Estimating Causal Effects with Observational Data: Evidence from Title IX on how Sports Impacts Kids

Betsey Stevenson

The Wharton School, University of Pennsylvania

Estimating Causal Effects with Observational Data

There are four main approaches economists use:

- **Cross-sectional regression analysis**
  - Uses observable measures of differences between people
  - LIMITATION: often unobserved differences between groups
- **Time series analysis**
  - Documents the correlation between the variables of interest over time; works best if there are sharp changes in a policy variable over time
  - LIMITATION: often many things are changing over time at once
- **Quasi-experiments**
  - Uses changes in the environment that create roughly identical treatment and control groups for studying the effect of that environmental change.
  - LIMITATION: Provides an estimate of the causal impact of a particular treatment; tells how outcomes change but not why
- **Structural modeling**
  - Constrains the data through assumptions based on economic theory
  - LIMITATIONS: difficult to estimate, results are only as good as the assumptions

Cross-Sectional Analysis: Effects of Sports Participation

Research Question: Does participating in sports in high school cause kids to complete more schooling than they otherwise would have?

Regression analysis of years of schooling explained by a

$$(Years\ of\ Schooling)_{i,s,t} = \alpha + \beta(Sports\ Participant)_{i,s,t} + x_{i,s,t}\,\delta + \varepsilon_{i,s,t}$$

What other independent variables do we want to include in this regression?

- Personal characteristics: race, age, IQ, self-confidence
- Family characteristics: Parents education, whether living with both parents, siblings, family income, family gets the newspaper, family has a library card
- Community characteristics: state, urban or rural area
- School characteristics: proportion of teachers with masters degree, proportion of students who are disadvantage, drop-out rate, attendance rate

- Some of these independent factors are easier to measure than others
- The ones that can't be measured cause bias in our causal estimates if they are correlated with our variable of interest

Cross-Sectional Analysis: Effects of Sports Participation

Independent Variable	Female			Male		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Dependent Variable: Years of Education[f] (OLS)						
Athletics	.984	.632	.436	1.14	.862	.471
	(.117)	(.111)	(.098)	(.119)	(.112)	(.097)
Adjusted R-squared	.117	.260	.431	.165	.297	.485
Controls						
Demographics	ü	ü	ü	ü	ü	ü
Family characteristics		ü	ü		ü	ü
School characteristics		ü	ü		ü	ü
Ability/achievement			ü			ü

- High school athletes get more education than non-athletes
- Difference shrinks as we take account of the fact that athletes are different in a number of observable ways
- But might athletes be different in unobservable ways as well?

Summary from the cross-section

- People who play sports in high school get more schooling
  - The raw correlation between sports and schooling is contaminated by omitted variables that are correlated with sports participation and have important effects on schooling
  - Controlling for a wide-array of personal, family, schooling, and community characteristics still suggests that athletics has an important and positive relationship with education
    - Conditional on observables athletes get .4 of a year more schooling
- Possibilities
  - Participating in sports changes people causing them to complete more schooling: Causation
  - People who participate in sports are the type of people who would have done well anyhow: Correlation (driven by a third factor such as ambition, energy, etc that isn't captured properly by the control variables)

