ECON 1960 - Identification Strategies

October, 2024

Difference-in-Differences (DiD)

$$y_{it} = \alpha_i + \gamma_t + \beta \cdot D_{it} + X_{it} + \epsilon_{it} \tag{1}$$

- y_{it} is the outcome of interest for unit i (county) in time t (year)
- Unit and time fixed effects α_i and γ_t
- X_{it} is a vector of time varying controls for units
- D_{it} is an indicator function where $D_{it} = \mathbb{1}_{i \in \{Treat\}} \cdot \mathbb{1}_{t \in \{Post\}}$
- Key assumption: Absent the treatment, treated units would experience the same change in outcomes as untreated units

- Stata code for estimating this:

```
global controls "population income"
reghdfe y D $controls, absorb(county year) cluster(county)
```

Dynamic DiD (Event Study)

$$y_{it} = \sum_{m=-7}^{7} \beta_m z_{i,t-m} + X_{it} + \alpha_i + \gamma_t + \epsilon_{it} \tag{2}$$

- $z_{i,t-m}$ is a dummy variable for the leads and lags of treatment status for a given unit i
- Key assumption:

$$E[\epsilon_{it} | z_{i,t-m}, X_{it}, \alpha_i, \gamma_t] = 0 \tag{3}$$

- Dual purposes:
  - Show how the effect of the treatment evolves over the course of the treatment
  - Placebo tests for the plausibility of parallel trends
- See lecture note for example Stata code for estimating this

Instrumental Variables/2SLS

First stage:

$$X_i = \pi_0 + \pi_1 Z_i + \pi_2 Controls_i + u_i \tag{4}$$

Second stage:

$$Y_i = \gamma_0 + \gamma_1 \hat{X}_i + \gamma_2 Controls_i + \varepsilon_i \tag{5}$$

- Y_i is the outcome of interest for unit i
- X_i is endogenous variable
- Z_i is instrumental variable
- $Controls_i$ is a vector of controls
- Key assumptions/conditions
  - **Relevance**: the instrument predicts the endogenous variable
  - **Validity/Exclusion restriction:** the instrument only affects the outcome through the endogenous variable

- Stata code for estimating this:

```
global controls "population income"
ivreg2 y $controls (x=z), robust
```

Regression Discontinuity

$$y_i = \beta_0 + \beta_1 \mathbb{1}_{X_i \geq \tau} + f(X_i) + \varepsilon_i \tag{6}$$

- y_i is the outcome for unit i
- X_i is the running variable
- τ is the cutoff value of the running variable

- Stata code for estimating this (for τ=30):

```
rdrobust y runningvar, c(30)
rdplot y runningvar, c(30) plot
```