
API-202B Empirical Methods II

Session #17:

Quasi-experimental methods:
Differences-in-differences + Fixed Effects Case Study

miguel_santos@hks.harvard.edu
@miguelsantos12

Today's class

- Let's do it all in one example:
 - Fixed effects
 - Differences-in-differences
 - Instrumental variables
- Takeaways

The labor market return to math coursework

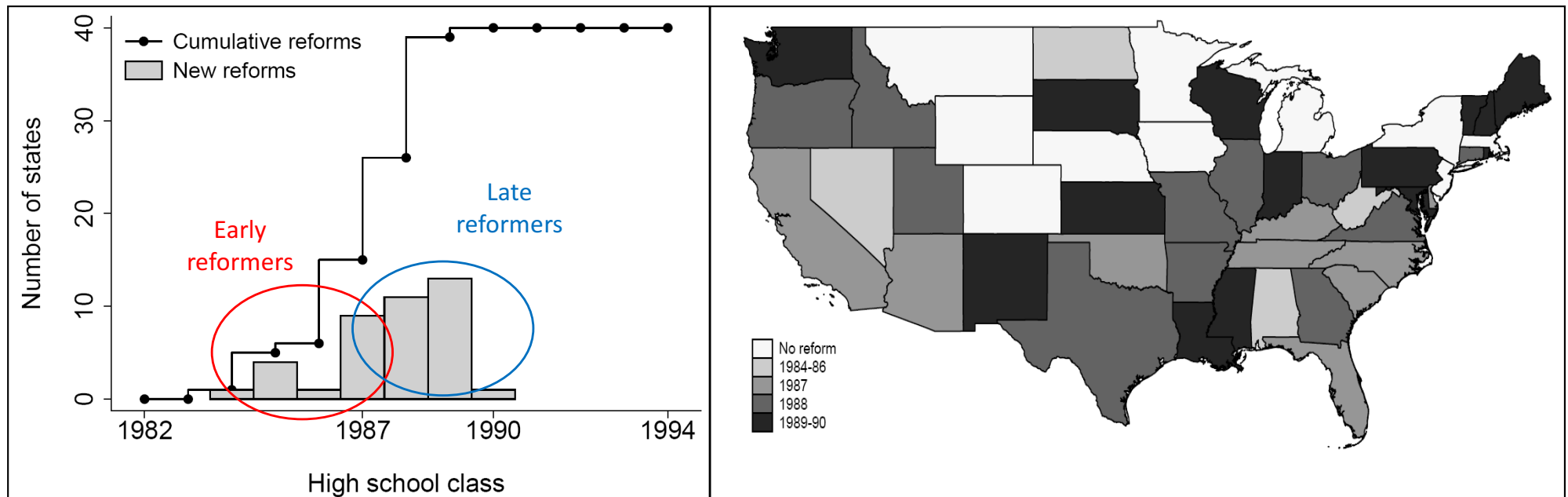
- The goal of this paper is to provide evidence about the labor market return (higher salaries) to high school math coursework
- The big challenge is that students who complete more math coursework differ in many ways from those who complete less math coursework
- Not all of these differences are observable in typical data sets (motivation, math skill, 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 when 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 earnings rise earlier in states with earlier reforms and later in states with later reforms?
- There are lots of variation over time and states to exploit:



The identification strategy

X = Math courses (the treatment we are interested in)

Z = State reforms to increase math minimum requirements (the source of variation in that treatment)

Y = Labor earnings in year 2000 (the outcome of ultimate interest)

The paper starts with this first stage 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*.

