Simple Panel Data Methods

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

Two-Period Panel Data Analysis

Int. for pd. 1 : β_0

" " pd. 2 : $\beta_0 + \delta_0$

- 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 for pd. 2; 1 → pd. 2, 0 → pd. 1

- 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.)

$v_{it} = a_i + u_{it}$ (composite error)

- Suppose

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

  - a_i : unobserved effect / fixed effect / unobs. heterogeneity
  - u_{it} : idiosyncratic / time-varying error
  - v_{it} :
- Example

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

e.g. infrastr., geog., area etc.

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 work
- One solution:

difference the data

e.g. if a_i and x_{it} are correlated

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

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

- The *first-differenced equation*

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i$$ (for t=2)

- Example

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

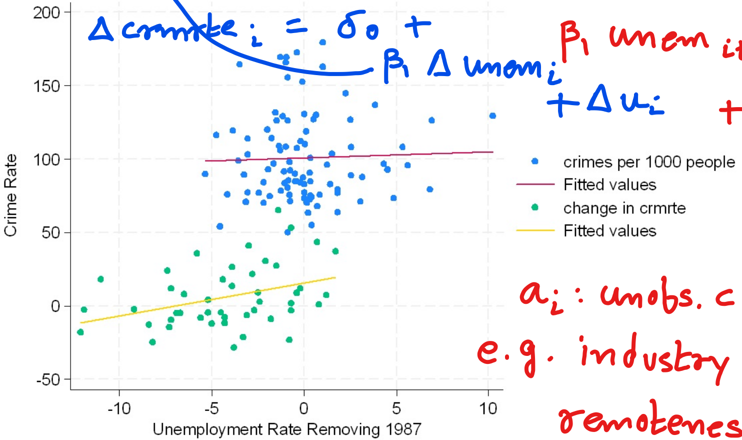
Two-Period Panel Data Analysis (cont.)

CRIME2

$\hat{\beta}_1 = 0.427$

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

$\Delta crmrte_i = \delta_0 + \beta_1 \Delta unem_i + \Delta u_i$ → $\hat{\beta}_1 = 2.22$



a_i: unobs. city effect e.g. industry compos^n, remoteness, geography

u_{it}: idiosyncratic errors such as weather, protests/activism

Two-Period Panel Data Analysis (cont.)

$(u_{i2} - u_{i1})$ uncorr. w/ $(x_{i2} - x_{i1})$

u_i should be " w x_i from all time pds.

- Note
  - Still need Δu_i to be uncorrelated with Δx_i → for unbiasedness/consistency
  - The *strict exogeneity* assumption ↓ strict exogeneity
  - Need variation in Δx_i across i — controls for all time-inv. factors

strict exog.

u & x uncorr. across all pds. → u_{i2} uncorr. w/ x_{i2} & x_{i1}; u_{i1} " " x_{i2} & x_{i1}

u & x uncorr. in same pd. → u_{i2} uncorr. with x_{i2}

x_{i2} → contemporaneous exogeneity

Differencing with More than Two Time Periods

d2 : dummy for t = 2
d3 : " " t = 3

- Model

If a_i corr. w/ x_{itj} ⇒ OLS → biased estimator

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

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

strict exogeneity

$corr(x_{itj}, u_{is}) = 0$ for all t, s, j

Differencing with More than Two Time Periods (cont.)

- Three years

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

- Subtracting

t=3

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

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

t=2

Differencing with More than Two Time Periods (cont.)

for t = 3:

$\Delta d2_t = -1$

$\Delta d3_t = 1$

for t = 2 & t = 3

for t = 2:

$\Delta d2_t = 1$

$\Delta d3_t = 0$

- More generally

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

- Alternatively, obtain the same β_j estimates from

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

Differencing with More than Two Time Periods (cont.)

CRIME4:

$\Delta \log(crmrte_{it}) = \alpha_0 + \alpha_3 d83 + \ldots + \alpha_T d87$

$\rightarrow t = 2, 3, \ldots, T$

- For $T > 3$

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

- Alternatively, obtain the same β_j estimates from

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

$+ \beta_1 \Delta \log(avgsen_{it}) + \beta_2 \Delta \log(polpc_{it})$

$+ \Delta u_{it}$

$\hat{\beta}_1 = -0.066$

an ↑ in avgsen by 1% ↓ crmrte ↓ 0.066%

Differencing with More than Two Time Periods (cont.)

u_{i1} u_{j1}
u_{i2} u_{j2}
$\vdots$ $\vdots$
u_{iT} u_{jT}

- 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 : with large N and small T cluster-robust std. errors are appropriate.

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