Identification Strategies

Francesco Ferlenga, Ali Lodermeier, Fu Jin, Jiayue Zhang

October 2024

Difference in Differences

The aim here is to document classic Difference-in-Differences (DiD) specifications and provide a guide on how to estimate these models in Stata.

Static DiD

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

Where y_{it} represents the outcome of interest for unit i (county) in time t (year) with unit and time fixed effects α_i and γ_t. X_{it} is a vector of time varying controls for units and D_{it} is an indicator function where $D_{it} = \mathbb{1}_{i \in \{Treat\}} \times \mathbb{1}_{t \in \{Post\}}$.

Code for estimating this is as follows:

```
encode(fips_state_county_code), gen(i)
gen t = year

global controls "population income"

gen first_treat = .
replace first_treat = treatment_year if !missing(treatment_year)

gen D = 0
replace D = 1 if !missing(treatment_year) & first_treat <= t

reghdfe y D $controls, absorb(i t) cluster(i)
local did_coeff = round(_b[D],.01)
```

Dynamic DiD

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

Where everything is as per before except with $z_{i,t-m}$ being a dummy variable for the leads and lags of treatment status for a given unit i. The identifying assumption is that:

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

```
encode(fips_state_county_code), gen(i)
gen t = year

global controls "population income"

gen first_treat = .
replace first_treat = treatment_year if !missing(treatment_year)

gen rel_time = t - first_treat // event time

gen never_treat = first_treat==.  // never treated group

sum first_treat
gen last_cohort = first_treat==r(max) // last treated

gen gvar = first_treat
recode gvar (. = 0)

 // leads
 cap drop F_*
 cap drop ref*
 cap drop stack

  summ rel_time
  local relmin = abs(r(min))
  local relmax = abs(r(max))

 // leads
 cap drop F_*
 forval x = 1/`relmin' {  // drop the first lead
  gen     F_`x' = rel_time == -`x'
  replace F_`x' = 0 if never_treat==1
 }


 //lags
 cap drop L_*
 forval x = 0/`relmax' {
  gen     L_`x' = rel_time ==  `x'
  replace L_`x' = 0 if never_treat==1
 }

 ren F_1 ref  // reference year

reghdfe y L_* F_* $controls, absorb(i t) cluster(i)
estimates store twfe
```

```
colorpalette tableau, nograph

event_plot    twfe, ///
 stub_lag(L_#) ///
 stub_lead(F_#) ///
together trimlead(10) trimlag(10) noautolegend ///
plottype(scatter) ciplottype(rspike) ///
lag_opt1(msymbol(+) msize(1.2) mlwidth(0.3) color(black)) ///
lag_ci_opt1(color(black) lw(0.15))

graph export "./output/analysis/figures/eventstudy.pdf", replace
```