Does School Spending Pay for Itself? Long-Run Effects on Personal Income

John Klopfer (HKU) and Corbin Miller (IRS)*

November 16, 2024
NTA Conference
PRELIMINARY: DO NOT DISSEMINATE

*The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and do not necessarily reflect the views or the official positions of the U.S. Department of the Treasury. Any taxpayer data used in this research was kept in a secured Treasury or IRS data repository, and all results have been reviewed to ensure that no confidential information is disclosed.

Introduction

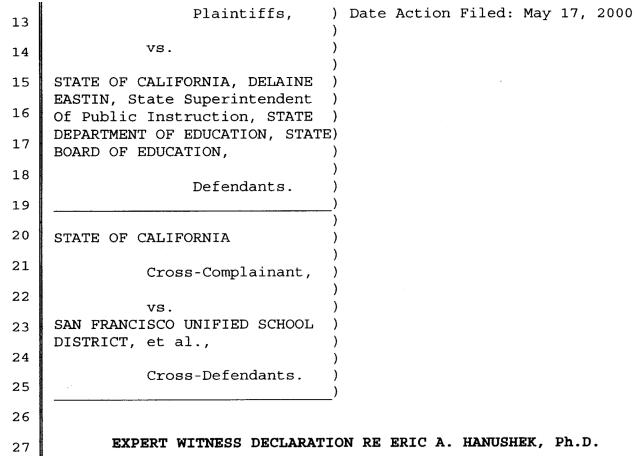
Does money matter in schools?

- Consequential: K-12 is 4 percent of GDP, 5 percent of employment, 18 percent of state and local government spending
- Debated since 1960s: Coleman (1968), Burtless (1996), Hanushek (2003, 2006), Hanushek and Handel (2023), Jackson and Mackevicious (2024)
  - Older work: causality is a challenge and biases are unsigned (e.g. family location choice and private investment, versus need-based funding)
  - Newer work: causality is addressed but challenges remain including power, sample size, spillovers, external validity

Introduction

Does money matter in schools?

- Policy venue: *state* courts and legislatures



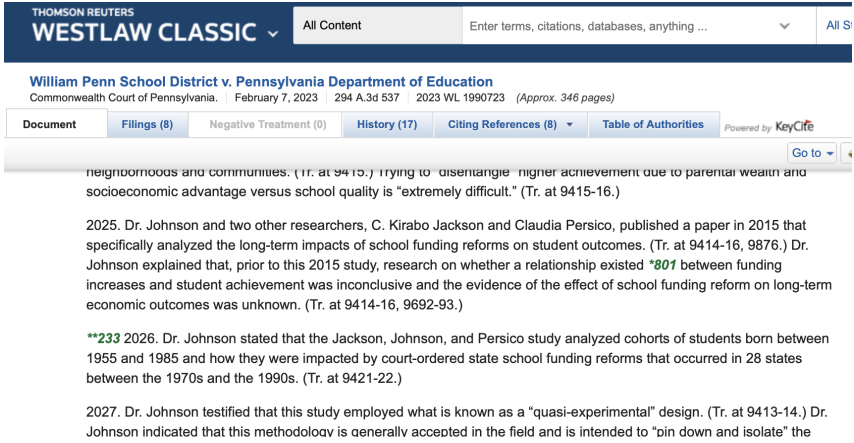
Plaintiffs,) Date Action Filed: May 17, 2000
)
vs.)
)
STATE OF CALIFORNIA, DELAINE)
EASTIN, State Superintendent)
Of Public Instruction, STATE)
DEPARTMENT OF EDUCATION, STATE)
BOARD OF EDUCATION,)
)
Defendants.)
_____)
)
STATE OF CALIFORNIA)
)
Cross-Complainant,)
)
vs.)
SAN FRANCISCO UNIFIED SCHOOL)
DISTRICT, et al.,)
)
Cross-Defendants.)
_____)

EXPERT WITNESS DECLARATION RE ERIC A. HANUSHEK, Ph.D.

Introduction

Does money matter in schools?

- Policy venue: *state* courts and legislatures



Introduction

Does money matter in schools?

- Policy venue: *state* courts and legislatures
- Policy alternatives: redistribution ("equity") or aggregate spending ("adequacy")
- Policy evidence: redistribution
  - Lafortune, Rothstein, and Schanzenbach (2018): redistribution and test scores
  - Jackson, Johnson, and Persico (2016), Jackson and Johnson (2019), Rothstein and Schanzenbach (2022): redistribution and adult earnings
  - Biasi (2023): redistribution and IGM
  - Redistribution appears to improve the relative standing of targeted populations, but not obvious that it is efficient

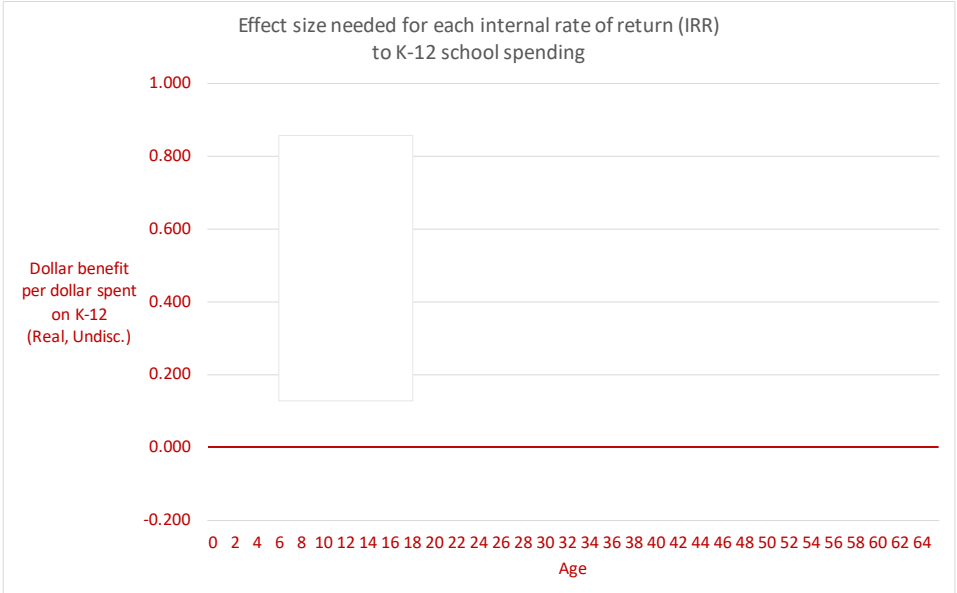
Today: evidence on aggregate spending...

Research questions

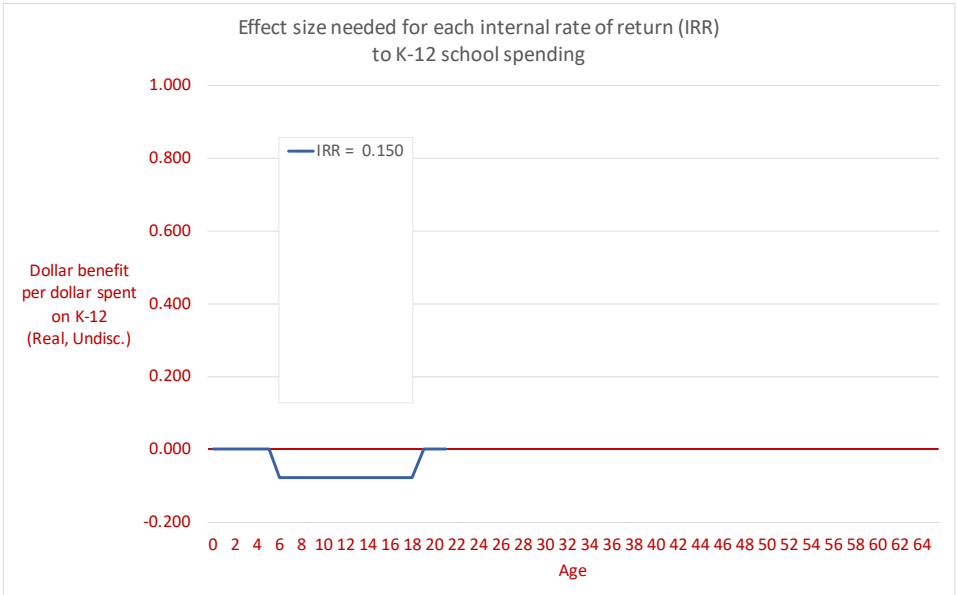
Three substantive cost-benefit questions:

1. When states spend more on K-12, do students earn more as adults?
2. If students earn more, do benefits outweigh costs?
3. If benefits outweigh costs, do they come soon enough that the investment can be financed at a gain?

