# HANDOUT 17 – DIFFERENCES-IN-DIFFERENCES + FIXED EFFECTS CASE STUDY

## AGENDA

- Fixed effects
- Difference-in-difference
- Instrumental Variables
- Takeaways

## BIBLIOGRAPHY FOR TODAY´S CLASS

- Goodman (2017). The Labor of Division: Returns to Compulsory High School Math Coursework. Focus on the Introduction and previous literature, and empirical results from pages 17-21 (*).
- Stock and Watson (2007), 10.3-10.6 (**).

## THE LABOR MARKET RETURN TO MATH COURSEWORK

- The goal of this paper is to provide evidence about the labor market return to high school math coursework.
- The big challenge is that students who complete more math coursework differ in many ways from those who complete less math coursework.
- Not all of these differences are observable in typical data sets (motivation, math skills, parental influence, etc.).
- We need to find an exogenous source of variation in math coursework.
- The 1983 "Nation at Risk" report provides this:
    - Report said U.S. high schools allowed students too much choice of coursework.
    - States reacted by raising the minimum number of math courses students needed to complete in order to earn a high school diploma.
    - States varied in whether and then they raised such requirements.

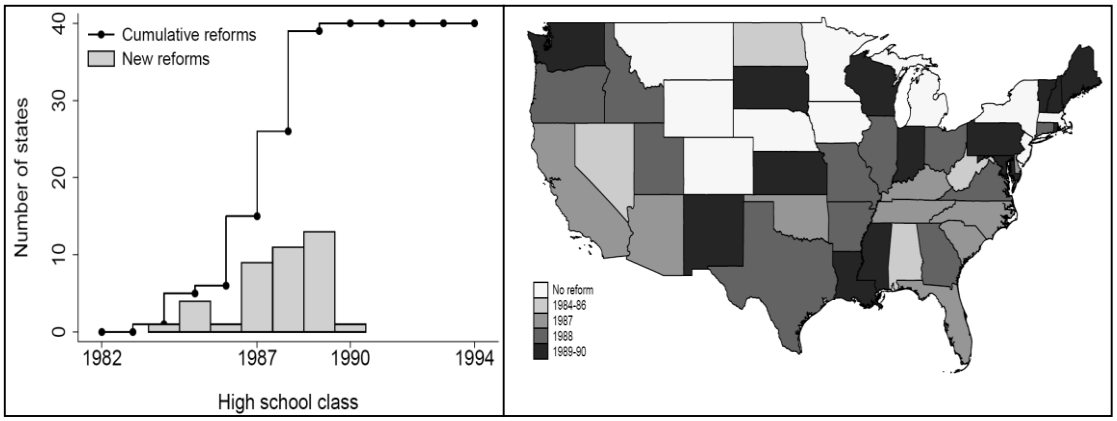
## THE DATA

- No single data set contains high school coursework and later labor market earnings for the time period of interest.
- The paper uses two separate data sets:
    - High school transcripts:
        - Nationally representative sample, state of high school.
        - From high school classes of 1982, 1987, 1990 and 1994.
        - Detailed coursework information.
        - Demographic information (race, gender).
    - 2000 U.S. Census:
        - Labor market earnings.
        - Age (to assign high school class).
        - State of birth (to assign state of high school).

- Demographic information (race, gender).

## THE IDENTIFICATION STRATEGY

- The paper exploits the fact that the timing of reforms varied by state.
- The paper asks whether the timing of changes in math coursework and in earnings seem related to the timing of each state's math reform:
  - Do math courses and earning rise earlier in states with earlier reforms and later in states with later reform?
- There are lots of variation over time and states to exploit:




X = _____ (the treatment we are interested in).

Z = _____ (the source of variation in that treatment).

Y = _____ (the outcome of ultimate interest).

The paper states with this _____ regression:

$$MathCourses_{isc} = \beta MathReform_{sc} + \mu_s + \delta_c + \varepsilon_{isc}$$

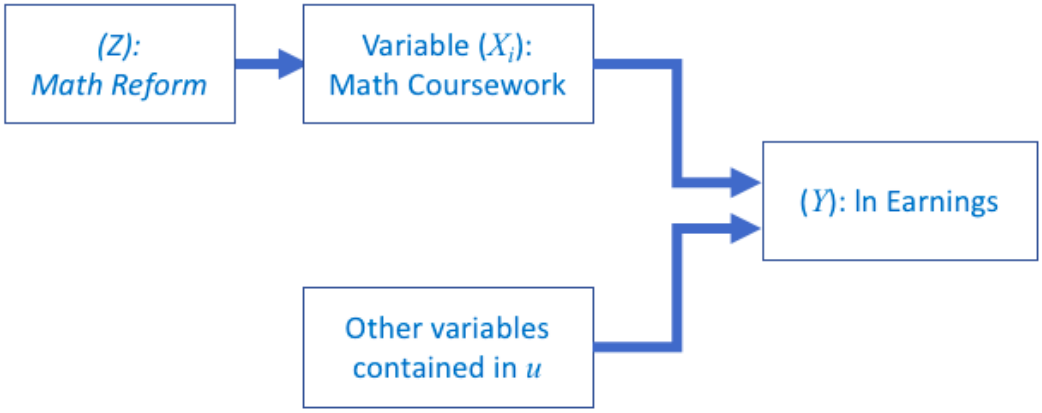
*MathCourses* = number of math courses completed by student *i*, attending high school in state *s*, from high school class *c*.

