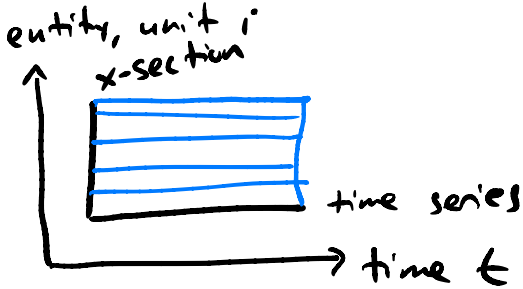
Lecture Note 8: Panel Data

Y_{it} ← panel data
← longitudinal data
← ordered



Y_{ij} ← unordered

$$Y_{ij} = \alpha + X_{ij}'\beta + Z_i'\gamma + U_{ij} \quad \text{for } i = 1, \cdots, N \text{ and } j = 1, \cdots, J$$

balanced panel

$$U_{ij} = \delta_i + \varepsilon_{ij}$$

family effect

individual error

OLS w/ clustering
random effects

$$Y_{ij} = \alpha + X_{ij}'\beta + Z_i'\gamma + \delta_i + \varepsilon_{ij}$$ ← error components model

Problems

① error correlation: $\text{cov}(U_{ij}, U_{ik}) = \text{cov}(\delta_i + \varepsilon_{ij}, \delta_i + \varepsilon_{ik}) = V[\delta_i] > 0$

② δ_i may be correlated with X_{ij} and Z_i

fixed effects

Fixed effects

→ "Brute force"

- dummy vars $D_i = \begin{cases} 1 & \text{if observation in } i \\ 0 & \text{otherwise} \end{cases}$
- then control for D's in regression:

$$Y_{ij} = \alpha + X_{ij}'\beta + \sum_{i=2}^{N} \lambda_i D_i + \varepsilon_{ij}$$

→ "Finesse"

- start with
- "between"
- "within" demeaned

$$Y_{ij} = \cancel{\alpha} + X_{ij}'\beta + \cancel{Z_i'\gamma} + \cancel{\delta_i} + \varepsilon_{ij}$$

$$-\left(\bar{Y}_i = \cancel{\alpha} + \bar{X}_i'\beta + \cancel{Z_i'\gamma} + \cancel{\delta_i} + 0\right)$$

$$Y_{ij} - \bar{Y}_i = (X_{ij} - \bar{X}_i)'\beta + \varepsilon_{ij}$$

same

→ In R: brute force → feols(y ~ x + factor(groupvar), data=df)

finesse → feols(y ~ x | groupvar, data=df)

Time in Panel Data

$ij \rightarrow it$

More general model: $Y_{it} = \alpha + X_{it}'\beta + \underline{Z_i'\gamma} + \underline{W_t'\lambda} + \underbrace{\delta_i + \tau_t + \varepsilon_{it}}_{U_{it}}$

Two-way fixed effect: $Y_{it} = \alpha + X_{it}'\beta + \delta_i + \tau_t + \varepsilon_{it}$

δ_i: unit FE ← dummies for units

τ_t: time FE ← dummies for time

In R:

```
feols(y ~ x + factor(group) + factor(time), data = df)
feols(y ~ x | group + time, data = df)
```

First differencing

$$Y_{it} = \cancel{\alpha} + X_{it}'\beta + \cancel{\delta_i} + \tau_t + \varepsilon_{it}$$

$$-\left(Y_{i,t-1} = \cancel{\alpha} + X_{i,t-1}'\beta + \cancel{\delta_i} + \tau_{t-1} + \varepsilon_{i,t-1}\right)$$

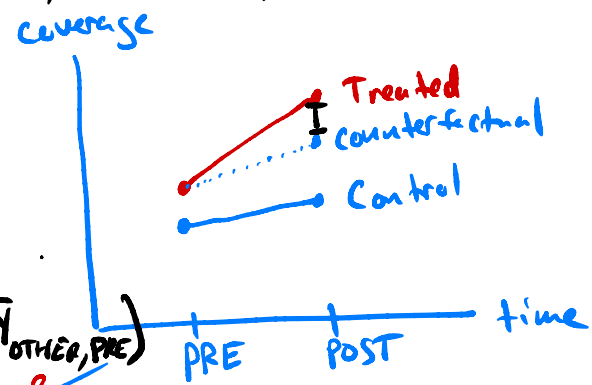
$$Y_{it} - Y_{i,t-1} = (X_{it} - X_{i,t-1})'\beta + (\tau_t - \tau_{t-1}) + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

$$\Delta Y_{it} = \Delta X_{it}'\beta + \underbrace{\Delta \tau_t}_{\text{new year FE}} + \underbrace{\Delta \varepsilon_{it}}_{\text{new error}}$$

In R: `feols(d(y) ~ d(x) | year, data = df, panel.id = ~state + year)`

Diff-in-diff

Y_{ist} - insurance coverage for person i, state s, year t



$$\Delta_s = \bar{Y}_{MA,POST} - \bar{Y}_{OTHER,POST}$$

$$\Delta_t = \bar{Y}_{MA,POST} - \bar{Y}_{MA,PRE}$$

$$\Delta\Delta = (\bar{Y}_{MA,POST} - \bar{Y}_{MA,PRE}) - (\bar{Y}_{OTHER,POST} - \bar{Y}_{OTHER,PRE})$$

$$\cancel{\beta_0} + \cancel{\beta_1} + \cancel{\beta_2} + \beta_3 - (\cancel{\beta_0} + \cancel{\beta_2}) \qquad \cancel{\beta_0} + \cancel{\beta_1} - \cancel{\beta_0}$$

"parallel trends"

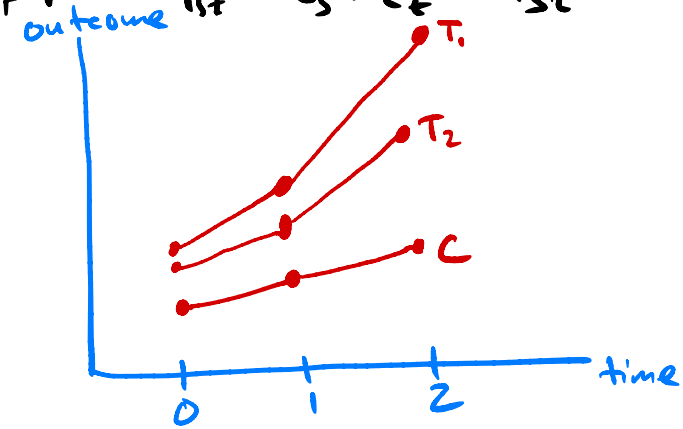
$$Y_{ist} = \beta_0 + \beta_1 POST_t + \beta_2 MA_s + \beta_3 POST_t \times MA_s + U_{ist}$$

$$\bar{Y}_{MA,POST} = \beta_0 + \beta_1 + \beta_2 + \beta_3, \quad \bar{Y}_{MA,PRE} = \beta_0 + \beta_2 \quad \bar{Y}_{OTHER,POST} = \beta_0 + \beta_1, \quad \bar{Y}_{OTHER,PRE} = \beta_0$$

$$\Delta\Delta = \beta_3$$

$$Y_{ist} = \underbrace{\tau_t + \delta_s}_{FE} + \beta POLICY_{st} + \varepsilon_{ist}$$

$$Y_{ist} = \underbrace{\tau_t + \delta_s}_{FE} + \beta POLICY_{st} + \varepsilon_{ist} = \alpha + \beta POLICY_{st} + \delta_s + \tau_t + \varepsilon_{ist}$$



Complications:

- staggered diff-in-diff
- clustered SEs