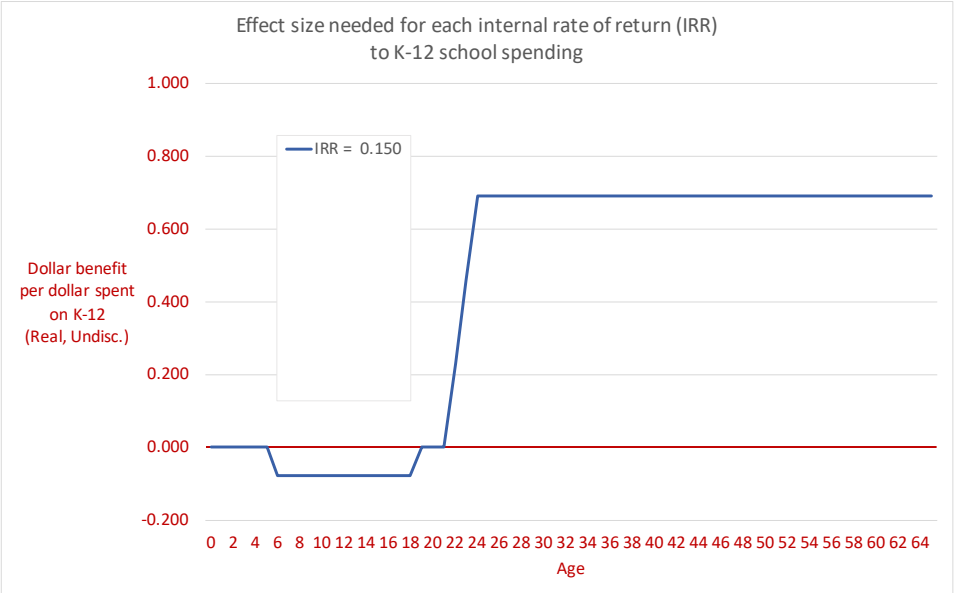
Research questions



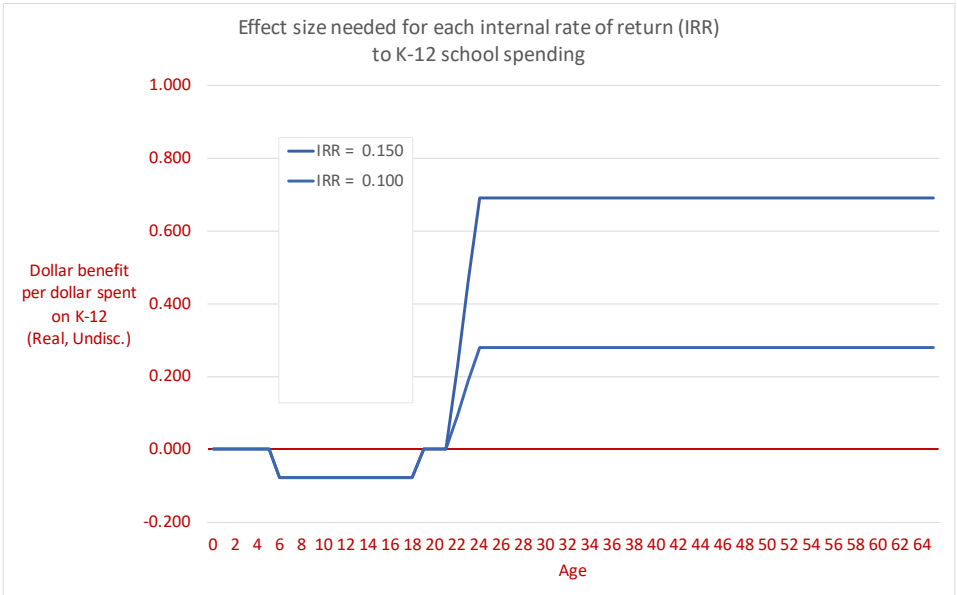
Research questions



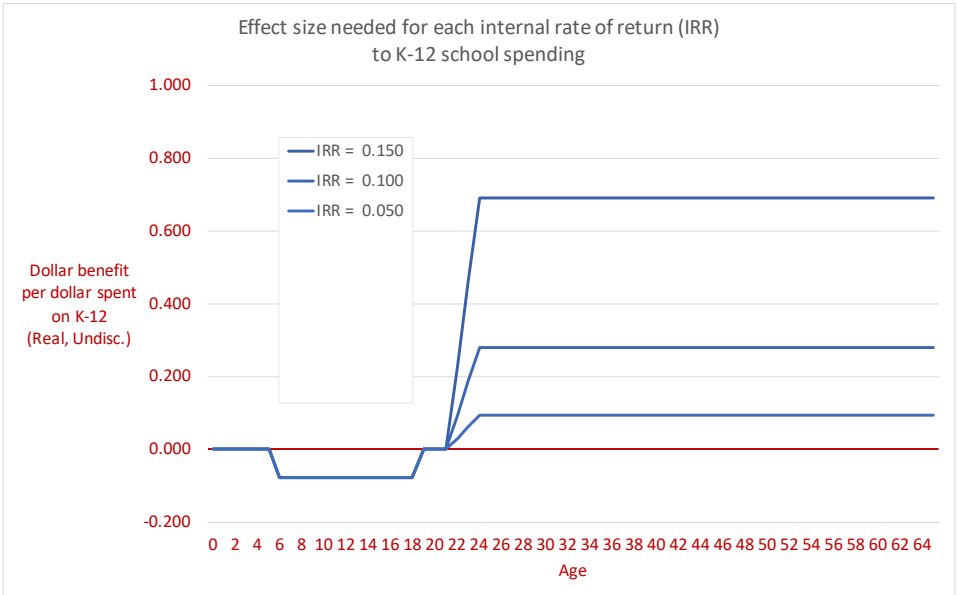
Research questions



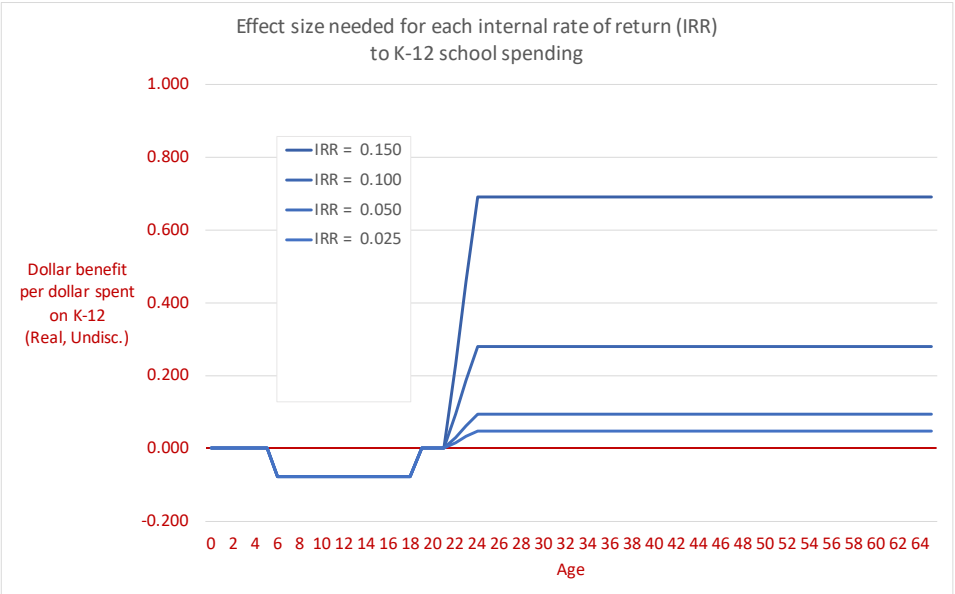
Research questions



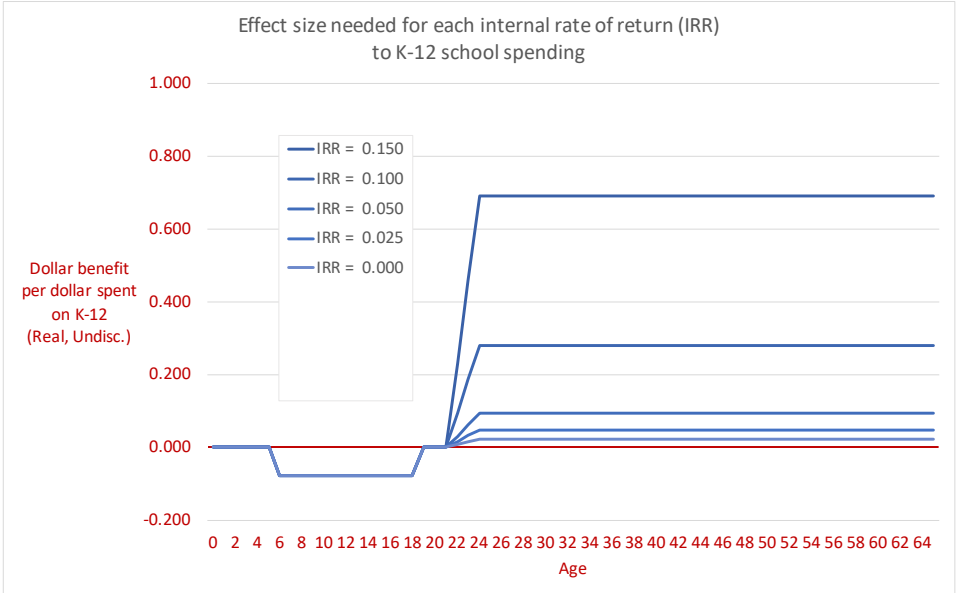
Research questions



Research questions



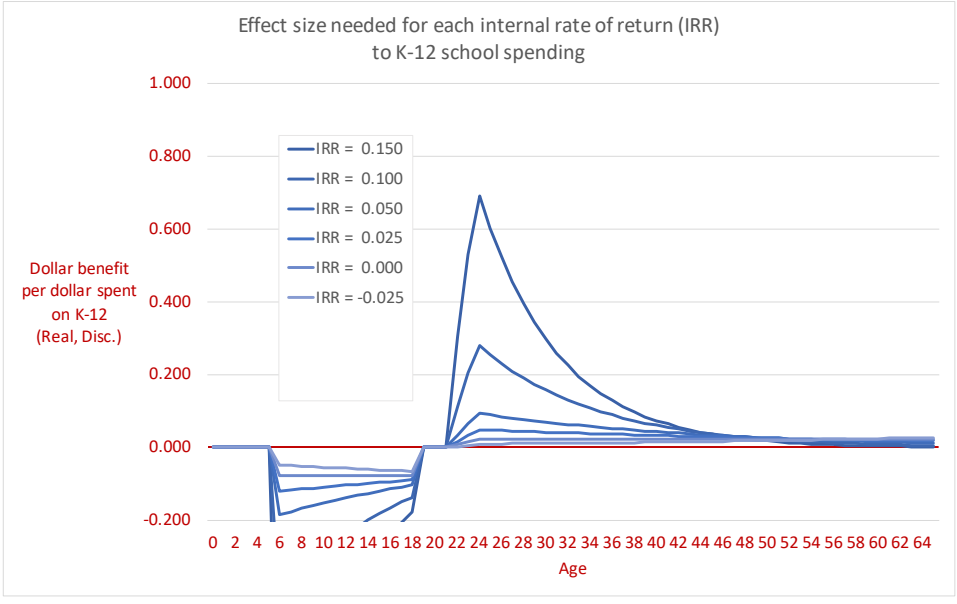
Research questions



Research questions



Research questions



This paper: Goal and practical contributions

Dollar-for-dollar, age-by-age, precise IV estimates of spending impact on earnings

X: Historical data on school spending

- Digitized annual data 1950-2020 (71 years) at the state-by-year-level
- Cumulative spending for cohorts born 1944-2002, and graduating 1962-2020