*MathReform* = 1 if student was subject to increased requirements state *s*, class *c*.

Important: regression includes high school class and state fixed effects.

## MATH COURSES AND EARNINGS

Why this is presumably a good source of exogenous variation? What is the problem?




- This is actually a difference-in-difference regression!
- To see why, imagine there where only 2 states and 2 time periods:

$$MathCourses_{isc} = \beta MathReform_{sc} + \mu_s + \delta_c + \varepsilon_{isc}$$

- Instead it's a 40 states (that had math reforms) and 4 years/cohorts.
- The state fixed effects control for any factor that differ by _____ in ways that are constant _____.
    - Critics cannot object that results are due to the fact that states enacting math reforms were higher income or otherwise better to begin with, because all estimates come from **within-state** changes in variables.
- The high school class fixed effects control for any factors that differ by _____ in ways that are constant _____.
    - Critics cannot object that results are due to the fact that reforms all happened when the national economy started to boom.
    - In fact. If all reforms happened simultaneously, this could just be a simple and likely wrong _____ comparison, but they did not!
- The coefficient from this multi-state, multi-year difference-in-difference regression measures whether:

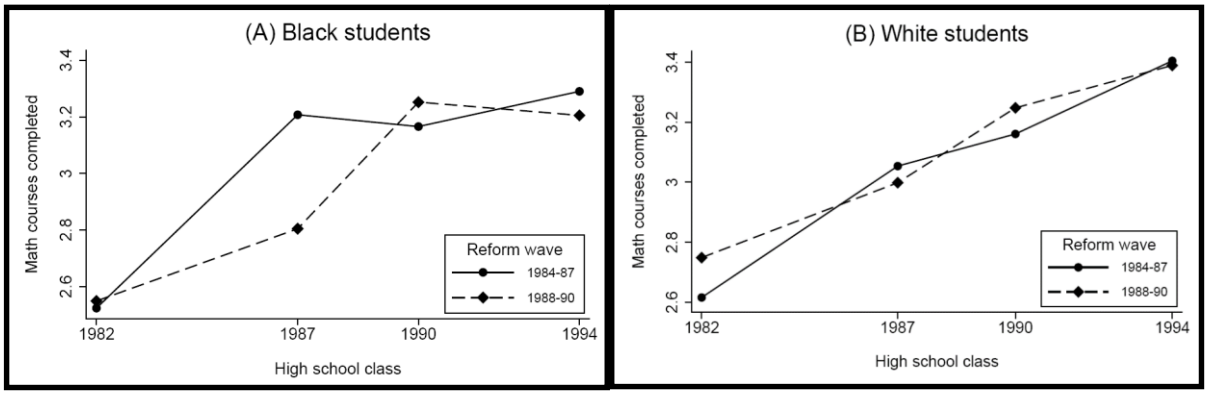
> The timing of within-state changes in treatment status is related to the timing of within-state changes in outcomes.

- Let's see if this appears to be true.
- Does the timing of state reforms seem related to:
    - Changes in complete math coursework of affected students?

- Changes in labor market earnings of affects students?

MATH COURSEWORK (FIRST STAGE RESULTS)

- Let's look at a graphical evidence from the transcript data first.
- The author splits data by race because it's the only proxy for socioeconomic status combined in both the transcript and Census data sets:



