Microeconometrics (Causal Inference)
Week 5 - Differences-in-differences I

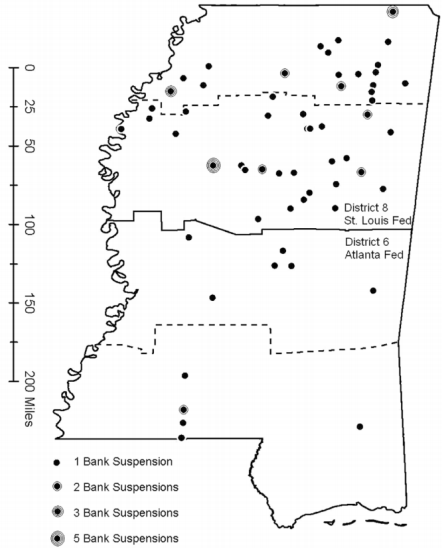
Joshua D. Merfeld
KDI School of Public Policy and Management

2023-09-18

What are we doing today?

- More commonly referred to as "DID" or "diff-in-diff"
- Most common method, likely because data requirements are least stringent
- Example in *Mostly Harmless*: offering credit to banks during the Great Depression (Richardson and Troost, 2009)
  - Set up: Two different federal reserve banks lent to neighborhood banks in Mississippi
  - Atlanta fed favored lending to banks in trouble
  - St. Louis fed favored the exact opposite

Richardson and Troost (2009) - Mississippi dividing line



Did the policy of extra lending save banks?

- Basic idea: compare what happened to Atlanta fed banks (southern Mississippi) with St. Louis fed banks (northern Mississippi)
- Could compare after lending, but what's the assumption here?

Did the policy of extra lending save banks?

- Basic idea: compare what happened to Atlanta fed banks (southern Mississippi) with St. Louis fed banks (northern Mississippi)
- Could compare after lending, but what's the assumption here?
- Assumption: same levels before intervention (very strict assumption)

In fact, pre-intervention levels are different!

TABLE 4
BANK SUSPENSIONS AND LIQUIDATIONS

Begin July 1		End June 30	Percentage of Banks Suspending: All (1)	Suspending: Federal Reserve District: 6th Atlanta (2)	Suspending: Federal Reserve District: 8th St. Louis (3)	Percentage of Banks Liquidating: All (4)	Liquidating: Federal Reserve District: 6th Atlanta (5)	Liquidating: Federal Reserve District: 8th St. Louis (6)
1929	to	1930	4.8	7.1	3.0	4.5	7.1	2.4
1930	to	1931	28.9	14.2	39.5	13.6	7.1	18.6
1931	to	1932	13.2	14.9	11.8	8.0	7.9	8.1
1932	to	1933	7.7	7.5	7.9	7.3	6.5	7.9
1933	to	1934	.9	.0	1.7	.9	.0	1.7
1929	to	1934[a]	49.8	38.7	59.2	30.9	26.8	34.4

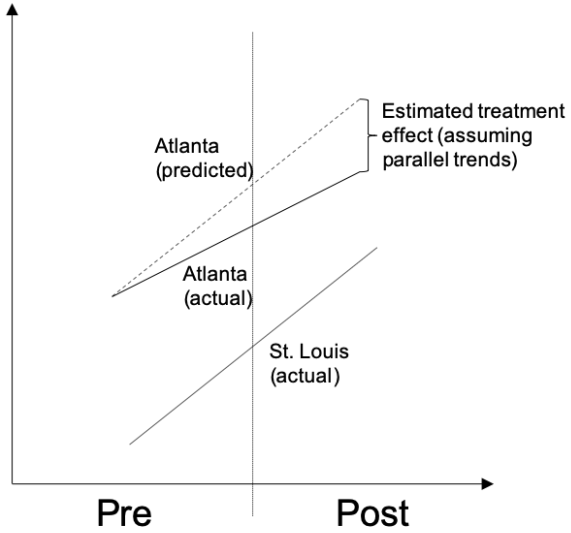
SOURCE.—*Rand McNally Bankers Directory* and National Archives and Records Administration Record Group 82. See Section II and Richardson (2006, 2007*a*, 2007*b*, 2008) for details.

[a] The last row indicates the percentage of banks operating on July 1, 1929, that either suspended or liquidated by June 30, 1933.

Did the policy of extra lending save banks?