Y: Administrative data for students' adult outcomes

- Full-population tax records from IRS for 1999-2021
- Earnings. Also: employers and unemployment; tuition, grants, loans; mortgage, family formation, migration

Z: Strong instruments for school spending

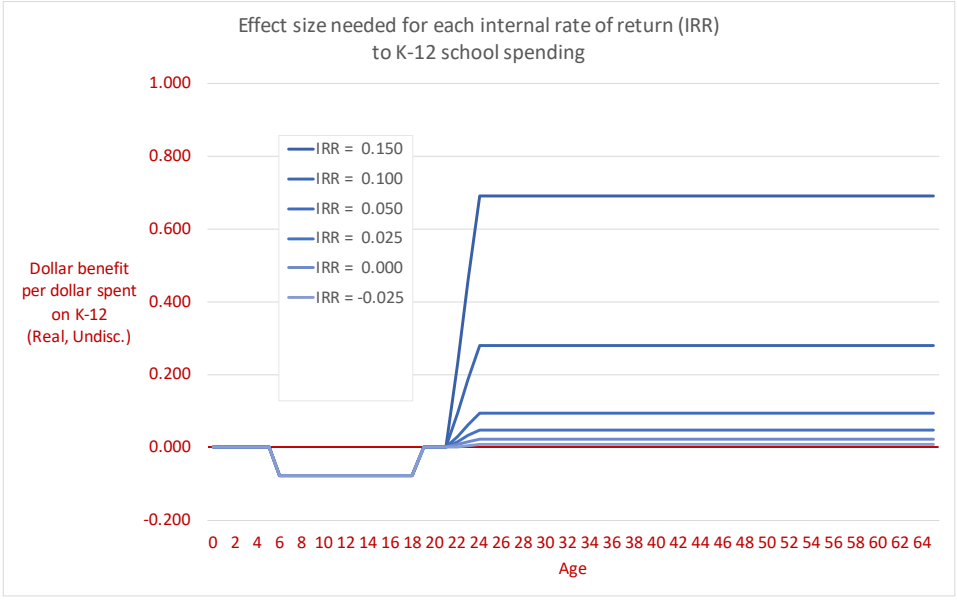
- Exogeneity: State-by-cohort timing of court-ordered school finance reforms
- Relevance: LASSO to avoid 'weak' court orders

This paper: Strengths

Estimates directly address the substantive cost-benefit questions (aggregate benefit, break-even, rate of return)

- Dollar-for-dollar: no assumptions on cost and benefit distribution, or weighting
- Age-by-age: no assumptions on benefit timing
- Precise: std. error 3-4 cents