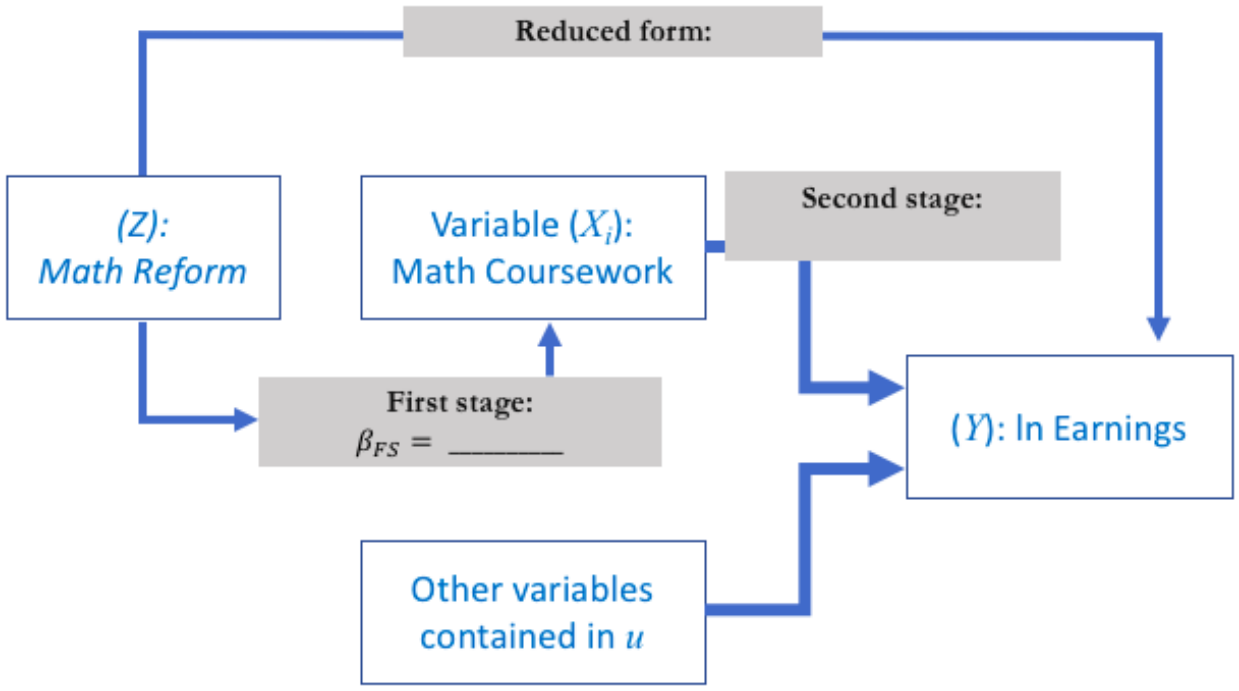

- Is reform timing connected to math coursework?

- Let's look at regression versions of these results.

Here are the *MathReform* coefficients:

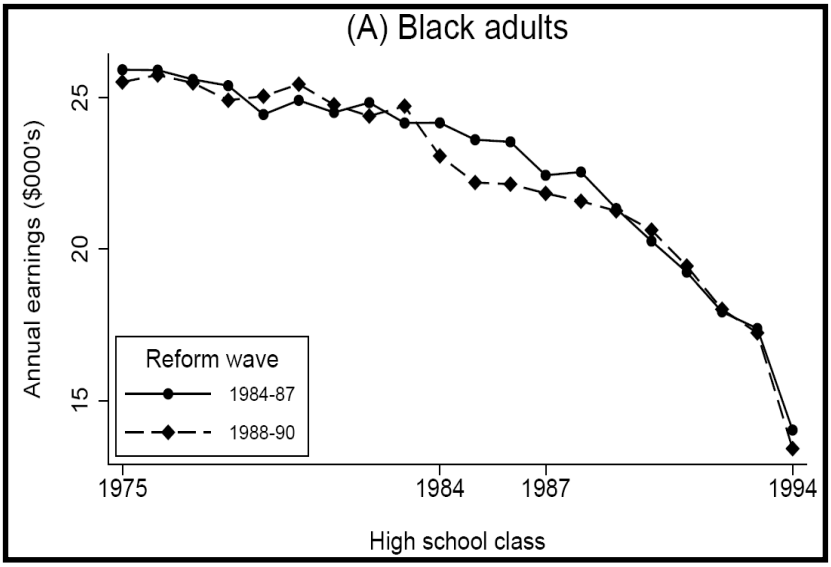
| | Black (1) | White (2) | All (3) |
|---|---|---|---|
| (A) Overall math courses | | | |
| Number of math courses | 0.354*** | 0.141 | 0.179** |
| | (0.067) | (0.084) | (0.076) |
| 1982 mean | 2.54 | 2.71 | 2.69 |

- For black students, the math reforms increased the number of completed math courses by _____, a _____ significant results.
- For white students, the math reforms increased the number of completed math courses by _____, a _____ significant results.
- The instrument only allows us to estimate the impact of math coursework on _____ students (i.e. there is no experiment being run on _____ students).




EARNINGS (REDUCED FORM RESULTS)

- Let's look at a graphical evidence from the 2000 Census earnings data:




- Does the "parallel trends" assumption appear to hold?