Quasi-Experiment: Title IX

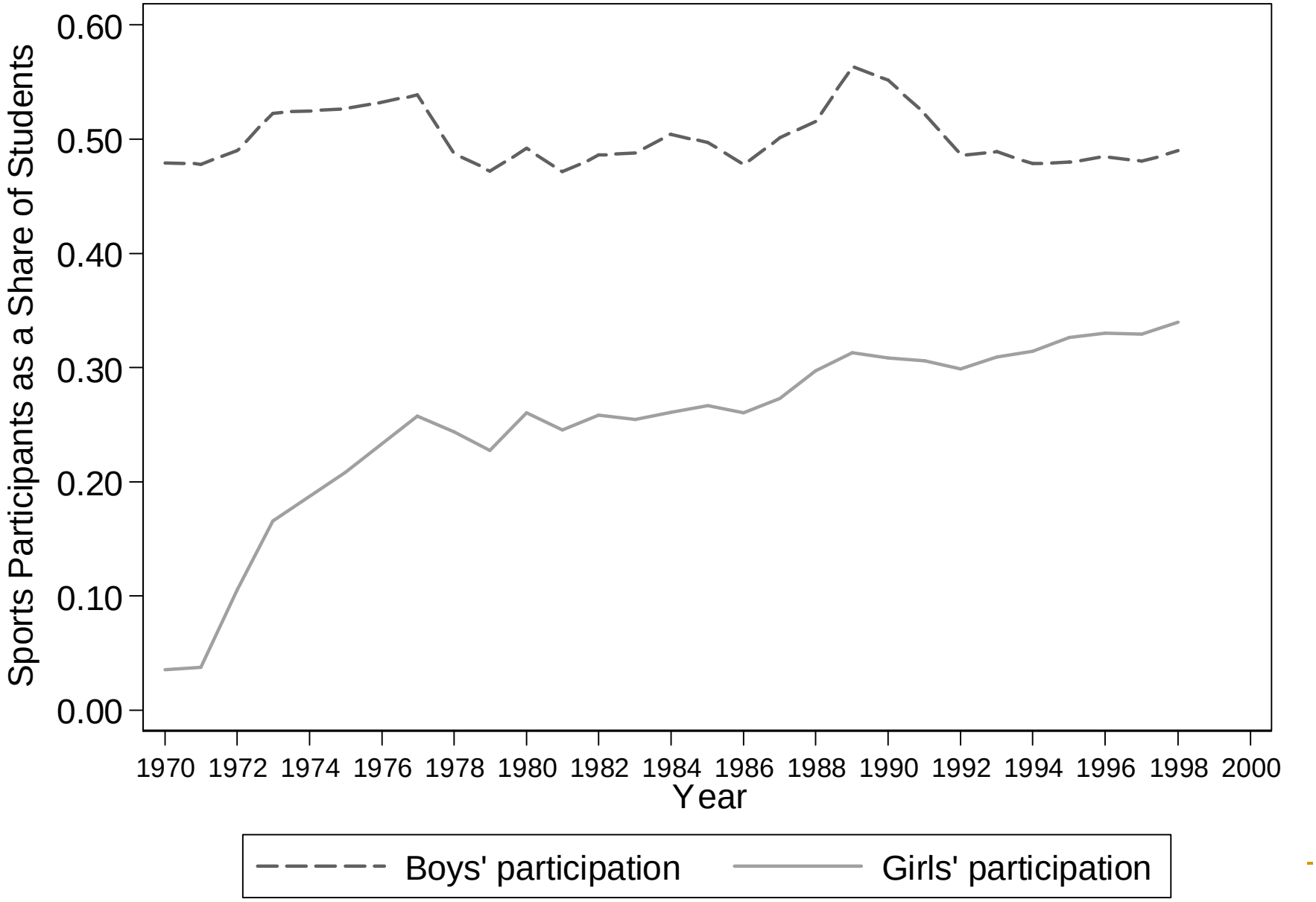
- In 1972 Congress enacted legislation that provided a mandate against gender discrimination in federally funded educational institutions.