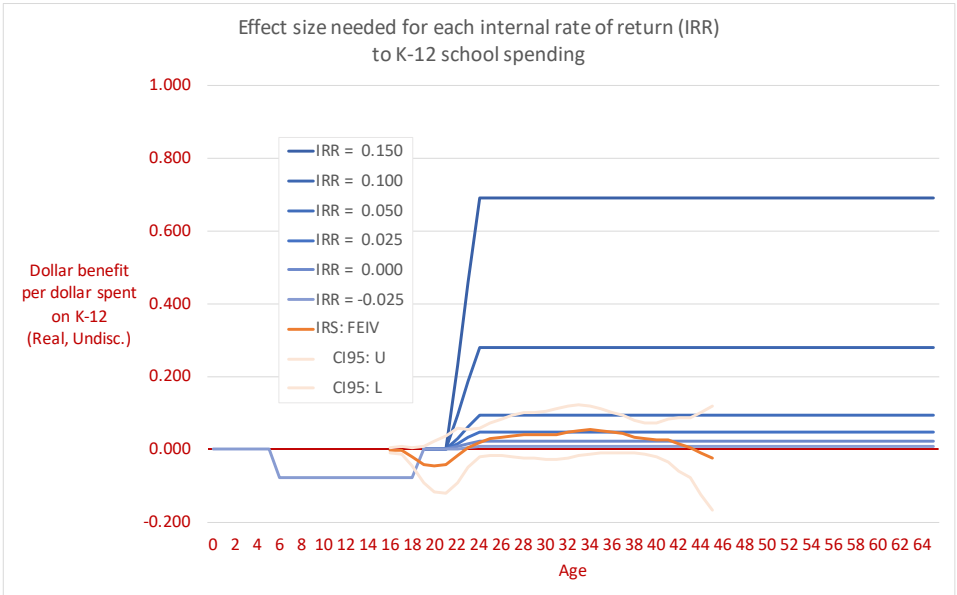
This paper: Rule out large returns



This paper: Rule out large returns



This paper: Rule out large returns



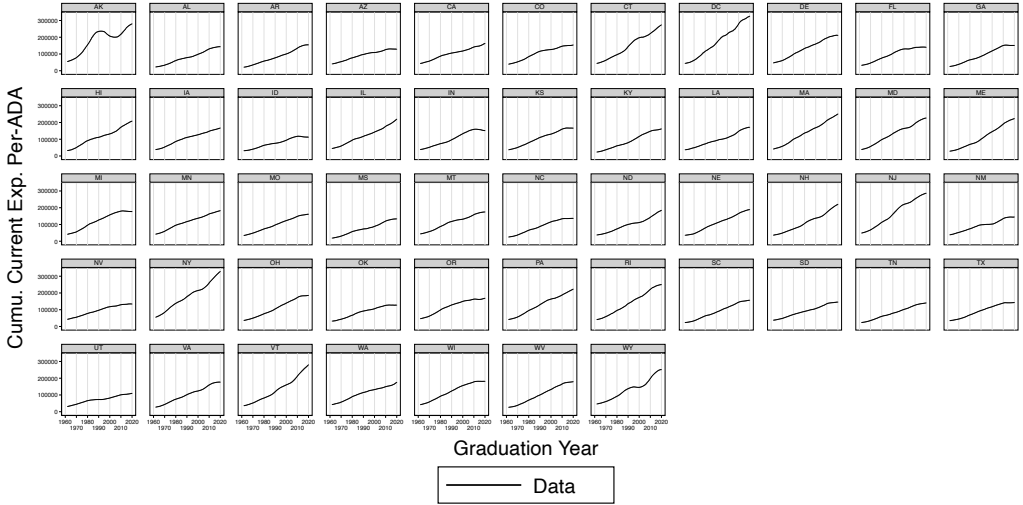
This paper: Reminders

Estimates address *only* the substantive cost-benefit questions: we weigh policy costs against benefits of court-ordered reforms

- not all benefits that matter are observed
- not all beneficiaries that matter are observed
- money may do more if targeted differently
- money may be targeted for equity, not efficiency: policy may harm some to help others (important not to conflate benefit-for-treated with harm-for-control)

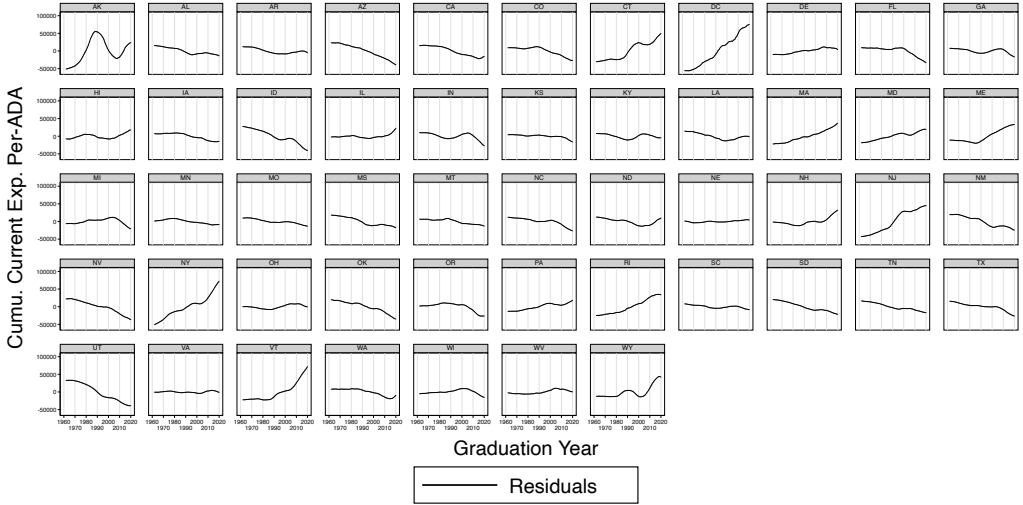
TAKEAWAY: results DO NOT show money can't or doesn't ever matter; results DO show that as-spent, in aggregate, court-ordered K-12 spending did not pay for itself

School spending



Graphs by Postal state abbreviation codes

School spending, state and year FE removed



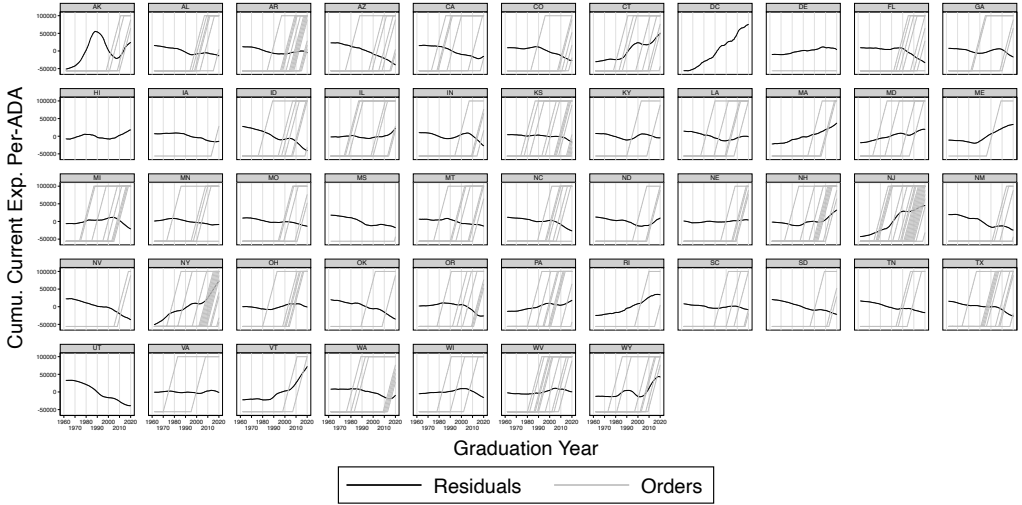
Graphs by Postal state abbreviation codes

Identification strategy: court orders and school finance reforms

Court orders:

- Families or school districts sue for funding under state constitution
- State judges order legislature and governor to change and fund formula (or don't)
- Legislature and governor comply (or don't)

Identification strategy: 50 states and DC (285 candidate events)



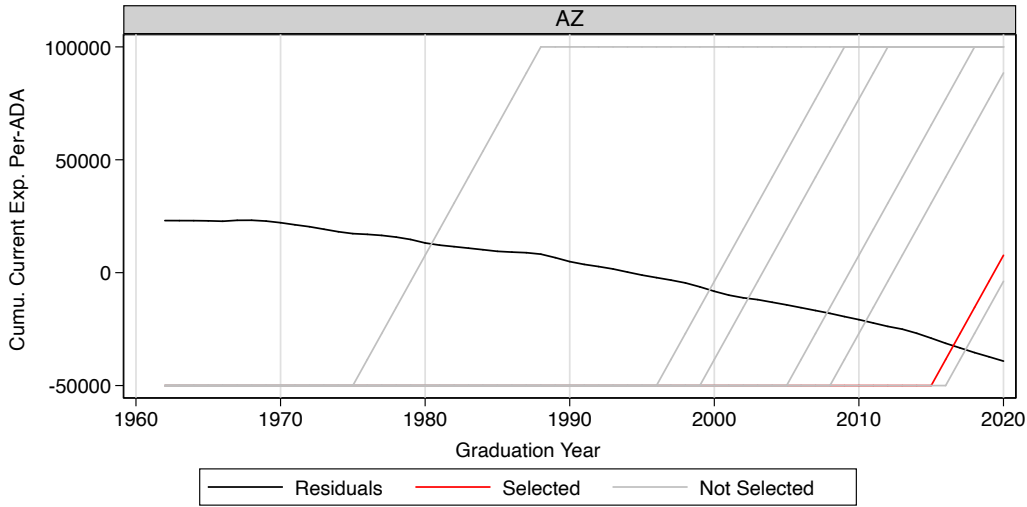
Graphs by Postal state abbreviation codes