- Instead, compare *changes* from before to after treatment
- Assumption: parallel trends
- If valid, the fact the districts were different prior to the treatment isn't a problem

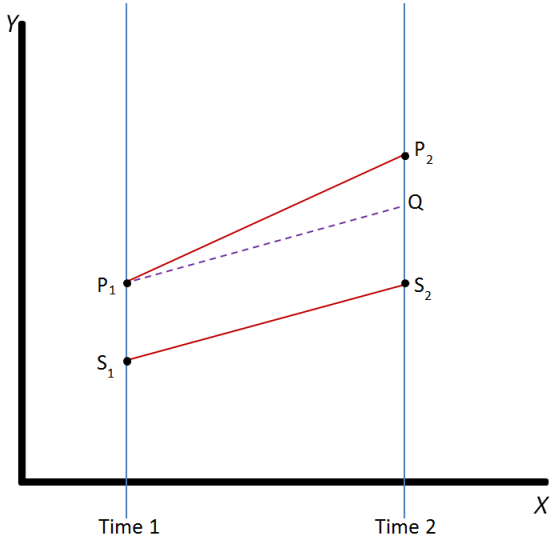
"Parallel trends"



Why is it “differences in differences”?

- Difference 1: St. Louis post minus St. Louis pre
- Difference 2: Atlanta post minus Atlanta pre
- Difference-in-differences: Difference 2 minus difference 1

“Differences in differences” graphically



Parallel trends assumption

- The key assumption in differences-in-differences is the parallel trends assumption
  - *If the treated group had not been treated, it would have changed by the same amount ("had the same trend") as the comparison group.*
- This is a counterfactual assumption: We cannot explicitly test it
- What can we do instead?

- The key assumption in differences-in-differences is the parallel trends assumption
  - *If the treated group had not been treated, it would have changed by the same amount ("had the same trend") as the comparison group.*
- This is a counterfactual assumption: We cannot explicitly test it
- What can we do instead?
  - We can test trends *before* treatment
  - Or in the case of this article, *after* treatment!

Richardson and Troost (2009) - Testing the assumption

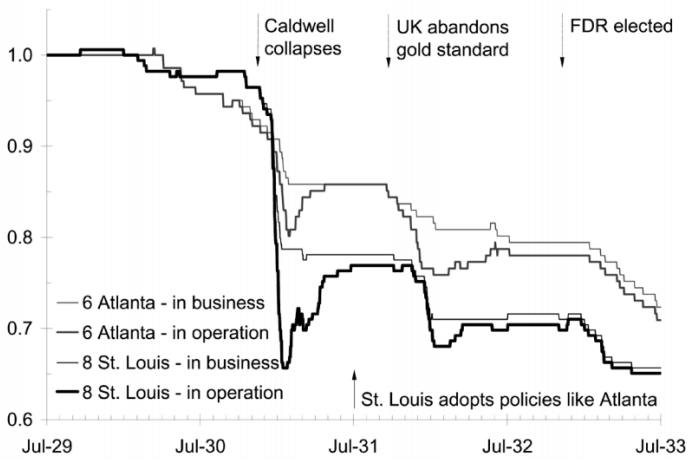


FIG. 3.—Percentage of banks in business and in operations in the 6th and 8th Federal Reserve Districts in Mississippi, July 1929 to June 1933. Source: See Section II.

- Can be estimated in a straightforward regression:

$$Y_{it} = \beta_0 + \beta_1 TREAT_i + \beta_2 POST_t + \beta_3 (POST_t \times TREAT_i) + \varepsilon_{it} \quad (1)$$

- Can be estimated in a straightforward regression:

$$Y_{it} = \beta_0 + \beta_1 TREAT_i + \beta_2 POST_t + \beta_3 (POST_t \times TREAT_i) + \varepsilon_{it} \quad (1)$$

  - β_0: pre mean for the comparison group
  - β_1: difference in the pre mean between the treated and untreated group
  - β_2: difference in means between the pre and post period for the comparison group
  - β_3: difference-in-differences estimate
    - This is the difference in the change from pre to post for the treated group relative to the comparison group

- Can add control variables

$$Y_{it} = \beta_0 + \beta_1 TREAT_i + \beta_2 POST_t + \tag{2}$$
$$\beta_3 (POST_t \times TREAT_i) + X_{it} + \varepsilon_{it} \tag{3}$$

- Adding controls can help control for differing trends ("conditional" parallel trends)
- Note: the interpretation of β_0 is no longer the same; others stay the same