Title IX of the Education Amendments to the 1964 Civil Rights Act:

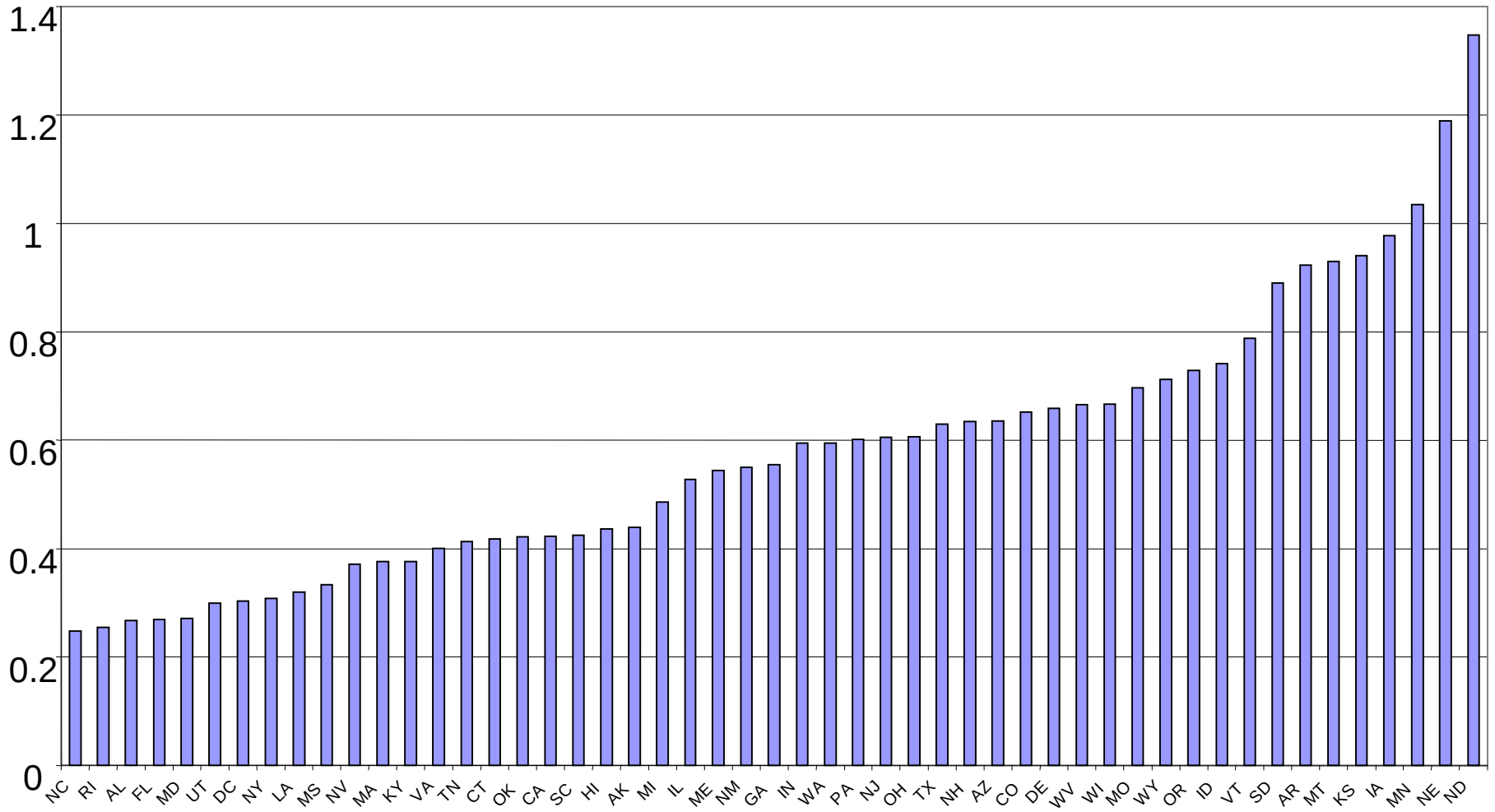
> ***No person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any educational program or activity receiving financial assistance.***

- Requires that girls be given the same ***opportunities*** as boys
- Policy gives a new generation of girls opportunities to play sports, but the size of the policy change varies across communities because some schools have large sports programs and others have small

Male and Female High School Sports Participation



Male Sports Participation Rate in 1971



Experimental Design

Whom does Title IX affect?

