Differences-in-Differences

Jörn-Steffen Pischke

LSE

October 10, 2016

Fixed effects versus differences-in-differences

- Recall how the fixed effects model assumes

$$E(Y_{0it}|A_i, t) = \alpha + \lambda_t + \gamma A_i$$

or

$$E(Y_{0it}|i, t) = \alpha_i + \lambda_t$$

- The differences-in-differences (DD) model makes a very similar assumption but conditions on a group level instead of an individual level effect

$$E(Y_{0ist}|s, t) = \gamma_s + \lambda_t$$

where s could be, for example, a state. This is sufficient for any treatment that happens at the state-time level.

- While the basic strategy is the same, the data requirements are much less. We don't need repeated observations on unit i (i.e. a panel). Repeated cross-sections sampling from the same aggregate units s are sufficient.

Compulsory schooling laws

- An example for a differences-in-differences setup would be the effect of compulsory schooling laws on schooling obtained in the US. These laws are set at the state level, and different states change the compulsory schooling laws at different times. For example, Florida raised its compulsory schooling requirement from 5 to 7 grades in 1935. Neighboring Georgia required 6 grades both before and after 1935.
- We can think of FL as the treatment state and GA as the control state.
- 1934 is a control period and 1935 is the treatment period.

Differences-in-differences

Let D_{st} denote a dummy for the treatment, i.e. a compulsory schooling requirement in FL.

$$Y_{ist} = \gamma_s + \lambda_t + \beta D_{st} + e_{ist}$$

Using this it is easy to see that

$$E[Y_{ist}|s = GA, t = 1935] - E[Y_{ist}|s = GA, t = 1934] = \lambda_{1935} - \lambda_{1934}$$

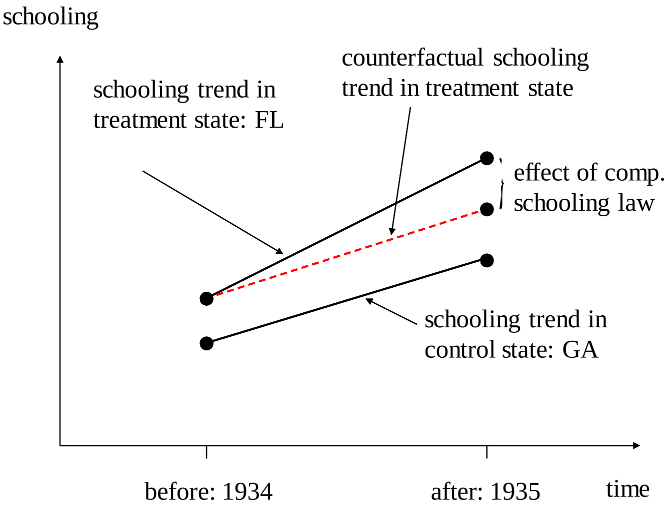
$$E[Y_{ist}|s = FL, t = 1935] - E[Y_{ist}|s = FL, t = 1934] = \lambda_{1935} - \lambda_{1934} + \beta$$

The population difference-in-difference is

$$[E(Y_{ist}|s = FL, t = 1935) - E(Y_{ist}|s = FL, t = 1934)]$$

$$-[E(Y_{ist}|s = GA, t = 1935) - E(Y_{ist}|s = GA, t = 1934)] = \beta$$

Identification in the differences-in-differences model



Regression DD

$$
\begin{aligned}
Y_{ist} &= \gamma_s + \lambda_t + \beta D_{st} + e_{ist} \\
&= \alpha + \gamma 1(s = FL) + \lambda 1(t = 1935) \\
&\quad + \beta 1(s = FL) \cdot 1(t = 1935) + e_{ist}
\end{aligned}
$$

where $1(\cdot)$ is the indicator function. Taking conditional expectations for different states and periods, and subtracting easily yields

$$
\begin{aligned}
\alpha &= E(Y_{ist}|s = GA, t = 1934) = \gamma_{GA} + \lambda_{1934} \\
\gamma &= E(Y_{ist}|s = FL, t = 1934) - E(Y_{ist}|s = GA, t = 1934) \\
&= \gamma_{FL} - \gamma_{GA} \\
\lambda &= E(Y_{ist}|s = GA, t = 1935) - E(Y_i|s = GA, t = 1934) \\
&= \lambda_{Nov} - \lambda_{Feb} \\
\beta &= [E(Y_{ist}|s = GA, t = 1935) - E(Y_i|s = GA, t = 1934)] \\
&\quad - [E(Y_{ist}|s = FL, t = 1935) - E(Y_i|s = FL, t = 1934)] .
\end{aligned}
$$

