AREC422 Notes

Youpei Yan

May 7, 2019

Panel Data Analysis Cont'd

Econometric way:
$price = \beta_0 + \delta_0 yrB + \beta_1 near\_inc + \delta_1 yrB \cdot near\_inc + u$, where $\hat{\delta}_1$ has the same value as before.
We should get the following: $\widehat{price} = 82 + 19yrB - 18near\_inc - 12yrB \cdot near\_inc$

5. Policy Analysis with DID

In a policy: we have a treatment group (T) & a control group (C).
In the previous example, the treated group is the group of houses near the incinerator, and the control group is the group away from the incinerator.

Period 1: both of them are untreated.
Period 2: T group is treated. (affected by a policy or an event)

$y = \beta_0 + \delta_0 yr2 + \beta_1 groupT + \delta_1 yr2 \cdot groupT + \beta_k x_k + u$

	C	T	Diff(T-c)
yr 1	β_0	$\beta_0 + \beta_1$	β_1
yr2	$\beta_0 + \delta_0$	$\beta_0 + \delta_0 + \beta_1 + \delta_1$	$\beta_1 + \delta_1$
yr2-yr1(DID)			δ_1

$\therefore \hat{\delta}_1 = (\bar{y}_{2T} - \bar{y}_{2C}) - (\bar{y}_{1T} - \bar{y}_{1C})$

III. 2-period panel data analysis

Still, there are so many factors unable or hard to control for, which will lead to OVB problems.
Some unobservables are constant over time for each of the individual.

Let these unobservables for person i be a_i, a_i is not changing over time, which is known as the fixed effects. t denotes the time period.

Fixed effects model will be:
$y_{it} = \beta_0 + \delta_0 yr2(+\delta_1 yr3 + ...) + \beta_1 X_{it} + a_i + u_{it}$
With multiple years' data, we can cancel a_i using the difference between the equation, so a_i can no longer bias the estimators.

The error term ($a_i + u_{it} = v_{it}$) is the composite error, in pooled OLS, $corr(a_i, X_{it}) \neq 0$ can lead to biased estimation.

Advantage of collecting panel data: allow a_i to be correlated with X_{it} and we "difference it away" when we observe the same individual multiple times.

Mathematically with 2-period panel:

In year B: $y_{iB} = (\beta_0 + \delta_0) + \beta_1 X_{iB} + a_i + u_{iB}$
In year A: $y_{iA} = (\beta_0 + 0) + \beta_1 X_{iA} + a_i + u_{iA}$

Difference of the two: $\Delta y_i = \delta_0 + \beta_1 \Delta X_i + \Delta u_i$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (\Delta X_i - \Delta \bar{X}) \Delta y_i}{(\Delta X_i - \Delta \bar{X})^2}$$

This is also called the first difference (FD) equation.

$\hat{\beta}_1$: FD estimator in the 2-period FE model. Without the differencing step, $\hat{\beta}_1$ will be biased.

Example: land use before/after CRP:

	yr1	yr2
C	no	no
T	no	effect

$land\_conv_{it} = \beta_0 + \delta_0 y2_t + \beta_1 CRP_{it} + a_i + u_{it}$, with
$y2_1 = 0$, $y2_2 = 1$, $CRP_{i1} = 0$, $CRP_{i2} = 0$ for C, $CRP_{i2} = 1$ for T.

a_i: $farm_i$'s fixed effects, such as soil quality, farmer's average income, local weather on average, slope of land, etc.
We don't need to control them anymore, as long as they are not changing in these two years.

$\Delta land\_conv = \delta_0 + \beta_1 \Delta CRP + \Delta u = -5 + 7\Delta CRP$

$\therefore$ CRP's effect is 7, i.e., holding everything else fixed, 7 acres of land will be retired in acreas where the program is in effect.

$\hat{\beta}_1 = \overline{\Delta y_{treat}} - \overline{\Delta y_{control}}$, the DID/FD estimator.