- Non-treated cohort ð women born prior to 1955
  - Those ages 25-34 in the 1980 census
- Treatment cohort ð those born after 1964
  - Those ages 25-34 in the 2000 census

How do States Differ?

- Individual states varied dramatically in terms of the participation rates of boys.
  - *States with higher levels ofmale participation needed more dramatic change in order to be in compliance with Title IX*.

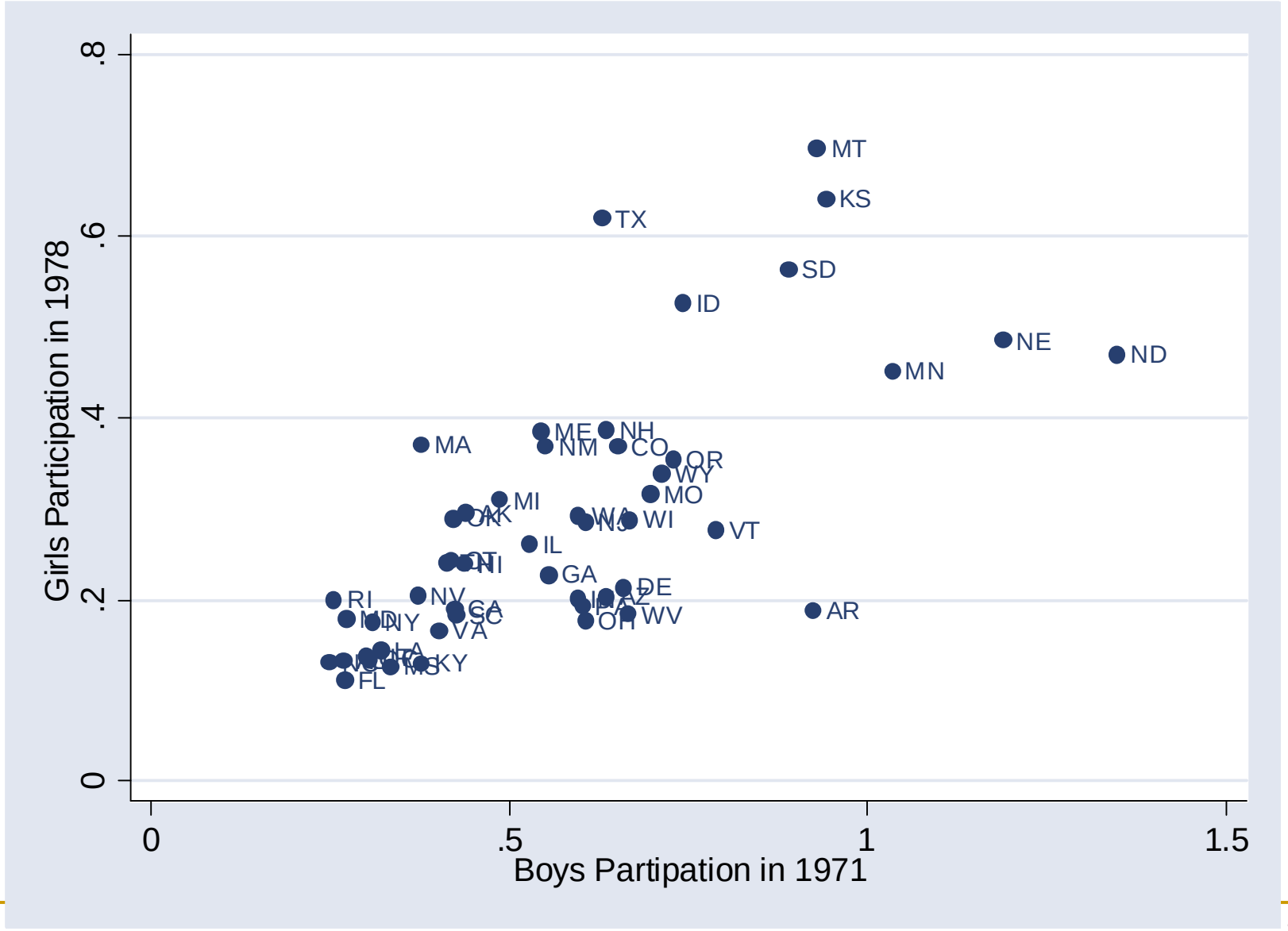
Difference-in-difference estimator

- combines time-series and cross-sectional analysis thereby controlling for cross-sectional differences and time-series differences where the:
  - Cross-sectional differences are stable over time
  - The time-series differences are the same across the cross-sectional groups

Isolating the exogenous change

- It is possible that some states increased girls' sports participation more than others because of other factors
- Solution: Intention-to-treat or instrumental variables
- Isolate the exogenous part of the policy change
  - **Instrumental Variable**
    - A variable that is correlated with the assignment to treatment but that does not solely determine the assignment
    - **Must** be uncorrelated with the error in the outcome equation
- Instrument for female participation after Title IX with male participation prior to Title IX
  - pre-Title IX (1971) level of boys participation interacted with the individual's cohort generates my experimental variation

