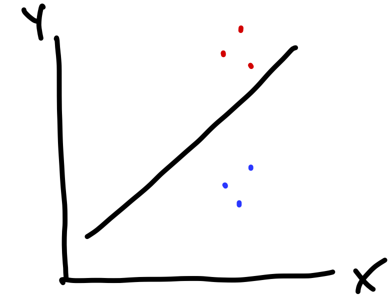
Lecture Note 8: Panel Data



Y_{it} ← longitudinal dataset

← "ordered" panel data

Y_{ij} ← "unordered" panel data

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

balanced panel

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

family component

individual component

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

OLS w/ clustered SEs

Random effects

Problems: ① errors correlated within i: $\mathrm{cov}(U_{ij}, U_{ik}) = V[\delta_i] > 0$

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

fixed effects →

Fixed effects

→ "Brute force"

→ Dummy for each i. $D_i = \begin{cases} 1 & \text{if obs. in } i \\ 0 & \text{o.w.} \end{cases}$

→ Use D_i's as covariates:

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

same

→ "Finesse" (demeaning)

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

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

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

Time in Panel Data

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

Two-way FE:

$$Y_{it} = \alpha + X_{it}'\beta + \underbrace{\delta_i}_{\text{unit FE}} + \underbrace{\tau_t}_{\text{time FE}} + \varepsilon_{it}$$

dummies for units → unit FE

time FE ← dummies for time periods

First diff:

$$
\begin{aligned}
& Y_{it} = \cancel{\alpha} + X_{it}'\beta + \cancel{\delta_i} + \tau_t + \varepsilon_{it} \\
- & \left(Y_{i,t-1} = \cancel{\alpha} + X_{i,t-1}'\beta + \cancel{\delta_i} + \tau_{t-1} + \varepsilon_{i,t-1}\right) \\
\hline
& \Delta Y_{it} = \Delta X_{it}'\beta + \underbrace{\Delta \tau_t}_{\text{new time FE}} + \underbrace{\Delta \varepsilon_{it}}_{\text{new error}}
\end{aligned}
$$

Diff-in-diff

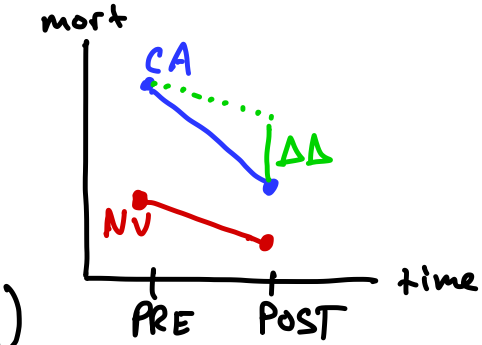
$$\Delta_s = Y_{CA,POST} - Y_{NV,POST}$$

$$\Delta_t = Y_{CA,POST} - Y_{CA,PRE}$$

$$\Delta\Delta = (Y_{CA,POST} - Y_{CA,PRE}) - (Y_{NV,POST} - Y_{NV,PRE})$$

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

$$Y = \beta_0 + \underbrace{\beta_1 CA}_{\text{state FE}} + \underbrace{\beta_2 POST}_{\text{year FE}} + \underbrace{\beta_3 CA \times POST}_{\text{POLICY}} + U$$



↑

counterfactual

↑

parallel trends