Standard event selection

Satisfying the IV relevance assumption: When there are many orders

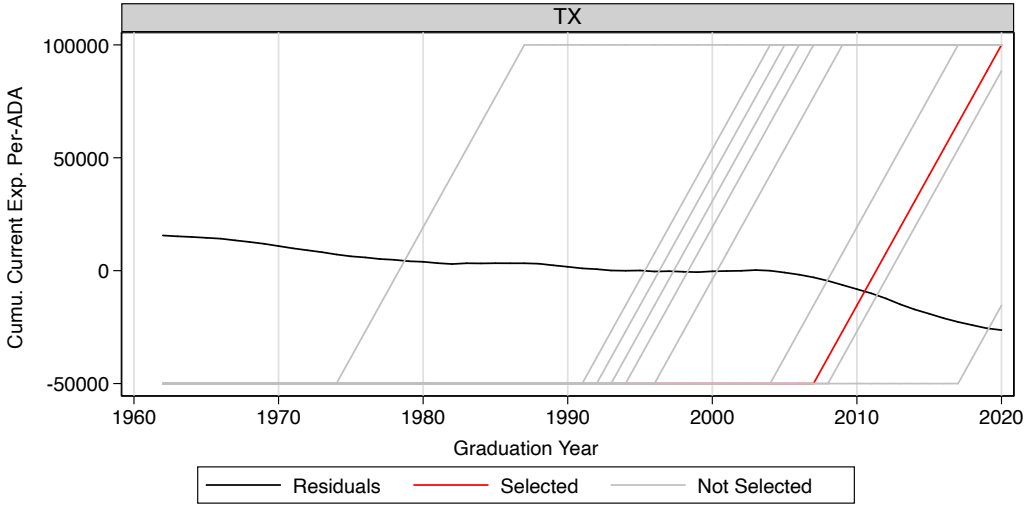
- Liscow's (2018) approach at the state-by-year level: selection of the "biggest" reform event in each state using Bai's (1997) test for structural breaks; following Lafortune, Rothstein, and Schanzenbach (2018), who caveat that...
  - ...if "each court order...was ineffective...our method will nevertheless pick one"

Standard event selection: Arizona (1/7 candidate events)



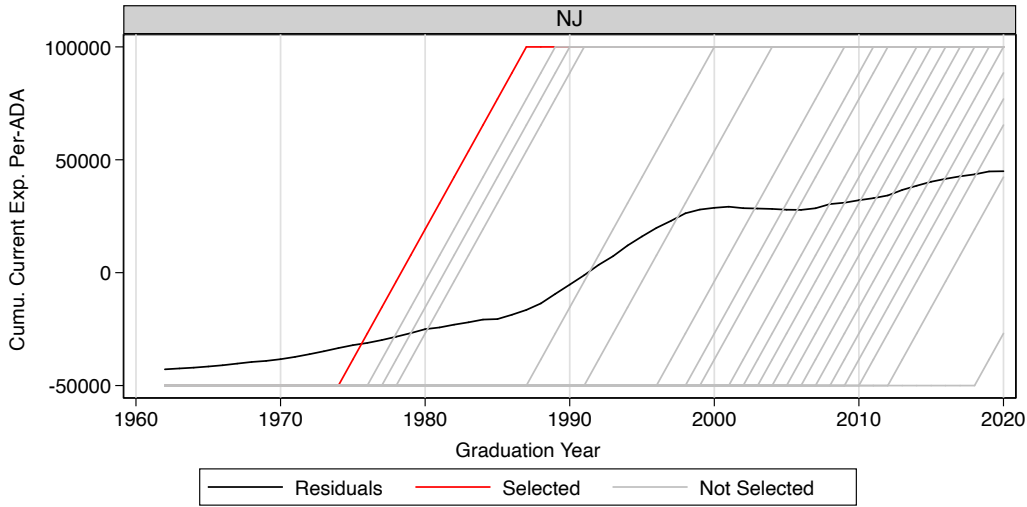
Graphs by Postal state abbreviation codes

Standard event selection: Texas (1/10 candidate events)



Graphs by Postal state abbreviation codes

Standard event selection: New Jersey (1/21 candidate events)

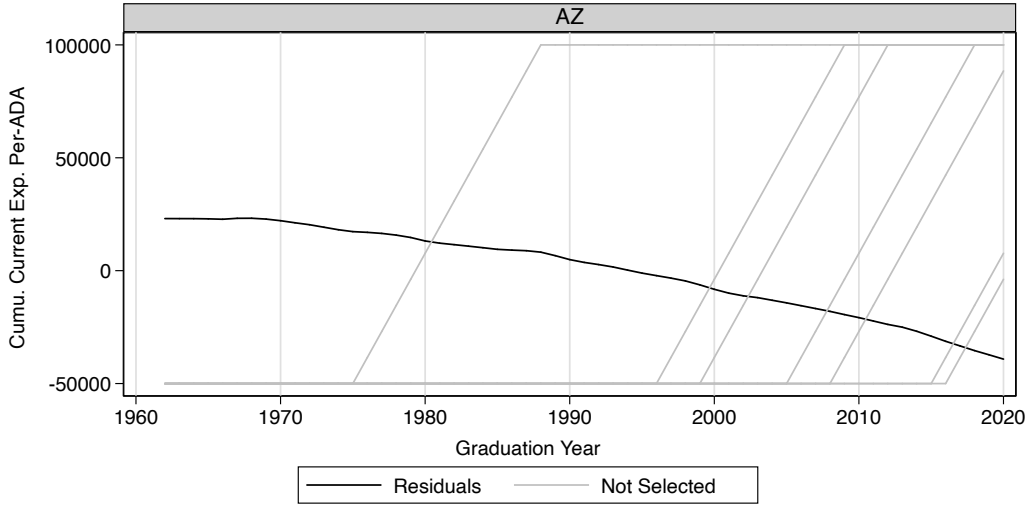


Our event selection

Satisfying the IV relevance assumption: When strongest is not strong

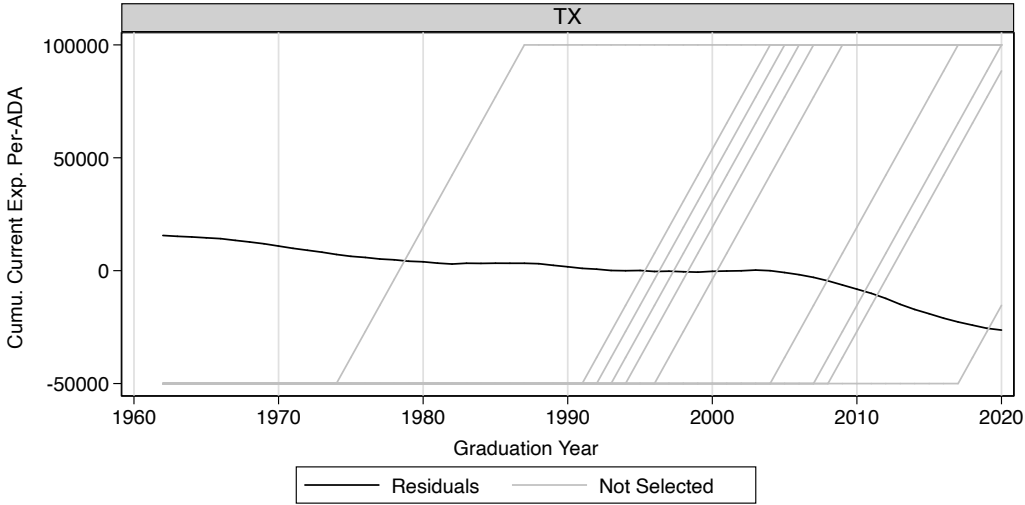
- Our approach at the state-by-cohort level: selection of the "big enough" reform events in each state using LASSO to both exclude irrelevant events and include multiple events where relevant
  - IV properties established in Belloni, Chen, Chernozhukov, and Hansen (2012)

