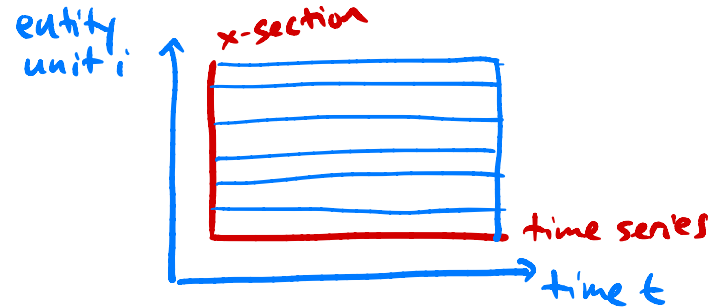
Lecture Note 8: Panel Data

Y_{it} ← longitudinal datset
← ordered panel data

Y_{ij} ← unordered panel data



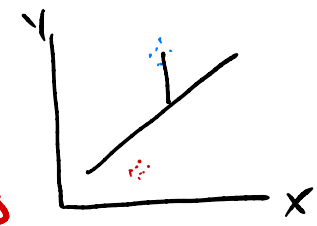
$$Y_{ij} = \alpha + X_{ij}'\beta + Z_i'\gamma + U_{ij} \qquad i=1,\cdots,N \qquad j=1,\cdots,J$$

balanced panel

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

δ_i: family component

ε_{ij}: individual component



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

OLS w/ clustering
Random effects

2 promblems:

1. error terms correlated within group: $cov(U_{ij}, U_{ik}) = cov(\delta_i + \varepsilon_{ij}, \delta_i + \varepsilon_{ik}) = V[\delta_i] > 0$
2. δ_i may be correlated with X_{ij} and Z_i

fixed effects

Fixed Effects

→ "Brute force"

→ Dummy variables for groups $D_i = \begin{cases} 1 & \text{if observation in } i \\ 0 & \text{otherwise} \end{cases}$

→ Use D_i's as control variables:

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

→ "Finesse" (demeaning)

→ start with: $Y_{ij} = \alpha + X_{ij}'\beta + Z_i'\gamma + \delta_i + \varepsilon_{ij}$

→ "between" : $-(\bar{Y}_i = \alpha + \bar{X}_i'\beta + Z_i'\gamma + \delta_i + 0)$

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

same

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

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

Time in Panel Data

$$Y_{it} = \alpha + X_{it}'\beta + Z_i'\gamma + W_t'\lambda + \delta_i + \tau_t + \varepsilon_{it}$$

Two-way fixed effect:

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

dummies for units → unit FE

time FE ← dummies for time periods

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

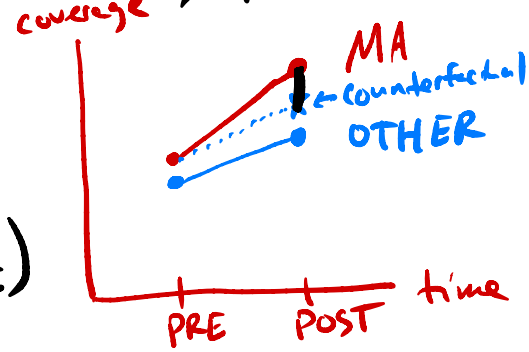
Y_{ist} = insurance dummy for individual i, state s, year t

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

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

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

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



"parallel trends"

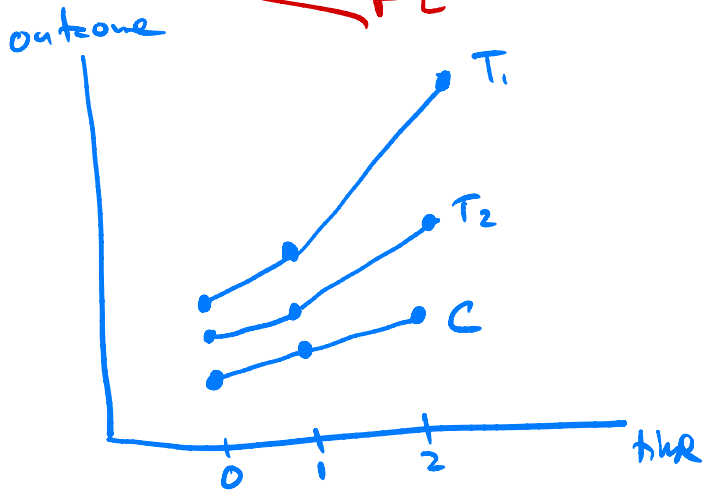
$$Y_{ist} = \beta_0 + \beta_1 MA_s + \beta_2 POST_t + \beta_3 MA_s \times POST_t + U_{ist}$$

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

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

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

FE

Complications:

- staggered diff-in-diff
- clustered SEs