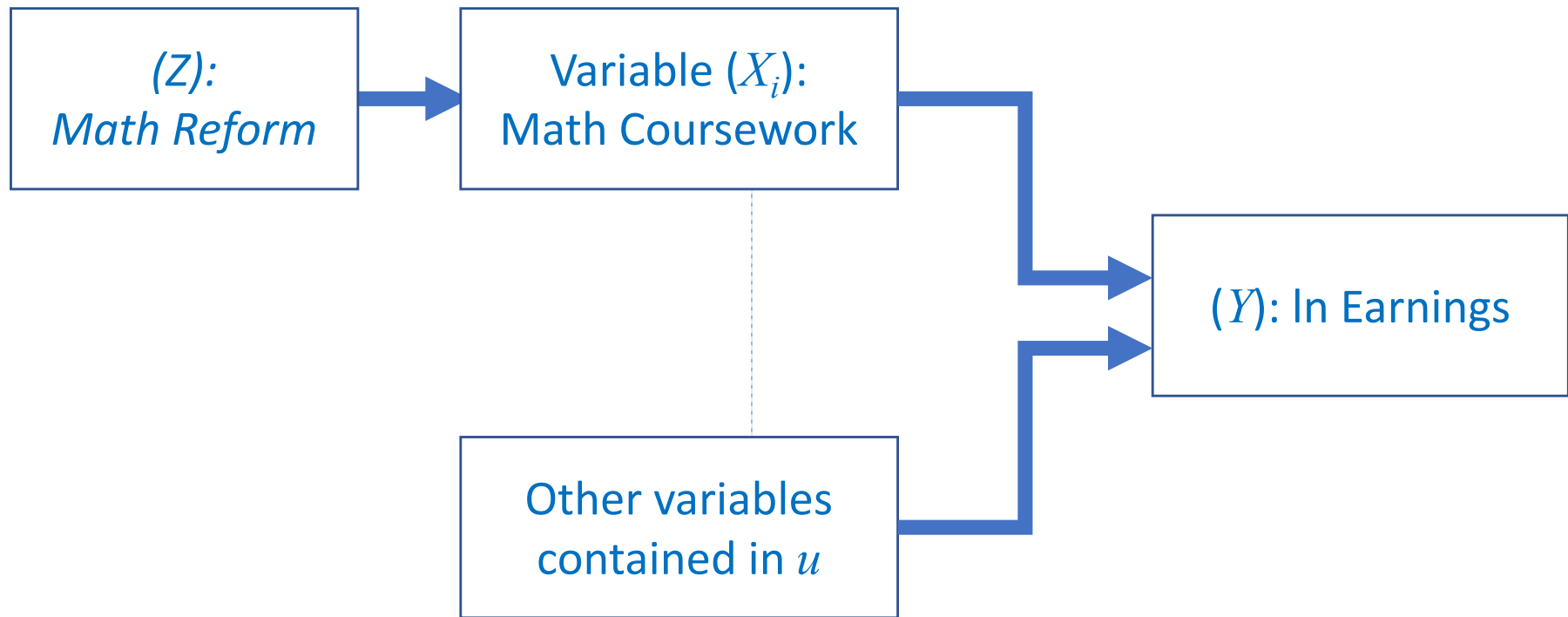
MathReform = 1 if student was subject to increased requirements state *s*, class *c*

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

Math courses and earnings

Why this is presumably a good source of exogenous variation? **Exogenous source of Math Courses Variation:** Once reform is made, timing of implementation varied by state. We assume that the timing was not related to factors affecting the relationship between income and math courses

What is the problem? **Students** receiving more math coursework differ from those who receive less in many other aspects (some of them not observable): Motivation, family factors, parental skills, math-skills.



The identification strategy

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

$$\mathit{MathCourses}_{isc} = \beta_{FS} \mathit{MathReform}_{sc} + \mu_s + \delta_c + \varepsilon_{isc}$$

- Instead it's 40 states (that had math reforms) and **4 cohort/years**.

The identification strategy

- The state fixed effects control for any factors that differ by **State** in ways that are constant through time (institutions, family values, attitudes, other pieces of legislations in place)
- 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 year/cohort class fixed effects control for any factors that differ by cohorts in ways that are constant across states.
- 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 would just be a simple and likely wrong before and after comparison, but they did not!