Our event selection: Arizona (0/7 candidate events)



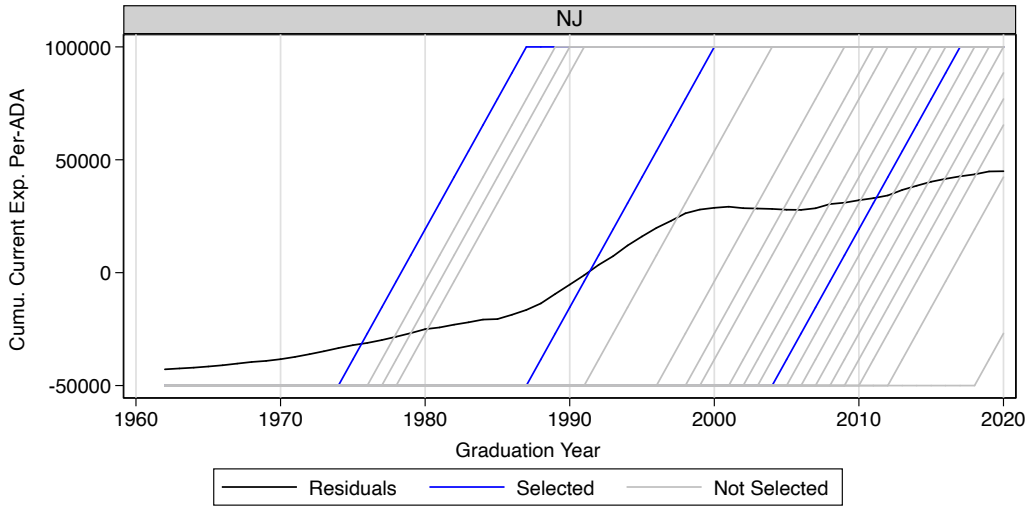
Graphs by Postal state abbreviation codes

Our event selection: Texas (0/10 candidate events)



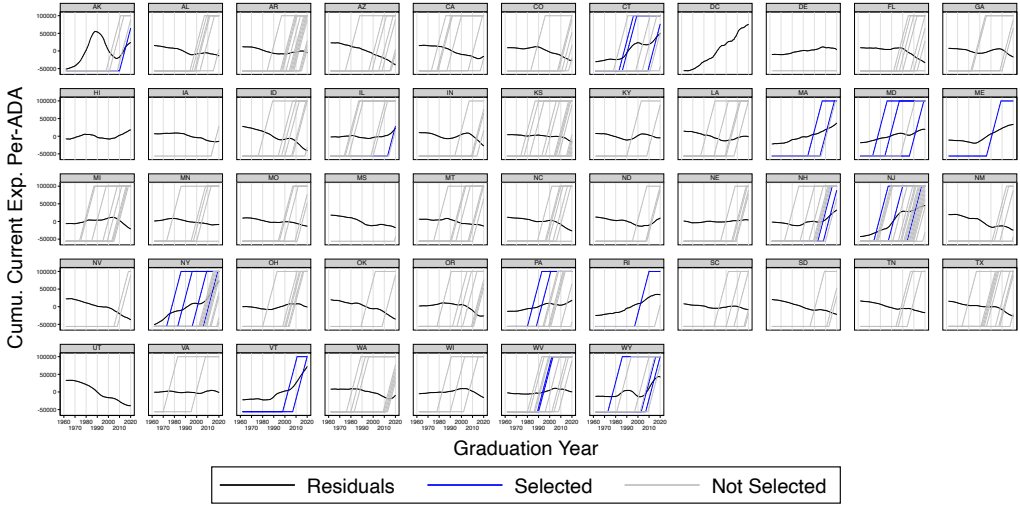
Graphs by Postal state abbreviation codes

Our event selection: New Jersey (3/21 candidate events)



Graphs by Postal state abbreviation codes

Our event selection: 50 states and DC (30/285 candidate events)



Graphs by Postal state abbreviation codes

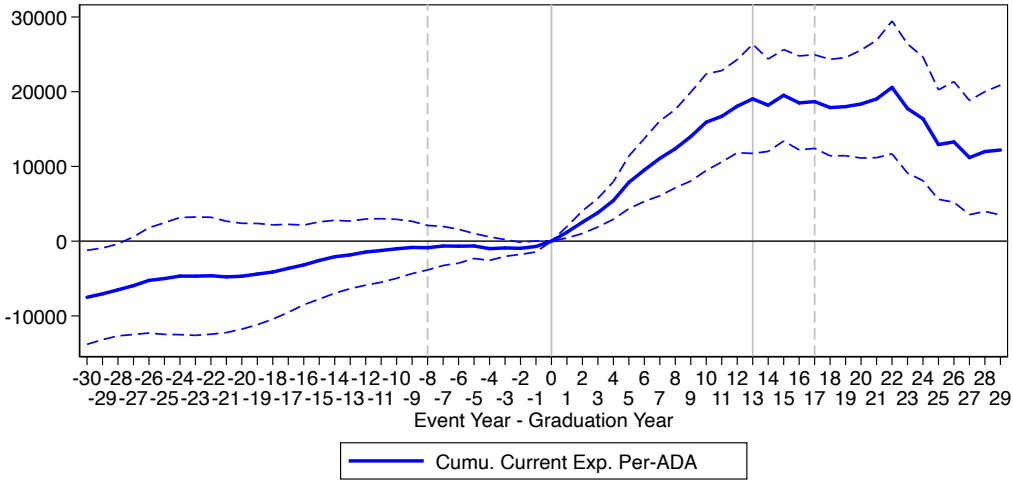
FE-IV specification: first stage, selected court orders and spending

First stage (event study)

$$E_{sg} = \sum_k \gamma_k \sum_{r \in R_s} 1[g + k = t_r] + \mu_s + \mu_g + \nu_{sg}$$

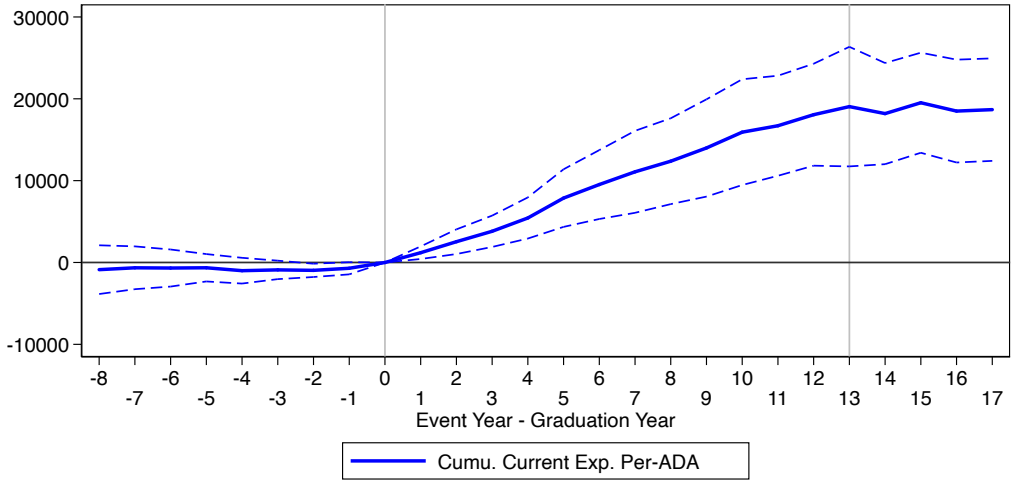
- E_{sg}: spending in state s, graduation cohort g
- t_r: year of reform $r \in R_s$, the set of selected court orders in state s
- γ_k: average effect of any given reform at lag k from the year of reform
- μ_s, μ_g: state and cohort fixed effects
- ν_s: time-varying error

First stage: event study



95 percent CI: standard errors clustered by state.
Specification includes state and birth year fixed effects. All lags included but only lags -30 through 29 shown.
Inner vertical lines show 0 and 13 years of exposure. Outer vertical lines show window for non-parametric instrument.

