 4 and half, next period 4, next period 3 and half in this way.

So, that means in the process of habit building, when it is increasing in that section also we showed that t^{th} period consumption is basically dependent on t – 1th period consumption. At the same time if the consumer is thinking of leaving that habit the consumer cannot leave it all of a sudden, the reduction will also become a slow process. So, all these examples show how y it that means the variable y in t^{th} period is basically dependent on its previous values.

These are some examples of dynamic panel data model, how dynamism is required. There are many such examples you can think of from your real life. Most of the economic relationships are dynamic in nature. That means lag dependent variable we must introduce in the model as one of our explanatory variables. That explains the current period value of y dependent variable.