The identification strategy

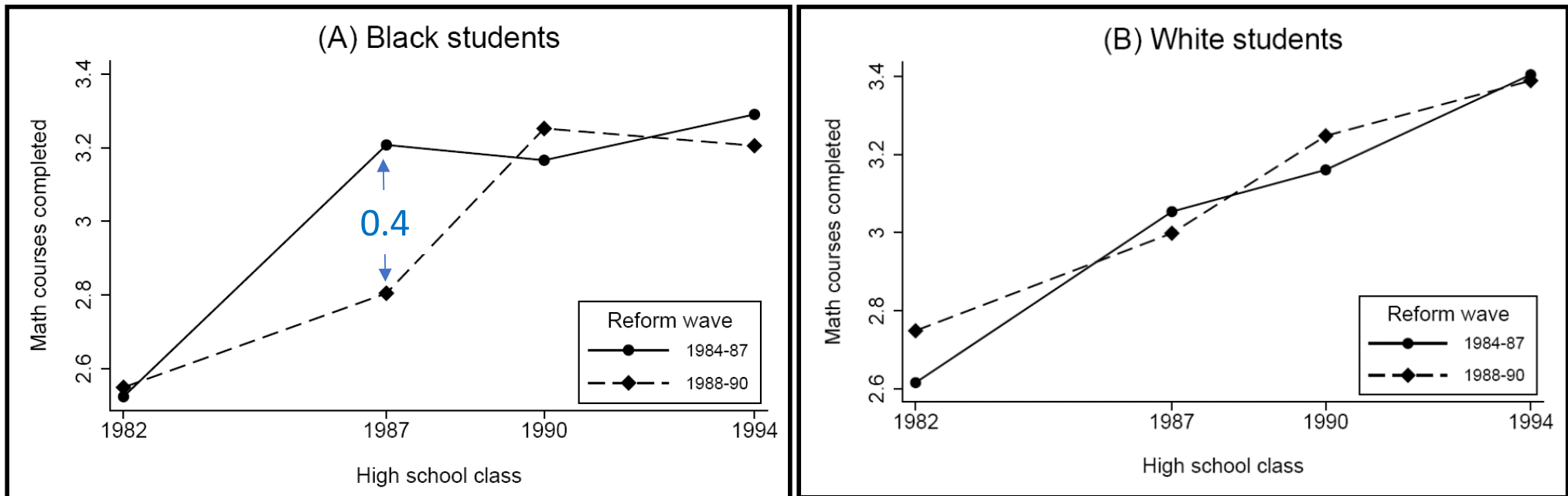
- The coefficient from this multi-state, multi-year difference-in-difference regression measures whether:

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 affected students?

Math coursework (first stage results)

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



Is reform timing connected to math coursework? For African-American students, it is clear that reform timing is indeed connected to number of math courses effectively completed.

Math coursework (first stage results)