First stage: event study (excluded instrument lags)



95 percent CI: standard errors clustered by state.
Specification includes state and birth year fixed effects. All lags included but only lags -8 through 17 shown.
Vertical lines show 0 and 13 years of exposure.

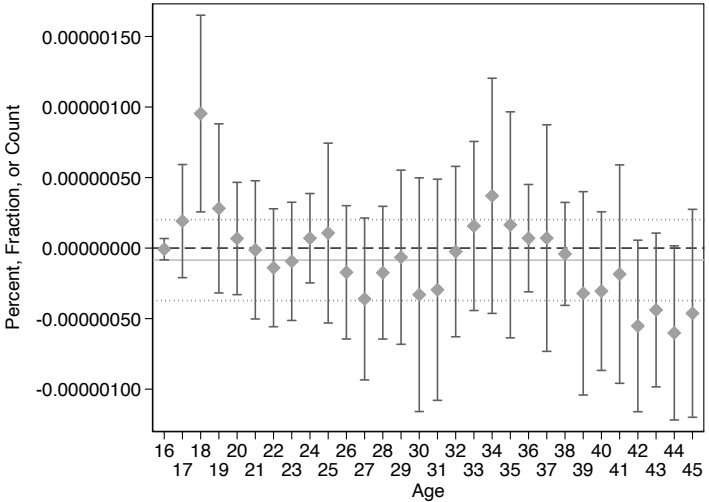
FE-IV specification: second stage, spending and adult outcomes

Second stage (by age)

$$
\begin{aligned}
Y_{isga} = & \beta_a E_{sg} \\
& + \sum_{k \notin K_{IV}} \beta_{ka} \sum_{r \in R_s} 1[g + k = t_r] \\
& + \alpha_{sa} + \alpha_{ga} \\
& + \epsilon_{isga}
\end{aligned}
$$

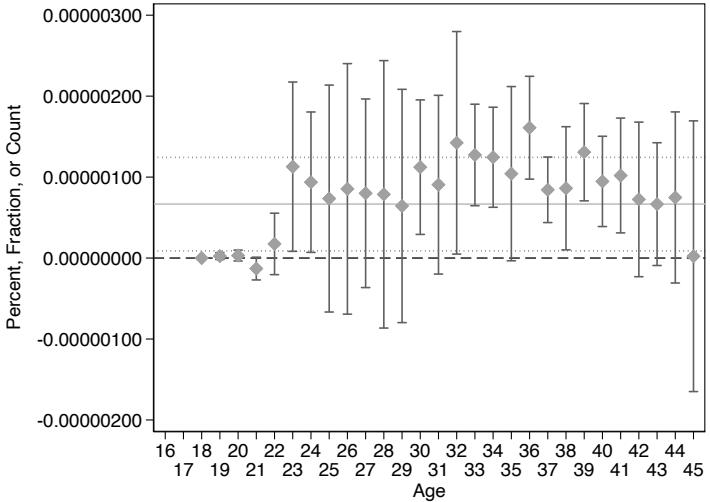
- Y_{isga}: outcome for individual i, in state s and graduation cohort g, at age a
- β_a: average effect of spending on outcome at age a
- K_{IV}: set of excluded instrument lags $-8 \leq k \leq 17$
- β_{ka}: average effect of control lags $k \notin K_{IV}$ on outcome at age a
- α_{sa}, α_{ga}: state and cohort fixed effects
- ϵ_{isga}: time-varying error

Results: K-12 spending raises higher-education attainment (ACS, 1:100)



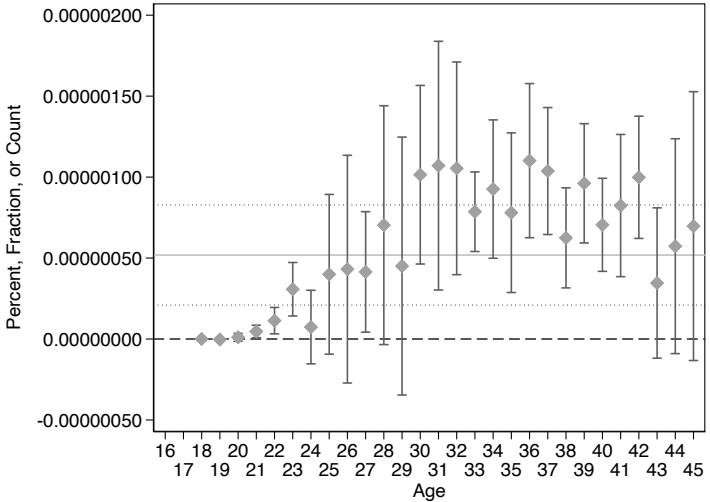
- High school degree +
  - No effect

Results: K-12 spending raises higher-education attainment (ACS, 1:100)



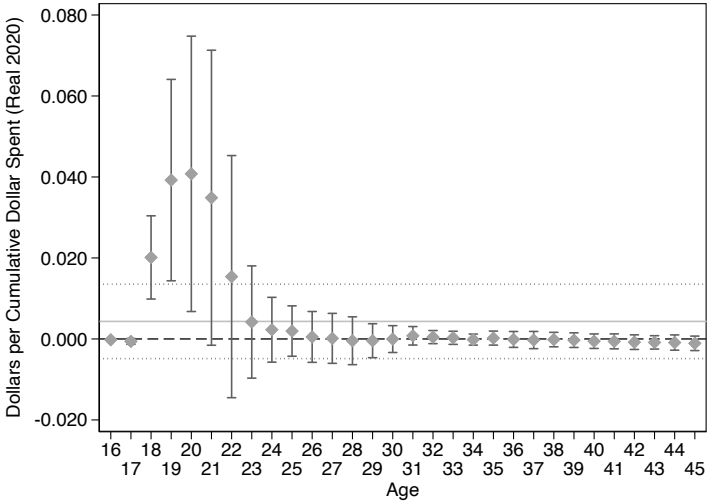
- High school degree +
  - No effect
- College degree +
  - 1.0 percent per $10,000

Results: K-12 spending raises higher-education attainment (ACS, 1:100)



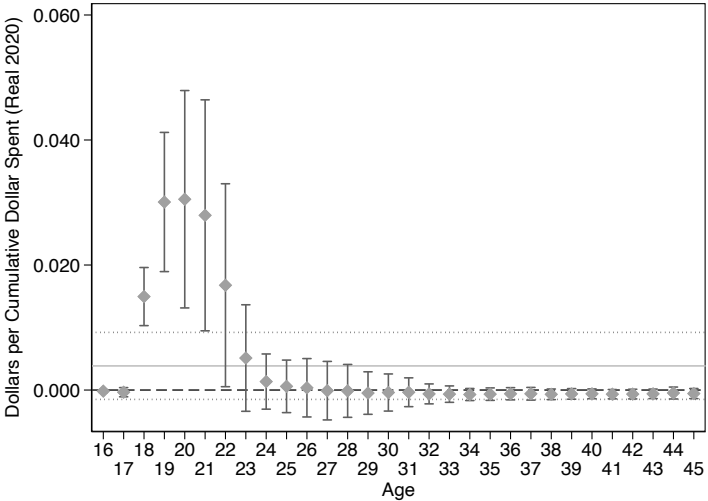
- High school degree +
  - No effect
- College degree +
  - 1.0 percent per $10,000
- Graduate degree +
  - 0.8 percent per $10,000
  - Mostly MA, not PhD or professional

Results: K-12 spending raises higher-education spending (IRS, 1:1)



