

# 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.

- 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

- 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
- ▶  $u_{it}$  : idiosyncratic error
- ▶  $v_{it}$  :

- Example

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

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

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

## Two-Period Panel Data Analysis (cont.)

not work  $\Rightarrow$  biased estimates

- Estimating  $\beta_1$

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

- Pooling the two years and performing OLS  $\rightarrow$  may not work e.g. if  $a_i$  and  $x_{it}$  are correl.
- One solution: difference the data

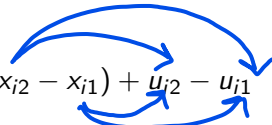
## Two-Period Panel Data Analysis (cont.)

- 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


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

- The *first-differenced* equation

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

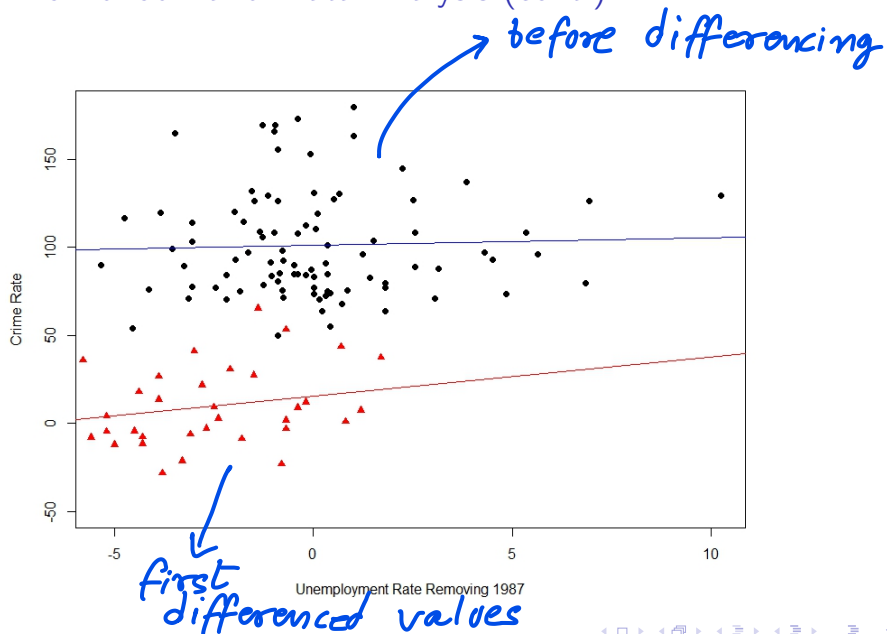
first-differenced  
estimate

- 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.)

$$(u_{i2} - u_{i1})$$

$$(x_{i2} - x_{i1})$$

$x_i$  should be uncorrelated with  $u_i$  from both periods

- Note

- ▶ Still need  $\Delta u_i$  to be uncorrelated with  $\Delta x_i$
- ▶ The *strict exogeneity* assumption
- ▶ Need variation in  $\Delta x_i$

contemporaneous

e.g.  $x \rightarrow \text{unem. rt.}$

$u \rightarrow \text{weather shock}$

exog.  $\Rightarrow x_i$  uncorr. with  $u_i$  in same pd.

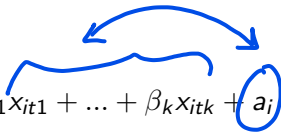
if past shocks affect

current empl.  $\Rightarrow$  strict exog. violated

# Differencing with More than Two Time Periods

OLS  
⇒ biased estimator  
correlated

- Model

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


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

-

## Differencing with More than Two Time Periods (cont.)

- Three years

$$y_{i3} = \delta_1 + \delta_3 + \beta_1 x_{i31} + \dots + \beta_k x_{i3k} + a_i + u_{i3}$$

$$y_{i2} = \delta_1 + \delta_2 + \beta_1 x_{i21} + \dots + \beta_k x_{i2k} + a_i + u_{i2}$$

$$y_{i1} = \delta_1 + \beta_1 x_{i11} + \dots + \beta_k x_{i1k} + a_i + u_{i1}$$

- Subtracting

For  $i$  (e.g. city, indiv., etc.)

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

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

strict exog. for  $T > 2 \Rightarrow x$  in any  
pd. is uncorr. w/  $u$  from all time  
pds.

## Differencing with More than Two Time Periods (cont.)

- More generally *Not my favorite!*

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

- Alternatively, obtain the same  $\beta_j$  estimates from

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

↓  
have intercept  
Time dummies from pd. 3 onwards

## Differencing with More than Two Time Periods (cont.)

- For  $T > 3$

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

- Alternatively, obtain the same  $\beta_j$  estimates from

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

## Differencing with More than Two Time Periods (cont.)

Large  $N$ , Small  $T$  : e.g. 1000 indiv. over 10 time pds.      Small no. of time pds. ( $T$ )

↳ not case

e.g. 25 countries

over 10 time pds.

- Standard errors

- ▶ For usual standard errors to be valid  $\Delta u_t$  should be uncorrelated over time
- ▶ Can test for such correlation
- ▶ Regardless of such correlation or heteroskedasticity → with

large  $N$  & small  $T$  cluster robust standard errors are appropriate

large no. of units observed ( $i$ )