Girls participation in 1978 versus Boys in 1971



Regression Specification

In other words we want to run:

$$(Outcome)_{i,s,t}$$

$$=$$

$$\alpha + \beta(\ Athl.\ Opportunity\ )_{s,t} + x_{i,s,t}\,\delta + Year\ FE + State\ FE + \varepsilon_{i,s,t}$$

Data:

- 5% Public Use Micro Sample (PUMS) of the 1980 & 2000 Census of Population
- 25-34 year old women who report having completed at least 10th grade.

Does increasing opportunities for women to play sports change female outcomes?

Is there a change in X across cohorts predicted by the athletic opportunity available to women during high school (identified by state of birth) instrumented by the male participation rate in that state prior to Title IX?

Where X is:

- ✓ Educational attainment
- ✓ Labor force participation
- ✓ Occupational Choice

Title IX and Education

- Female educational attainment rises with the opportunity to play sports
  - States with a 10 percentage point greater increase in the female athletic participation rate in a state had:
    - Increase in overall educational attainment of .039 years
    - Increase in probability of some post-secondary education of 1.3 percentage points
    - Increase of 0.8 percentage points in the probability of getting at least a college degree
  - Multiply by 3 to get aggregate effects of Title IX (raised female participation roughly by 30 percentage points)
  - Compare to total rise of 0.7 years in female education over this period