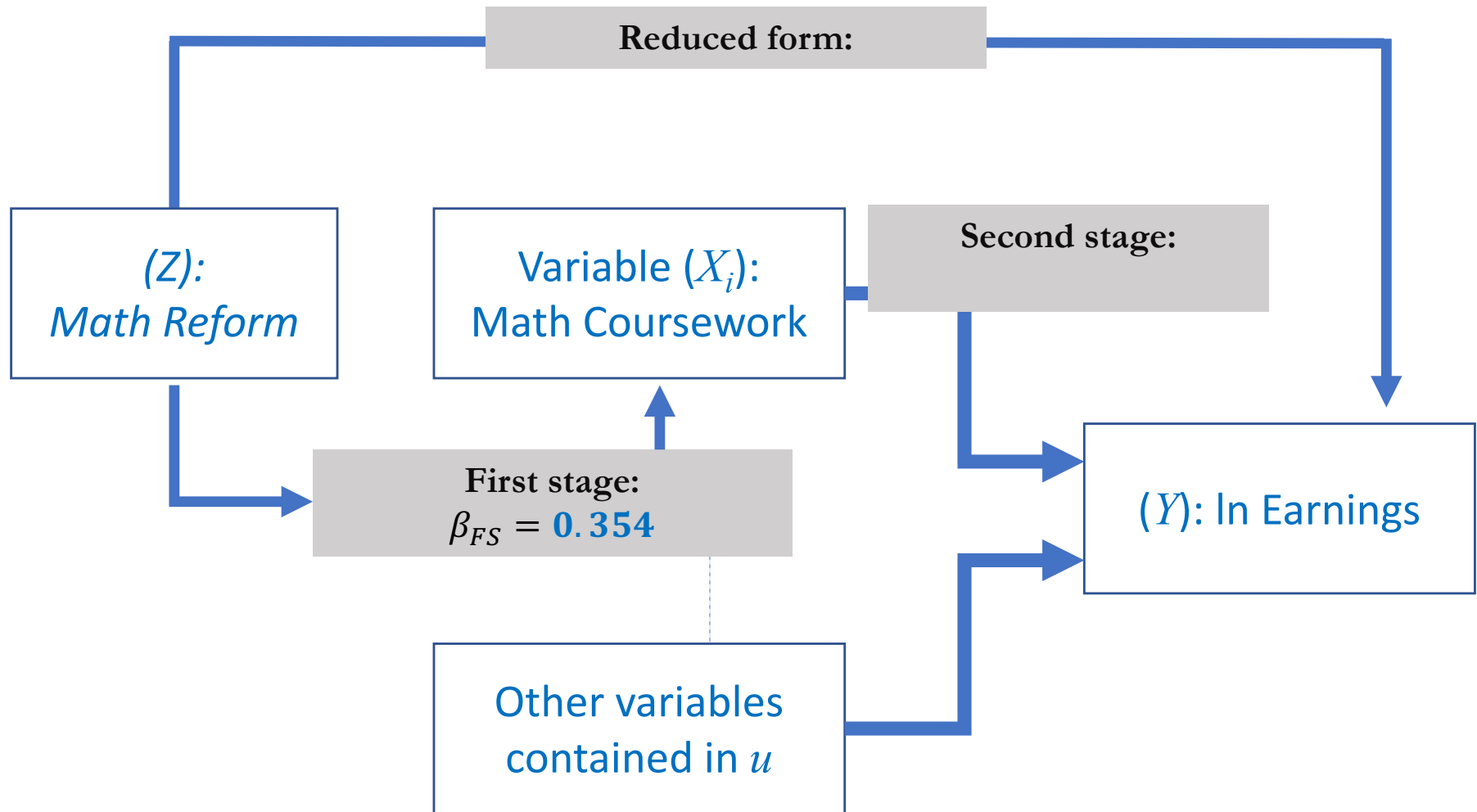
- Let's look at regression versions of these results

Here are the *MathReform* coefficients:

