AREC422 Notes

Youpei Yan

May 2, 2019

Panel Data Analysis

I. Pooling cross section vs. Panel

Data types: cross section, time series, multi-year cross section, panel.
So far, we've only seen cross sectional data: a data sample in a certain day/year from a population.
There is no correlation in the error terms across different observations.
Assumption: $corr(u_i, u_j) = 0$

If we randomly select a sample from a population and re-visit the same individuals (firms/farms/people/counties...) at several subsequent points in time, it will be panel data (longitudinal data).

In panel data, unobserved factors for each individual are correlated over time, and can give biased estimators if we do not take that into account.

II. Pooling independent cross sections across time.

We pool the yearly/monthly/daily data into a big dataset, because they're random draws into the population, errors are not correlated.

Advantage: increase the sample size: more precised estimators & test statistics have more power.
Challenges: distributions of key variables (wages, productions, prices...) can change over time. We should allow the intercept to change over time.

1. Year dummy variables:
which can answer the question: after controlling for other observed factors, what happens to the LHS var. over time?

e.g. $house\_price = \beta_0 + \beta_1 room + \beta_2 size + \beta_3 NO_2 + \beta_4 y2006 + \beta_5 y2007 + \beta_6 y2008 + u,$ when we have house price data from 2015 to 2018.
- Why do we include only $y2006$-$y2008$? To avoid perfect collinearity (treat $year = 2015, 2016, 2017, or 2018$ as a categorical variable).
- Interpretation in front of the yearly dummy: comparison between groups as before.
If $\beta_4 = -1000$: holding other factors fixed, the average housing price in 2016 is $1,000 lower comparing to it in 2015 (base group year).

2. Year dummy interaction with a key variable:
which can answer the question: what is the effect of this key variable has changed over a certain time period.

e.g. say we have only 2017 and 2018 data:
$house\_price = \beta_0 + \beta_1 room + \beta_2 size + \beta_3 NO_2 + \beta_4 y2008 + \beta_5 NO_2 \cdot y2008 + u$

In 2017: $house\_price = \beta_0 + \beta_1 room + \beta_2 size + \beta_3 NO_2 + u$
In 2018: $house\_price = (\beta_0 + \beta_4) + \beta_1 room + \beta_2 size + (\beta_3 + \beta_5) NO_2 + u$
$\therefore \beta_5$ means how the return to another year of pollution has changed over a year.

Testing if there is any change: $H_0 : \beta_5 = 0$ vs. $H_1 : \beta_5 > 0$.

3. Chow Test for structural changes across time:
Similar to m groups before, we have T periods. If we just care about changes in slopes:

$$F_{(T-1)k,[n-T(k+1)]} = \frac{(SSR_r - SSR_{ur})/[(T-1)k]}{SSR_{ur}/[n - T(k+1)]}$$

4. Difference-in-Differences (DID) estimator

e.g. Collect data of housing price & the location of the house relative to a garbage incinerator in 2 years.
The rumor of building the incinerator is after year A, and the construction started in year B.
We want to know the impact of locating near the incinerator to the price of houses, but the incinerator could be placed in a low-price region. How to control for it?

There could be unobserved demographic information: neighborhood issues, # crimes, distribution of schools, etc. Fortunately, these issues should not change too much in 2 years.

We can use the DID estimator to control these persisted unobservables in a DID model.

In year A: $\widehat{price} = \hat{\beta}_0 + \hat{\beta}_1 near\_inc = 82 - 18 near\_inc$
In year A: $\widetilde{price} = \tilde{\beta}_0 + \tilde{\beta}_1 near\_inc = 101 - 30 near\_inc$

2nd year - 1st year: the difference between the 2 slopes: $(-30 - (-18)) = -12$
$\hat{\delta} = \tilde{\beta}_1 - \hat{\beta}_1$ is the estimated effect of the incinerator on values of houses near the incinerator site.
This is called the DID estimator, because $\tilde{\beta}_1$ is the difference between price of houses near & further away from an incinerator in year B, $\hat{\beta}_1$ is the difference between price of houses near & further away from an incinerator in year A.