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 pds.

intercept pd. 1 : β_0
" 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 → 0 for pd. 1, → 1 " 2

- 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. heterogeneity
  - u_{it} : time-varying error / idiosyncratic error
  - v_{it} : $a_i + u_{it}$: 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 "work" e.g. if a_i & x_{it} are correlated
- One solution:

difference the data

x_{it} endog. ⇒ bias

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*

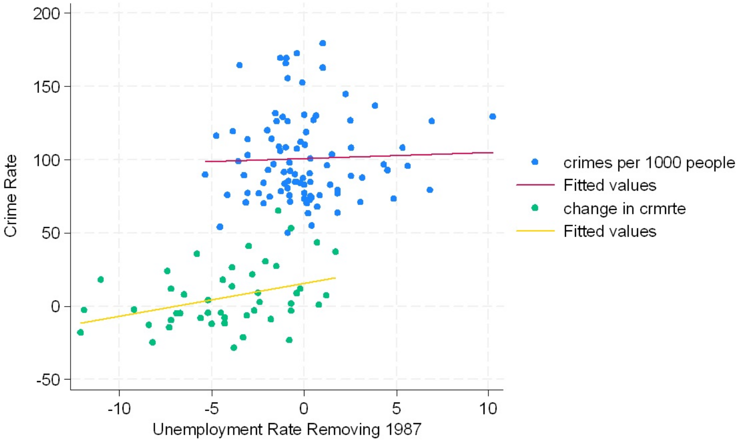
first-differenced estimator

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

- 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.)



Two-Period Panel Data Analysis (cont.)

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

u & x need to be strictly exogenous.

" uncorr. across all time pds.

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

$\Rightarrow$ strict exogeneity

(x_{i2}) x_{i1} must be uncorr. w/ u_{i2} (u_{i1})

u & x uncorr. in just the same pd. $\Rightarrow$

u_{i2} " w/ x_{i2}

u_{i1} " w/ x_{i1} $\Rightarrow$ contemporaneous exog.

Differencing with More than Two Time Periods

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

- Model

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

time-invariant

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

If a_i corr. with x_{it1} → OLS ⇒ biased estimator

↳ x_1 for obs. i. in pd. 2

strict exog. ⇒ $\text{corr}(x_{itj}, u_{is}) = 0$ 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

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 $t=2$

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

Differencing with More than Two Time Periods (cont.)

For $t=2$ $\quad \Delta d2_t = 1$ & $\Delta d3_t = 0$.

$t = 3$ $\quad \Delta d2_t = -1$ & $\Delta d3_t = 1$

- 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.)

* allow heterosk. & arbitrary correl^n within a cross-sectional unit (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 → with large N (no. of units) and small T (no. of time pds.) cluster-robust std. errors* are appropriate.