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

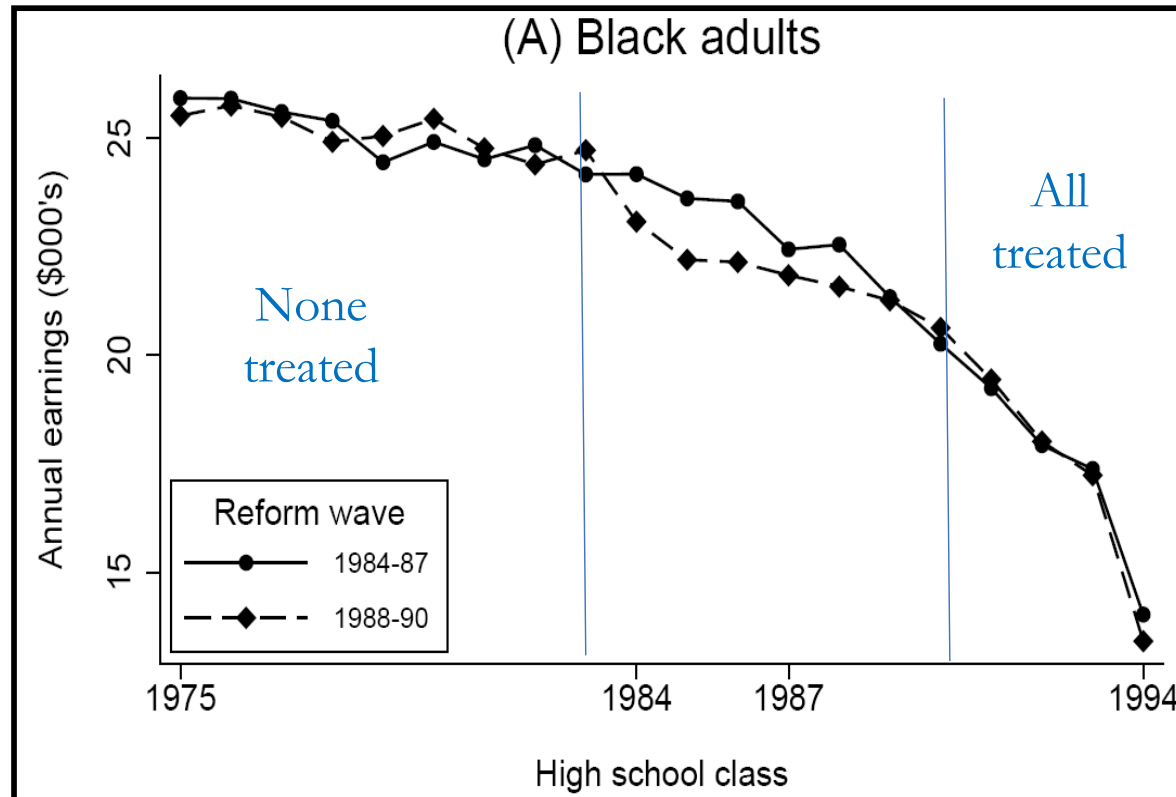
- For black students, the math reforms increased the number of completed math courses by [0.354](#), a [statistically significant](#) result.
- For white students, the math reforms increased the number of completed math courses by [0.141](#), a [statistically insignificant](#) result.
- The instrument only allows us to estimate the impact of math coursework on [African American](#) students (i.e. there is no experiment being run on [white](#) students)

Math courses and earnings



Earnings (reduced form results)

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



- Does the “parallel trends” assumption appear to hold? **Yes!** When early reformers and late reformers were either not treated (before Nation at Risk in 1983) or were all treated (after 1990) their average earnings moved in parallel.

Earnings (reduced form results)

- 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 measures the annual earnings of individual i born in state s from high school class c

