Panel Data Methods

$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + u_{it}$

1. Basic

Ch.13 Pooling & Panel Data Methods

1. Pooling Independent Cross Sections across Time
2. Policy Analysis with Pooled Cross Sections
3. Two-Period Panel Data Analysis
4. Policy Analysis with Two-Period Panel Data*
5. Differencing with More than Two Time Periods

Review Structure of economic data

- Pooled cross section
  - It can pool random cross sections to account for time differences –e.g. table 1.4.
- Panel data
  - It consists of the ***same*** random individual observations over time –e.g. table 1.5.

13.1 Pooled Cross Sections

Reason for using pooled cross sections is

- To get bigger sample sizes
- To investigate the effect of time
  - See example 13.1
- To investigate whether relationships have changed over time
  - Interaction term
  - See example 13.2
  - The Chow test for structural change

13.2 Policy Analysis with Pooled

- A ***natural experiment*** always has
  - a ***control group***, which is not affected by the policy change, and
  - a ***treatment group***, which is affected by the policy change.
- One can then simply compare the change in outcomes across the treatment and control groups to estimate the policy effect.

Difference-in-Differences

- The equation of interest is

  $y = \beta_0 + \delta_0 d2 + \beta_1 dT + \delta_1 d2dT + other\ factors$ (13.10)
  - where *d2* denotes a dummy variable for post-policy change, and
  - *dT* is a dummy variable for treatment group.
- ***Difference-in-Differences Estimator***

  $\hat{\delta}_1 = (\bar{y}_{2,T} - \bar{y}_{2,C}) - (\bar{y}_{1,T} - \bar{y}_{1,C})$ (13.11)
  - see example 13.3 & 13.4

13.3 Two-Period Panel Data

- Panel data can be used to address some kinds of omitted variable bias.
  - If we can think of the omitted variables as being fixed over time, then we can model as having a ***composite error***.
- Suppose the population model is

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + a_i + u_{it} \quad (13.13)$$

  - a_i captures all unobserved, time-constant factors that affect y_{it}.

Econometrics 7

Unobserved Fixed Effects

- ***Composite error***, $v_{it} = a_i + u_{it}$
  - If a_i is correlated with the x's, OLS will be biased, since a_i is part of the error term.
- In estimation with panel data, we can difference-out the unobserved fixed effect

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i \quad (13.17)$$

  - This model has no correlation between the x's and the error term, so no bias.
  - Δx_i must have some variation across i.

Econometrics 8

13.5 Differencing with Multiple Periods

- A general fixed effect model (T=3)

$$y_{it} = \delta_1 + \delta_2 d2_t + \delta_3 d3_t + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + a_i + u_{it} \quad (13.28)$$

- We can simply estimate by OLS, assuming the Δu_{it} are uncorrelated over time.

$$\Delta y_i = \delta_2 \Delta d2_t + \delta_3 \Delta d3_t + \beta_1 \Delta x_{i1} + \ldots + \beta_k \Delta x_{ik} + \Delta u_i$$

  - So if 3 periods, then subtract period 1 from period 2, period 2 from period 3 and have 2 observations per individual.

Econometrics 9