Advantages of the regression formulation

$$Y_{ist} = \gamma_s + \lambda_t + \beta D_{st} + e_{ist}$$

1. The regression gives you a standard error and t-statistic on β.
2. Can easily extend the DD framework to more than two states and periods: e.g. Acemoglu and Angrist (2000) use individuals born between 1910 - 1940 in any of the lower 48 states. If we use more than 2x2 states/periods:
3. Can use a multivalued treatment indicator

$$Y_{ist} = \gamma_s + \lambda_t + \beta CS_{st} + e_{ist}$$

   where CS_{st} takes on values from 6 to 9 for the number grades required, or multiple dummies for the different treatments.
4. Can add covariates

$$Y_{ist} = \gamma_s + \lambda_t + \beta CS_{st} + X_{st}\delta + e_{ist}$$

Covariates in the DD model

- Start with the regression

$$Y_{ist} = \gamma_s + \lambda_t + \beta CS_{st} + e_{ist}.$$

We could run this on the micro data or aggregate to the state level

$$Y_{st} = \gamma_s + \lambda_t + \beta CS_{st} + e_{st}.$$

Both regressions (the 2nd weighted by the number of obs. in the cell) give the same estimates since regressors just vary at the group level.

- For the same reason, only covariates at the state/year level matter for identification

$$Y_{ist} = \gamma_s + \lambda_t + \beta CS_{st} + X_{st}\delta + e_{ist}$$

- We might want to include individual level covariates

$$Y_{ist} = \gamma_s + \lambda_t + \beta CS_{st} + X_{ist}\delta + e_{ist}.$$

The within state variation doesn't matter for identification but may reduce standard errors.

Assessing DD identification

The key identifying assumption in DD models is that the treatment states have similar trends to the control states in the absence of treatment.

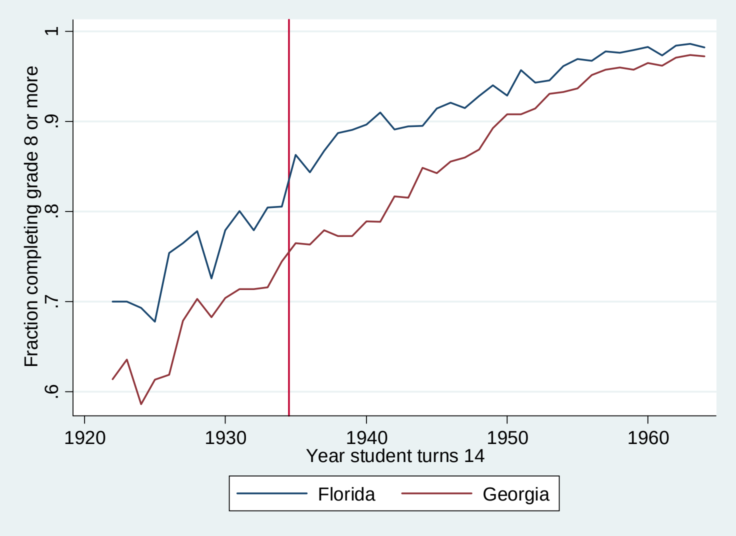
- With only one treatment and control group, graph your results, and look at trends in periods with before the treatment.
- With many treatment and control groups:
  - and a binary treatment, estimate treatment impacts at different dates

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{j=-m}^{q} \beta_j D_{st+j} + e_{ist}$$

    where D_{st} is now an indicator for whether the treatment got switched on in year t. This estimates q leads and m lags of the treatment. The leads should all be zero.
  - include state specific trends.

Graph for two states

Florida raises compulsory schooling from 5 to 7 grades in 1935



Some DD estimates

Years of schooling on child labor laws

Regressor	(1)	(2)	(3)	(4)	(5)
CL7	0.51	0.80	0.15	0.09	0.04
	(0.43)	(0.36)	(0.36)	(0.10)	(0.04)
CL8	1.37	1.11	0.77	0.20	0.05
	(0.28)	(0.22)	(0.33)	(0.10)	(0.05)
CL9	1.56	2.39	−0.28	0.37	0.06
	(0.31)	(0.21)	(0.37)	(0.12)	(0.04)
State effects		✓		✓	✓
Year effects			✓	✓	✓
State trends					✓

Some more DD estimates

Years of schooling on child labor laws

	Dependent variable			
	completes 8+ years		completes 10+ years	
Regressor	(1)	(2)	(3)	(4)
CL7	0.01	0.009	0.01	0.001
	(0.02)	(0.004)	(0.02)	(0.008)
CL8	0.03	0.007	0.03	0.002
	(0.02)	(0.005)	(0.01)	(0.009)
CL9	0.05	0.010	0.05	0.005
	(0.02)	(0.004)	(0.02)	(0.009)
State effects	✓	✓	✓	✓
Year effects	✓	✓	✓	✓
State trends		✓		✓