Simple Panel Data Methods

1. Two-Period Panel Data Analysis
2. Differencing with More than Two Time Periods

Two-Period Panel Data Analysis

Same units observed over 2 time pds.

Intercept for pd. 1 $= \beta_0$

- Model

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + v_{it}$$

  - i : person, firm, city, etc. and t : time period
  - $d2$: dummy var. for pd. 2 — 1 for pd. 2, 0 for pd. 1

Intercept for pd. 2 $= \beta_0 + \delta_0$

- Example

$$crime_{it} = \beta_0 + \delta_0 d2_t + \beta_1 unem_{it} + v_{it}$$
$$prod_{it} = \beta_0 + \delta_0 d2_t + \beta_1 expo_{it} + v_{it}$$

Two-Period Panel Data Analysis (cont.)

- Suppose

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + a_i + u_{it}$$

  - a_i : unobserved effect ; fixed effect ; unobs. heterogen- -eity
  - u_{it} : idiosyncratic error
  - v_{it} : time-varying error $= a_i + u_{it}$: composite error

- Example

$$crime_{it} = \beta_0 + \delta_0 d2_t + \beta_1 unem_{it} + city_i + u_{it}$$
$$prod_{it} = \beta_0 + \delta_0 d2_t + \beta_1 expo_{it} + mqual_i + u_{it}$$

Two-Period Panel Data Analysis (cont.)

- Estimating β_1

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + a_i + u_{it}$$

- Pooling the two years and performing OLS : may not be unbiased if e.g. a_i and x_{it} are correlated
- One solution: difference the data

Two-Period Panel Data Analysis (cont.)

- Two years

$$y_{i2} = (\beta_0 + \delta_0) + \beta_1 x_{i2} + a_i + u_{i2}$$
$$y_{i1} = \beta_0 + \beta_1 x_{i1} + a_i + u_{i1}$$

- Subtracting

$$y_{i2} - y_{i1} = \delta_0 + \beta_1 (x_{i2} - x_{i1}) + u_{i2} - u_{i1}$$

β_1 : first-differenced estimator

- The *first-differenced equation*

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i$$

- Example

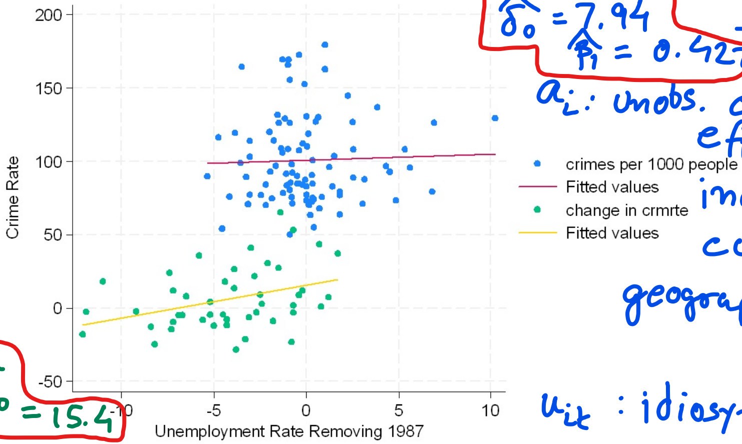
$$\Delta crime_i = \delta_0 + \beta_1 \Delta unem_i + \Delta u_i$$
$$\Delta prod_i = \delta_0 + \beta_1 \Delta expo_i + \Delta u_i$$

Two-Period Panel Data Analysis (cont.)

$crmrte_{it} = \beta_0 + \delta_0 d87_t + \beta_1 unem_{it} + a_i + u_{it}$

$\boxed{\hat{\delta}_0 = 7.94 \quad \hat{\beta}_1 = 0.427}$

a_i: unobs. city effect → industry composition, geography, etc.