- The regression version of this looks nearly identical to the first stage, except that we use earnings as an outcome:

$$Earnings_{isc} = \beta_{RF} MathReform_{sc} + \mu_s + \delta_c + \varepsilon_{isc}$$

*Earnings* measure the annual earnings of individual *i* born in state *s* from high school class *c*.

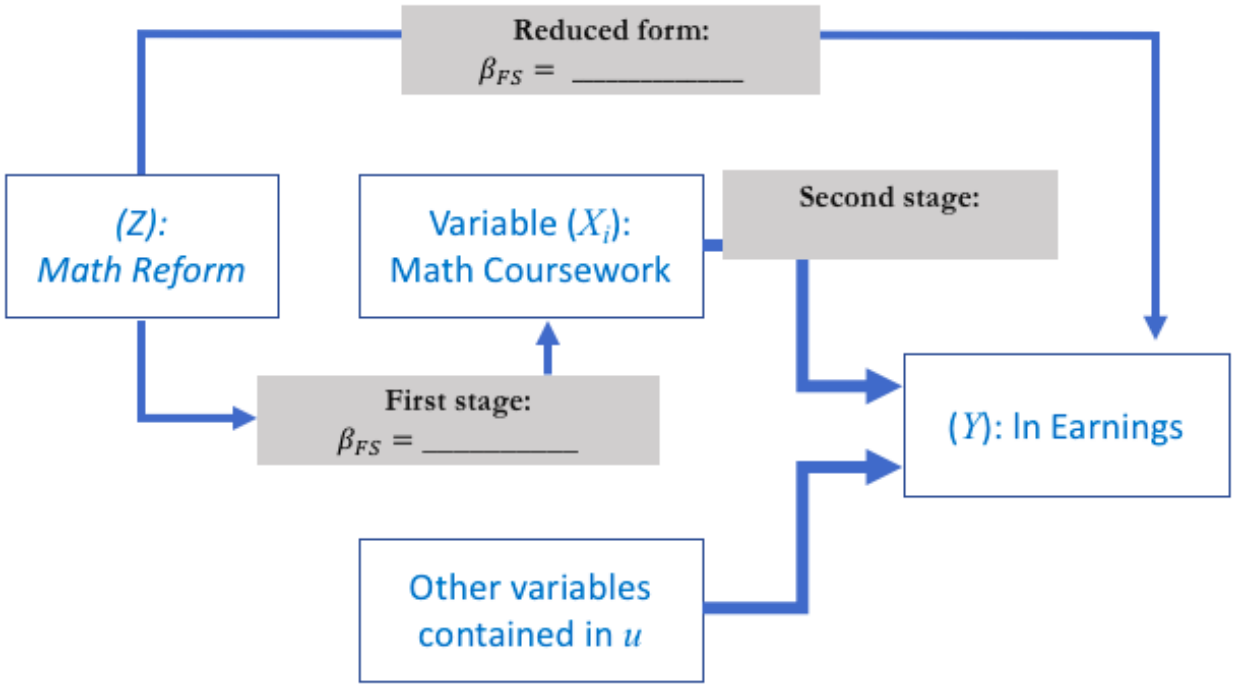
All other controls are the same.

*ln(earnings)* is used as outcome.

- Let's look at regression versions of these results.
- Here are the *MathReform* coefficients:

| | Black (1) | White (2) | All (3) |
|---|---|---|---|
| (A) High school graduates | | | |
| Ln(earnings) | 0.033** | -0.004 | 0.002 |
| | (0.014) | (0.006) | (0.005) |
| 1982 mean | 9.84 | 10.15 | 10.11 |
| N | 238,038 | 1,563,067 | 1,801,105 |
| Earnings | 0.635** | -0.071 | 0.083 |
| | (0.284) | (0.160) | (0.147) |
| 1982 mean | 21.76 | 32.91 | 31.08 |
| N | 280,299 | 1,753,217 | 2,033,516 |

- The math reforms increased black adults' earnings by _____ or _____, a _____ significant result.
- The math reforms _____ white adults' earnings.




IMPACT OF MATH COURSEWORK ON EARNINGS

- First stage estimate suggests increased math requirements cause black students to complete _____ more math courses than they would have otherwise.
- The reduced form estimate suggests that increased math requirements increased earnings of black adults by _____ percent.
- Thus, each additional math course increased earnings by _____ percent.
- This is high but represents only half the estimated return to a year of high school for these students.

TAKEAWAYS

- With multiple units (states) and time periods, we can still do difference-in-difference estimation, using fixed effects.
- Exploiting the differential timing of multiple reforms is more convincing than a single reform.
- Critiques of internal validity now have to be: "Here is another factor changing in the same states and at the same times as the policy of interest". That's possible but harder criticism to make!
- Do you have concerns about internal validity here?

- What about external validity?