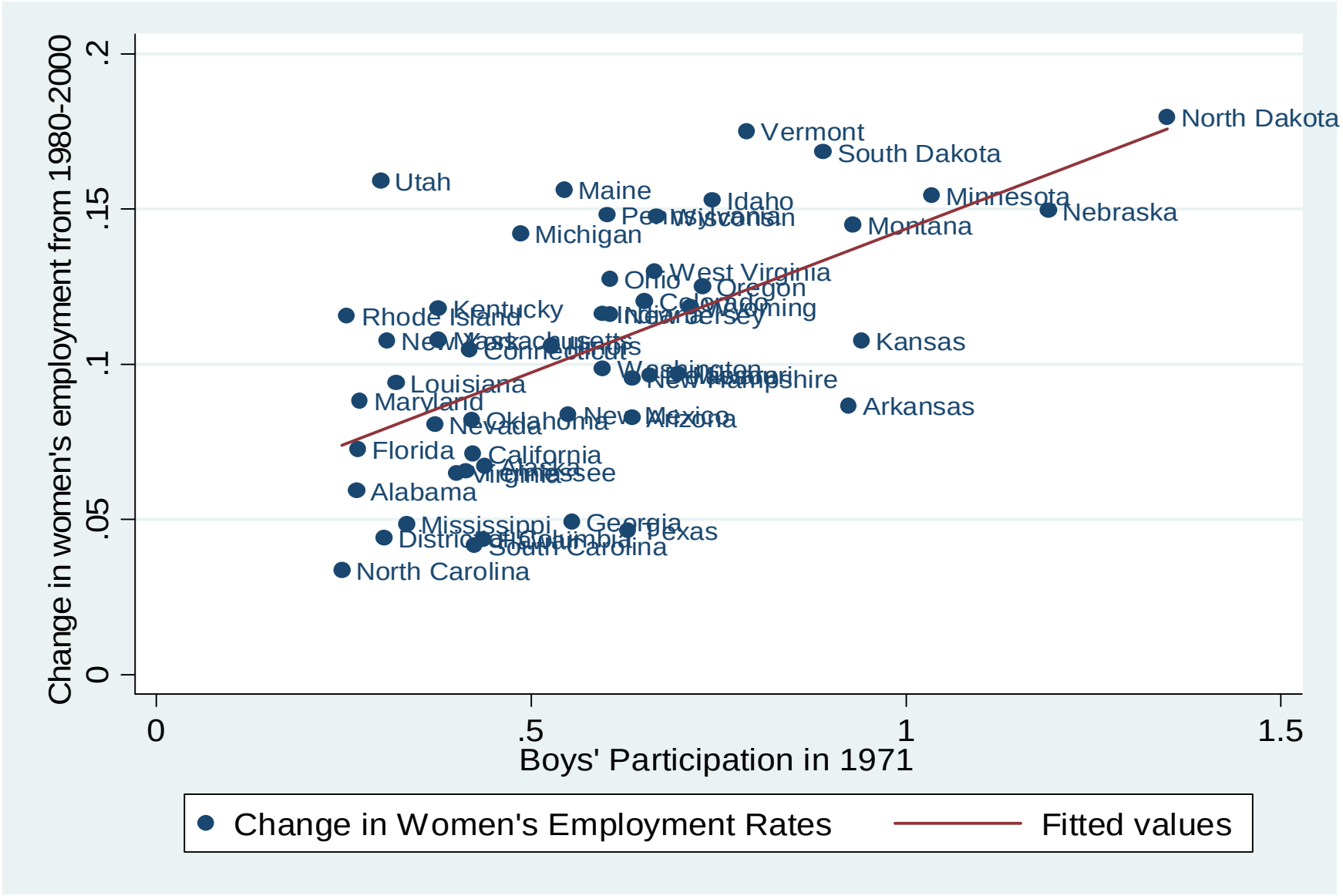
Data

- Data on sports participation comes from the *High School Athletic Participation Survey* conducted by the *National Federation of State High School Associations*.
- The state high school associations, which include the 50 states plus the District of Columbia, report annual data on the number of participants in each sport played in their particular state.
- Data is by gender, by sport
- Aggregate numbers of participants are thus given for each state
- “Participation rate” is number of participants divided by number of students

Title IX and Employment

- Female labor force participation rises with the opportunity to play sports
  - States with a 10 percentage point greater increase in the female athletic participation rate in a state had an increase in employment of 1-2 percentage points
  - The growth in employment happened disproportionately in male occupations compared with female occupations for states with more opportunities for girls to play sports
  - The variation across states in the opportunity to play sports can explain all of the variation across states in the growth in women working in sports
- Multiply by 3 for aggregate effect of Title IX
- Compare to an overall rise in labor force participation by this group of 10 percentage points

Change in Women's Employment and the Pre-Title IX Male Sports Participation Rate



Robustness Checks

- Compare those who stayed in the state to those who left
- Instrument using weather and school size instead of male participation
- Examine differences in female high school behavior before Title IX (probability of taking math, etc)
- Examine transition cohort
- Effects on boys