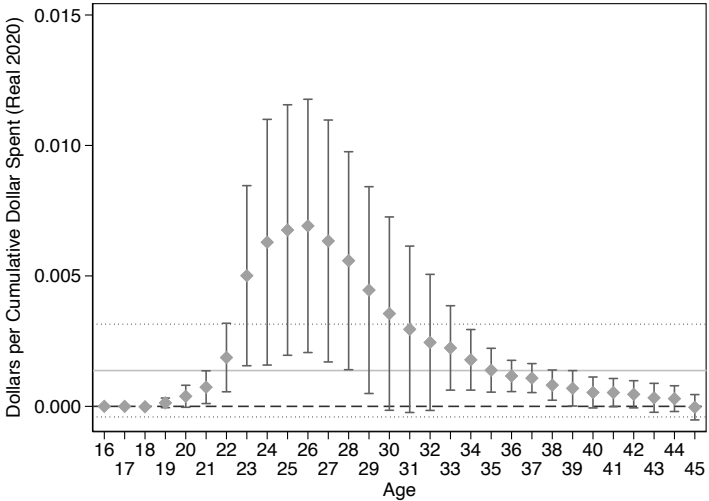
- Tuition (1098-T)
  - By 30: 15.8 cents per dollar

Results: K-12 spending raises higher-education spending (IRS, 1:1)



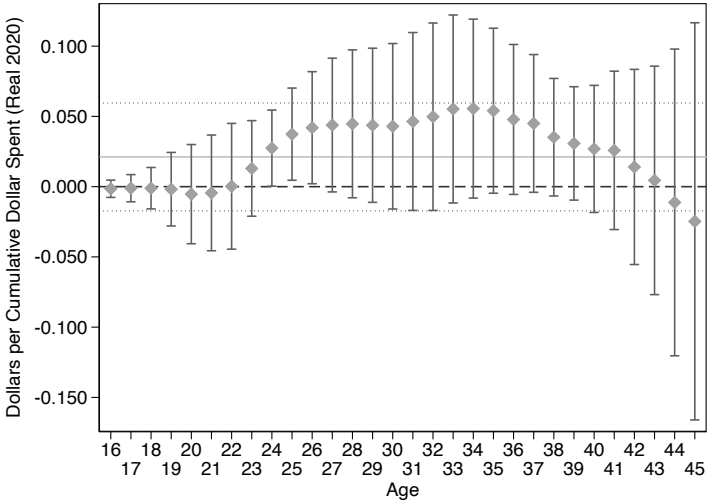
- Tuition (1098-T)
  - By 30: 15.8 cents per dollar
- Grants (1098-T)
  - By 30: 12.6 cents per dollar
  - Included in tuition

Results: K-12 spending raises higher-education spending (IRS, 1:1)



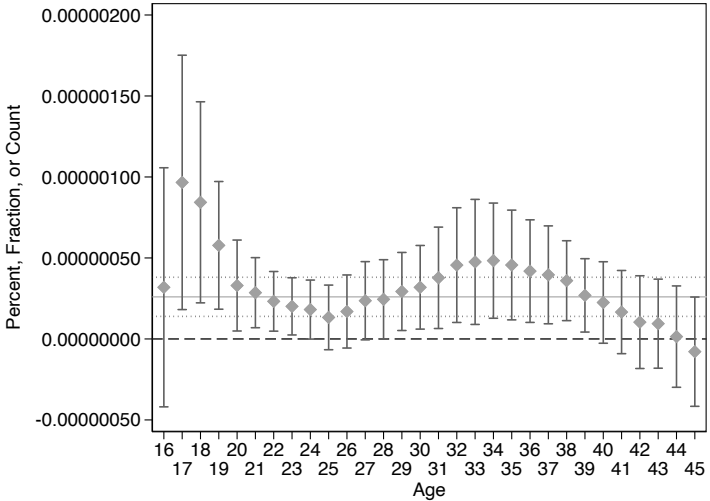
- Tuition (1098-T)
  - By 30: 15.8 cents per dollar
- Grants (1098-T)
  - By 30: 12.6 cents per dollar
  - Included in tuition
- Loan interest, fees (1098-E)
  - By 45: 6.5 cents per dollar
  - Not included in tuition

Results: K-12 spending raises earnings (IRS, 1:1)



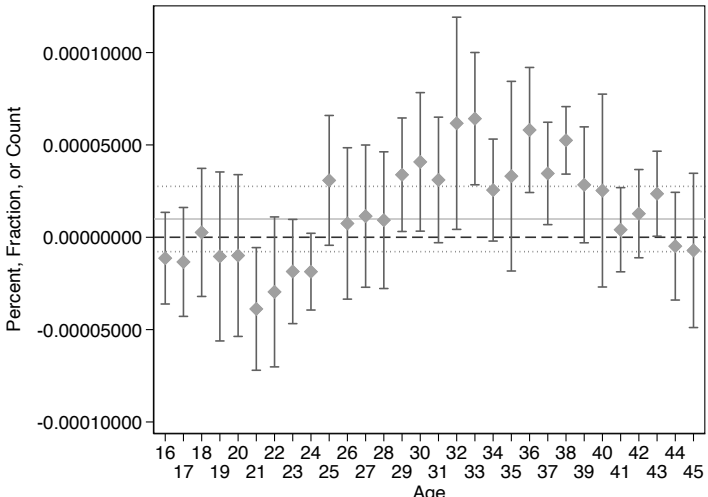
- Earnings (W-2)
  - By 30: up 28.0 cents per dollar [-24.6, 80.6]
  - By 45: up 73.5 cents per dollar [-79.8, 226.7]

Results: K-12 spending lowers leisure (IRS, 1:1)



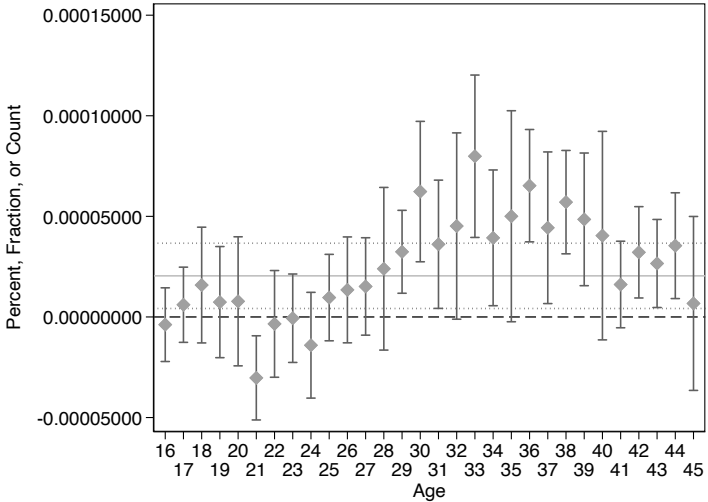
- Employment (W-2)
  - At 34: up 0.5 percent per \$10,000

Results: K-12 spending lowers leisure (ACS, 1:100)



- Weeks worked
  - By 30: down 0.1 weeks per \$10,000 (sum: 465)
  - By 45: up 4.3 weeks per \$10,000 (sum: 1,071)

Results: K-12 spending lowers leisure (ACS, 1:100)



- Weeks worked
  - By 30: down 0.1 weeks per \$10,000 (sum: 465)
  - By 45: up 4.3 weeks per \$10,000 (sum: 1,071)
- Hours worked per week
  - At 19: up 0.1 hours per \$10,000 (mean: 22)
  - At 35: up 0.5 hours per \$10,000 (mean: 35)

Conclusion: does money matter?

Three substantive cost-benefit questions:

1. Benefit? Significant, with caveats
   - Cumulative earnings up by 73.5 cents per dollar of K-12 spend (though a portion of this is due to greater labor input) – CI are -79.8 to 226.7
   - Cumulative spending on higher education up by 22.3 cents per dollar of K-12 spend
2. Break-even? Point estimates no, upper CI late
   - Earnings alone: over age 32, for upper end of CI (that is, at a 15-to-27-year lag)
   - Earnings net of spending on higher education: same
3. NPV>0? Point estimates no, upper CI low
   - Earnings alone: under IRR=3.7 percent (compare: 20-to-30-year Treasury real yield is 2.3 percent as of this week)
   - Earnings net of spending on higher education: under IRR=4.1 percent

Conclusion: does money matter?

What we DO show (for the case of court-ordered reforms):

- that K-12 spending did not pay for itself

What we DON'T show:

- that there are no benefits
- that there are no beneficiaries
- that K-12 spending cannot pay for itself
- that it even should

Thank you!

Miller: corbin.l.miller@irs.gov
Klopfer: johnklopfer@gmail.com