All other controls are the same

$\ln(\text{earnings})$ is used as outcome

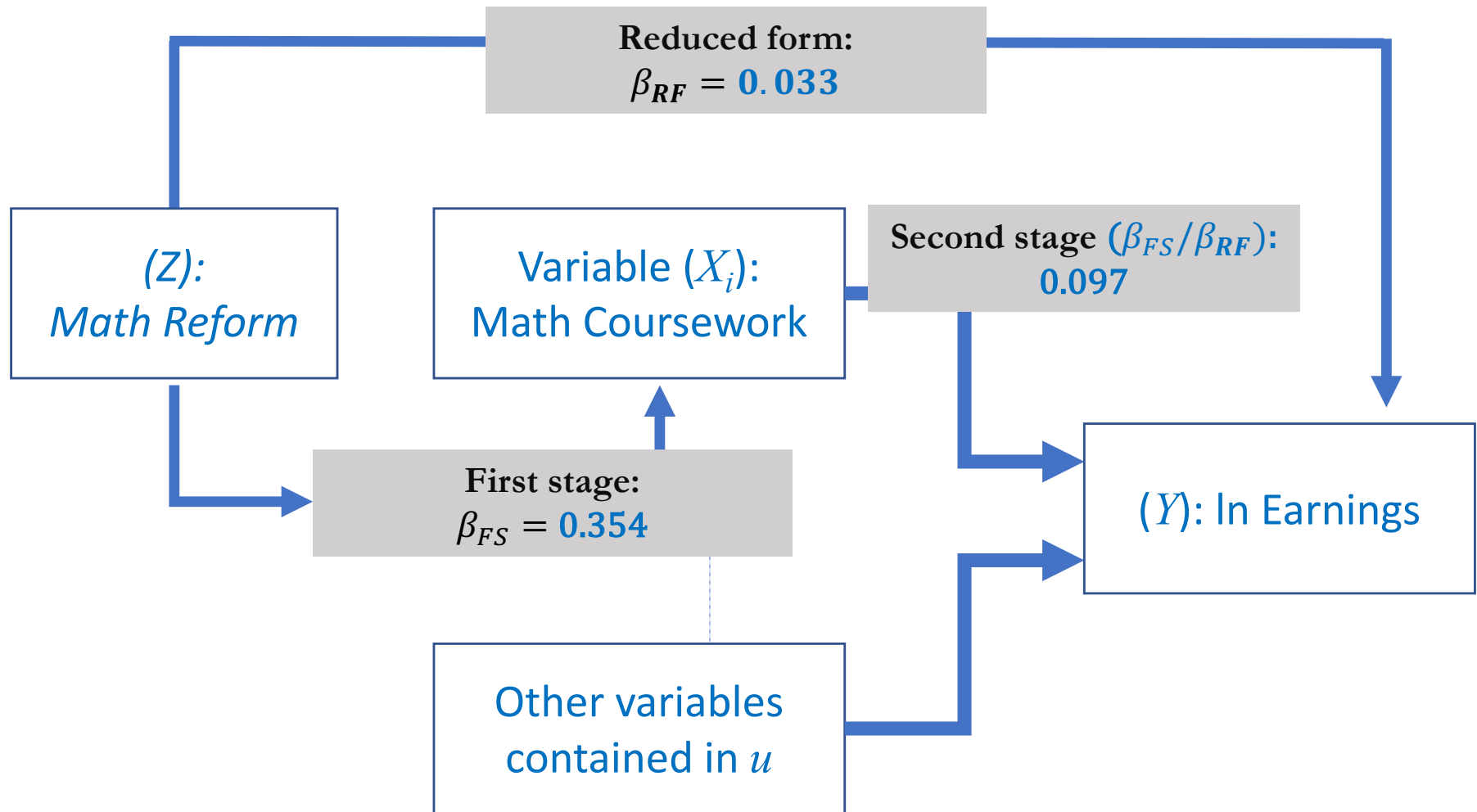
Earnings (reduced form results)

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

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

- The math reforms increased African American highschool graduates' earnings by 3.3% or 635 dollars, a statistically significant result.
- The math reforms did not have any impact on white adults' earnings.

Math courses and earnings



Impact of Math Coursework on Earnings

- First stage estimate suggests increased math requirements caused black students to complete 0.354 more math courses than they would have otherwise
- The reduced form estimate suggests that increased math requirements increased earnings of black adults by 3.3% percent.
- Thus, each additional math course increased earnings by 9.7% 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 a possible but harder criticism to make!
- Do you have concerns about internal validity here? There would have to be a change implemented within state at the same time of the education reform, that also impacts earnings. In general this is not implausible, but what makes it harder in this case, is that to question internal validity we would need one of those changes occurring in all states at the time of reform implementation, which is hard given that states reformed at different years.
- What about external validity? Results seem to be relevant only to those groups that study the minimum math requirements established by the law. To those already studying more than the minimum (presumably whites in this paper), changing the minimum will not change the outcome.