1 Recitation Six – Intro to Econometrics

2 Topic One: What are panel data? Why use them?

Types of Data:

- Panel data are data where we can follow multiple *entities* across multiple *time periods.*
- Cross-section data are data where we observe multiple *entities* at one time period.
- Time-series data are data where we observe one *entity* across multiple *time periods.*

What is special about panel data?
Recall our old "assumed" true model:

$$y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

where our "parameter of interest" in β_1 (i.e. what we care about is the effect of X_i on y_i) and ϵ_i is the unobserved effects.
Writing this in panel data form:

$$y_{it} = \beta_0 + \beta_1 X_{it} + \epsilon_{it}$$

Now, our estimated coefficients from running this regression are unbiased and consistent under the 4 LSAs. Does this satisfy LSA #1? It matters what is in the ϵ_{it} term! If there are variables that are in ϵ_{it} that correlate with X_{it} and have a direct effect on y_{it}, $\hat{\beta}_1$ will suffer from omitted variable bias. Usually, this is where we stop and say "oh, well". In panel data, we have a method to remove some of this omitted variable bias. How?

We can split up the error term into its component parts:

$$\epsilon_{it} = \alpha_i + \gamma_t + u_{it}$$

where α_i is the unexplained component of y that is specific to an entity, γ_t is unexplained component of y that is specific to a time period, and u_{it} is a mean 0 random variable. These α and γ variables are a combination of all of the omitted variables that effect y and are specific to entities or time periods. So, rewriting our "true" equation:

$$y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + \gamma_t + u_{it}$$

Using our panel data techniques, we can either estimate or remove the α and γ, so now all we have to worry about for OVB are variables left in u_{it}, which is a much smaller set of variables than those in ϵ_{it}. α_i (i.e. *entity fixed effects*) accounts for variables that change over entity, but not over time. γ_t (i.e. *time fixed effects*) accounts for variables that change over time, but not over entity. Variables unaccounted for (left in u_{it}) are only those that change both over time and over entities.

3 Topic Two: Types of Models

3.1 First Differencing

First differencing is a way to get rid of the "time fixed effects" and "entity fixed effects". It does so by canceling the entity fixed effects and estimating the change in time fixed effects are estimated by the constant. Say our true model is one with entity and time fixed effects. That is,

$$y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + \gamma_t + u_{it}$$

Our "true" first difference model is then:

$$y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + \gamma_t + u_{it}$$

$$-(y_{it-1} = \beta_0 + \beta_1 X_{it-1} + \alpha_i + \gamma_{t-1} + u_{it-1})$$

$$\Delta y_{it} = \beta_1 \Delta X_{it} + \Delta\gamma_t + \Delta u_{it}$$

So, under two assumptions (and the standard LSAs): 1) $E[\Delta u_{it}|\Delta X_{it-1}] = 0$ and 2) $\Delta\gamma_t$ is constant over time, running a regression of Δy_{it} on ΔX_{it} will give us unbiased and consistent estimates of β_1. $\Delta\gamma_t$ will be estimated by the intercept. Note that the first assumptions is true if our original LSA1 was true and that our second assumption is always true when T=2.

3.2 Dummy Variables

One way to estimate α_i and γ_t is to include dummy variables for each of their possible values. That is:

$$\alpha_i = \alpha_1 D_{i=1,t} + \alpha_2 D_{i=2,t} + \alpha_3 D_{i=3,t} + ...$$

and

$$\gamma_t = \gamma_1 D_{i,t=1} + \gamma_2 D_{i,t=2} + \gamma_3 D_{i,t=3} + ...$$

So, to estimate the true fixed effects model, we can plug in this dummy representation of the fixed effects and estimate using OLS!

$$y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_1 D_{i=1,t} + \alpha_2 D_{i=2,t} + \alpha_3 D_{i=3,t} + ... + \gamma_1 D_{i,t=1} + \gamma_2 D_{i,t=2} + \gamma_3 D_{i,t=3} + ... + u_{it}$$

Now, we can't run exactly this model because of multicollinearity between β_0 and the dummy variables, so we must fix this by the following:
If only one type of fixed effects used:
I)1) Exclude intercept.
OR I)2) Exclude one dummy.
If both types of fixed effects used:
II)1) Exclude intercept and one dummy.
OR II)2) Exclude one entity dummy and one time dummy.

Interpretations of estimates are slightly different in each of these implementations. Say I=3, T=3 for this example.

I)1)$y_{it} = \beta_1 X_{it} + \alpha_1 D_{i=1,t} + \alpha_2 D_{i=2,t} + \alpha_3 D_{i=3,t} + u_{it}$

$$\hat{\alpha_i} \to \beta_0 + \alpha_i$$

That is, $\hat{\alpha_i}$ is an estimate of the intercept specific to entity i – the expected value of y if X=0 and . In other words, an estimate of the overall intercept plus the expected increase in y coming from unobservables specific to entity i.