INSTRUMENTAL VARIABLES ESTIMATES

THE EFFECTS OF FEMALE ATHLETIC PARTICIPATION ON EDUCATIONAL ATTAINMENT

	(1)	(2)	(3)
Wald Estimator (IV) Causal Effect of Sports Participation[a]	.392** (.180)	.387** (.172)	.550*** (.191)
Reduced Form Results: Differential Effects of Title IX on Years of Education, by State[b]	.168*** (.066)	.168*** (.064)	.216*** (.051)
First-Stage Results: Changes in Female Sports Participation Generated by Title IX[c]	.429*** (.059)	.434*** (.052)	.392*** (.071)
Controls (All regressions include as controls a saturated set of dummy variables for state of birth, year of sample, age, race, and ethnicity.)			
Economic conditions	No	Yes	Yes
Year*Region of Birth Fixed Effects	No	No	Yes
Observations (standard errors are clustered at the level of 100 state-cohort cells)	1,544,870	1,544,870	1,544,870

Effects of the Title IX Generated Rise in Female Athletic Participation on the Probability of Attending College

	Completed high school		Attended at least 1 year of college		Attended at least four years of college			Post-college education
	(1)	(2)	(1)	(2)	(3)	(4)	(5)	(6)
Wald Estimator (IV) Causal Effect of Sports Participation	.002 (.006)	.057*** (.023)	.133*** (.046)	.121*** (.041)	.076** (.038)	.054* (.031)	.013 (.014)	.044** (.019)
Reduced Form Results: Differential Effects of Title IX on College-going propensities, by State [b]	.004 (.014)	.029*** (.009)	.057*** (.015)	.046*** (.011)	.032** (.015)	.021* (.011)	.006 (.006)	.017*** (.005)
First-Stage Results: Changes in Female Sports Participation Generated by Title IX [c]	.429*** (.059)	.392*** (.071)	.429*** (.059)	.392*** (.071)	.429*** (.059)	.392*** (.071)	.429*** (.059)	.392*** (.071)
Controls (All regressions include as controls a saturated set of dummy variables for state of birth, year of sample, age, race, and ethnicity.)								
Economic conditions	No	Yes	No	Yes	No	Yes	No	Yes
Year*Region of Birth Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations (standard errors are clustered at the level of 100 state-cohort cells)	1,544,870	1,544,870	1,544,870	1,544,870	1,544,870	1,544,870	1,544,870	1,544,870