$\boxed{\hat{\delta}_0 = 15.4}$

u_{it}: idiosyncratic errors e.g. weather shocks, protests/activism

$\Delta crmrte_i = \delta_0 + \beta_1 \Delta unem_i + \Delta u_i$

$\boxed{\hat{\beta}_1 = 2.22}$

Two-Period Panel Data Analysis (cont.)

$(u_{i2} - u_{i1})$ should be uncorrelated with $(x_{i2} - x_{i1})$; u_i should be uncorr. with x_i from both time pds.

- Note
  - Still need Δu_i to be uncorrelated with Δx_i
  - The *strict exogeneity* assumption
  - Need variation in Δx_i

u and x are uncorr. in same time pd. $\Rightarrow$ contemporaneous exogeneity

u & x uncorr. across all time pds. $\rightarrow$ strict exogeneity

x : strictly exogenous

Differencing with More than Two Time Periods

$d2$: dummy var. for pd. 2
$d3$: " " for pd. 3

- Model

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

  - i : individual units and $t = 1, 2,$ and 3

If a_i corr. w/ x_{itj} → (pooled) OLS ⇒ biased estimators

strict exogeneity ⇒ $\text{corr}(x_{itj}, u_{is}) = 0$ for all t, s, j

Differencing with More than Two Time Periods (cont.)

- Three years

t=3
t=2
t=1

$$y_{i3} = \delta_1 + \delta_3 + \beta_1 x_{i31} + \ldots + \beta_k x_{i3k} + a_i + u_{i3}$$
$$y_{i2} = \delta_1 + \delta_2 + \beta_1 x_{i21} + \ldots + \beta_k x_{i2k} + a_i + u_{i2}$$
$$y_{i1} = \delta_1 + \beta_1 x_{i11} + \ldots + \beta_k x_{i1k} + a_i + u_{i1}$$

- Subtracting

for t=3

$$y_{i3} - y_{i2} = \delta_3 - \delta_2 + \beta_1 (x_{i31} - x_{i21}) + \ldots + \beta_k (x_{i3k} - x_{i2k}) + u_{i3} - u_{i2}$$

for t=2

$$y_{i2} - y_{i1} = \delta_2 + \beta_1 (x_{i21} - x_{i11}) + \ldots + \beta_k (x_{i2k} - x_{i1k}) + u_{i2} - u_{i1}$$

for t=1 : all missing values

Differencing with More than Two Time Periods (cont.)

- More generally

$$\Delta y_{it} = \delta_2 \Delta d2_t + \delta_3 \Delta d3_t + \beta_1 \Delta x_{it1} + ... + \beta_k \Delta x_{itk} + \Delta u_{it}$$

- Alternatively, obtain the same β_j estimates from

$$\Delta y_{it} = \alpha_0 + \alpha_3 d3_t + \beta_1 \Delta x_{it1} + ... + \beta_k \Delta x_{itk} + \Delta u_{it}$$

Differencing with More than Two Time Periods (cont.)

- For $T > 3$

$$\Delta y_{it} = \delta_2 \Delta d2_t + ... + \delta_T \Delta dT_t + \beta_1 \Delta x_{it1} + ... + \beta_k \Delta x_{itk} + \Delta u_{it}$$

- Alternatively, obtain the same β_j estimates from

$$\Delta y_{it} = \alpha_0 + \alpha_3 d3_t + ... + \alpha_T dT_t + \beta_1 \Delta x_{it1} + ... + \beta_k \Delta x_{itk} + \Delta u_{it}$$

Differencing with More than Two Time Periods (cont.)

cluster-robust std. err. : allows heterosk. & arbitrary correl^n within a cross-sectional unit (i.e. cluster) over time but not across cross-sectional units.

- Standard errors
  - For usual standard errors to be valid Δu_{it} should be uncorrelated over time
  - Can test for such correlation
  - Regardless of such correlation or heteroskedasticity $\rightarrow$ with large N & small T cluster-robust std. errors are appropriate.