I)2)$y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_2 D_{i=2,t} + \alpha_3 D_{i=3,t} + u_{it}$

$$\hat{\alpha_i} \to \alpha_i - \alpha_1$$

($D_{1,t}$ was left out) That is, $\hat{\alpha_i}$ is an estimate of the expected difference in y between entity i and the "left out entity," all else equal.

Similar interpretations are used for models where both time and entity fixed effects are used.

3.3 Fixed Effects/Demeaning

Instead of estimating the α_i and γ_t, we can remove them by demeaning.

I show the case where we only have entity fixed effects here, but the derivation of why this works is below in the equivalencies section. And note that there is a entity and time demeaning process that can be carried out – it is slightly more complicated however.

Take our assumed "true" model:

$$y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + u_{it}$$

In this true model, note that we can define the mean of the entities over time over both sides:

$$\frac{1}{T}\sum_{t=1}^{T} y_{it} = \frac{1}{T}\sum_{t=1}^{T}(\beta_0 + \beta_1 X_{it} + \alpha_i + u_{it})$$

Equivalently:

$$\frac{1}{T}\sum_{t=1}^{T} y_{it} = \beta_0 + \beta_1 \frac{1}{T}\sum_{t=1}^{T} X_{it} + \frac{1}{T}\sum_{t=1}^{T} \alpha_i + \frac{1}{T}\sum_{t=1}^{T} u_{it})$$

Which we write as:

$$\bar{y}_{i\cdot} = \beta_0 + \beta_1 \bar{X}_{i\cdot} + \bar{\alpha}_{i\cdot} + \bar{u}_{i\cdot}$$

Now, note that since α_i is the same for every time period, $\bar{\alpha}_{i\cdot} = \alpha_i$.

Take the true model and substracting the meaned equation, we find that:

$$y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + u_{it}$$
$$-(\bar{y}_{i\cdot} = \beta_0 + \beta_1 \bar{X}_{i\cdot} + \alpha_i + \bar{u}_{i\cdot})$$

$$y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + u_{it}$$
$$-(\bar{y}_{i\cdot} = \beta_0 + \beta_1 \bar{X}_{i\cdot} + \alpha_i + \bar{u}_{i\cdot})$$
$$(y_{it} - \bar{y}_{i\cdot}) = \beta_1 (\bar{X}_{i\cdot} - \bar{X}_{i\cdot}) + \bar{u}_{i\cdot})$$

Note that we removed the α_i from the equation. Therefore, if the 4 OLS assumptions apply to this demeaned model, then our estimates of betas using the demeaned model are unbiased and consistent.

4 Topic 3: Equivalencies

1) In the Special Case where T=2, First Differencing is equivalent to a Time Fixed Effects model.

2) The demeaning and dummy variable models are ALWAYS equivalent. The way I like to think this is that the demeaning method estimates the α and γ parameters first, then plugs them back into the model before running OLS. The dummy variable model does this all in one step. Note: this is why the demeaning is much faster for a computer to estimate.

To make things slightly simpler, I will show this for a model with only entity fixed effects. This method can be easily generalized to models with both fixed effects.

Our model is

$$y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + u_{it}$$

With this structure, it seems obvious why we could estimate using a dummy variable method. From above, our interpretation of α_i is:

$$\alpha_i = E[y_{jt} - \beta_0 - \beta_1 X_{jt} - u_{jt} | j = i]$$

A natural estimator for this is:

$$\hat{\alpha}_i = \sum_{j=1}^{I} 1\{j = i\}[\frac{1}{T}\sum_{t=1}^{T} y_{jt} - \beta_0 - \beta_1 \frac{1}{T}\sum_{t=1}^{T} X_{jt}]$$

where $1\{j = i\}$ equals one if j=i and zero otherwise. Plugging this in to our original equation:

$$y_{it} = \beta_0 + \beta_1 X_{it} + \sum_{j=1}^{I} 1\{j = i\}[\frac{1}{T}\sum_{t=1}^{T} y_{jt} - \beta_0 - \beta_1 \frac{1}{T}\sum_{t=1}^{T} X_{jt}] + u_{it}$$

Doing a tiny bit of algebra:

$$y_{it} - \sum_{j=1}^{I} 1\{j = i\} \frac{1}{T} \sum_{t=1}^{T} y_{jt} = \beta_1 [X_{it} - \sum_{j=1}^{I} 1\{j = i\} \frac{1}{T} \sum_{t=1}^{T} X_{jt}] + u_{it}$$

Which is our entity demeaned model!