Conclusion

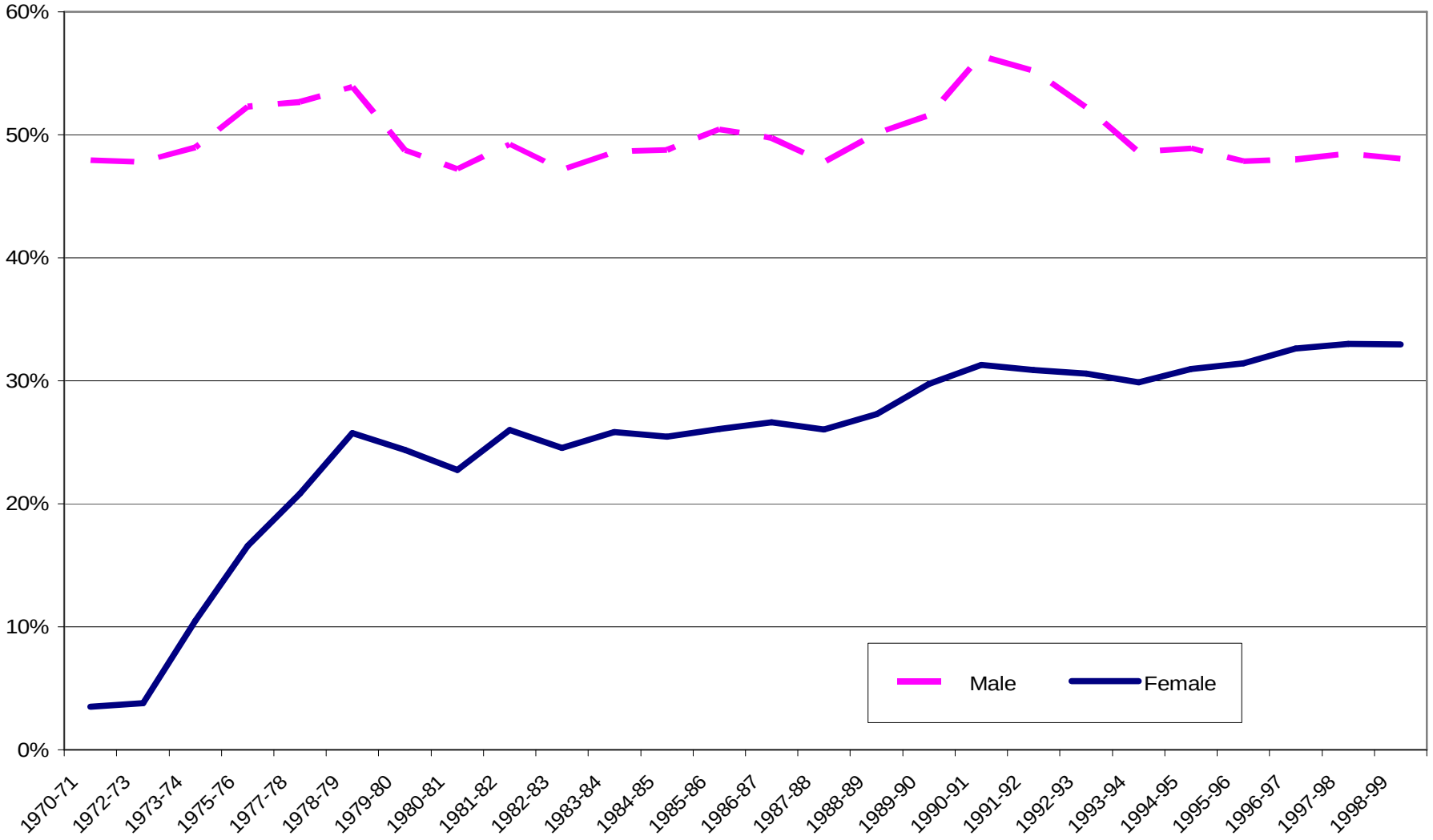
- "Quasi-experimental" evidence that high school sports yield traits that are productive in the labor market for the marginally treated woman
- States with higher male participation rates prior to Title IX have greater increases in female educational attainment, labor force participation, and have more female entrance into previously male-dominated high-skill jobs
- Title IX explains about 20 percent of the increase in women's education and about 40 percent of the rise in employment for 25-to-34-year-old women.

Time Series Analysis: Title IX

- The law passes in 1972 and states are given until 1978 to come into compliance
- In 1992 Franklin v. Gwinnett County Public Schools makes schools liable for violations and allows for monetary damages
- Discrete breaks: 1972, 1978, 1992

Male and Female High School Sports Participation

(As a Percentage of Female High School Enrollment)



Cross-Sectional Relationship

Effects of High School Participation in Extra Curricular Activities on Log Wages

Independent Variable	Female				Male			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel B: Dependent Variable: Log of Wages[a] (OLS)							
Athletics[b]	.141	.106	.072	.076	.190	.142	.076	.073
	(.039)	(.039)	(.038)	(.039)	(.034)	(.034)	(.034)	(.033)
Non-vocational Clubs[b]				-.018				.010
				(.039)				(.033)
Vocational Clubs[b]				-.032				-.061
				(.040)				(.036)
Adjusted R-squared	.092	.140	.210	.211	.142	.202	.275	.277
Controls[c]								
Demographics	ü	ü	ü	ü	ü	ü	ü	ü
Family characteristics		ü	ü	ü		ü	ü	ü
School characteristics		ü	ü	ü		ü	ü	ü
Ability/achievement			ü	ü			ü	ü