What Impacts Can We Expect from School Spending Policy? Evidence from Evaluations in the U.S.

C. Kirabo Jackson and Claire Mackevicius
Northwestern University

Introduction

- Education is one of the largest single components of government spending (OECD, 2020).
- Recent evidence strongly suggests that money matters, but school finance litigation and policy decision hinge on *the extent* to which money matters.
  - On this front there are many unanswered questions.

- How much does money matter *on average*?
- How heterogeneous are true marginal spending effects (not just sampling error)?
  - Do effects from one context inform another? Are studies of older policies informative for policy today?
- Do marginal effects vary by spending type, geography, and student populations?
- What predictions can be made about *future* school-spending policies?

- Answers to these questions requires information from a diversity of settings and policies → Difficult with a single policy.

What we do

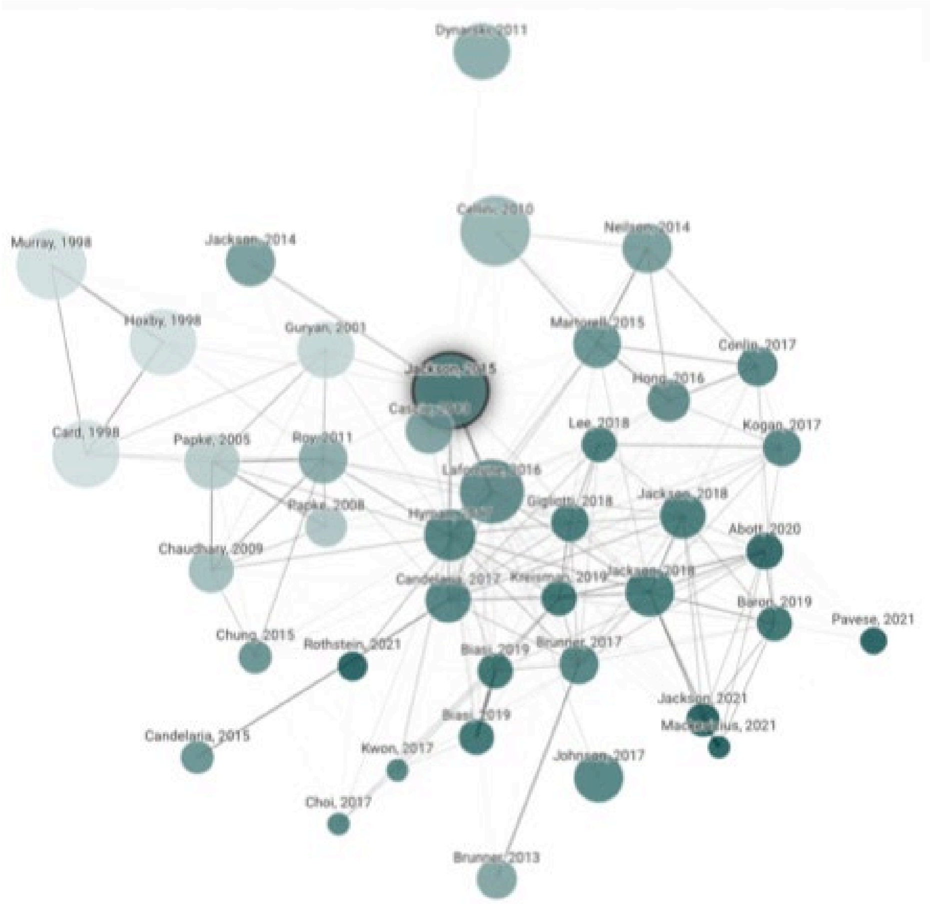
- Perform formal meta-analysis on a comprehensive set of design based (i.e., credibly causal) studies on the effect of school spending on student outcomes.
  - Measure the average marginal effect of school spending on test scores and educational attainment.
  - Measure true heterogeneity (not explained by sampling errors) in marginal effects to speak to generalizability.
  - Test for average differences across a few key dimensions.
  - Test the assumption about the distribution of true effects.
  - Make relatively precise policy predictions using our estimates of the average and heterogeneity.

Data for Analysis

- We use 31 studies that estimate the impact of spending on student outcomes and meet our inclusion criteria
  - Included studies had to be based on a valid policy instrument for school spending.
    - Be credibly causal (uses quasi-experimental variation with some testing of assumptions)
    - Show a statistically significant change in school spending.
      - Our conclusions are robust to excluding studies that may be weakly identified.

- Outcomes include:
  - Test scores
  - High school graduation, college enrollment → Educational Attainment

Getting a comprehensive set of studies



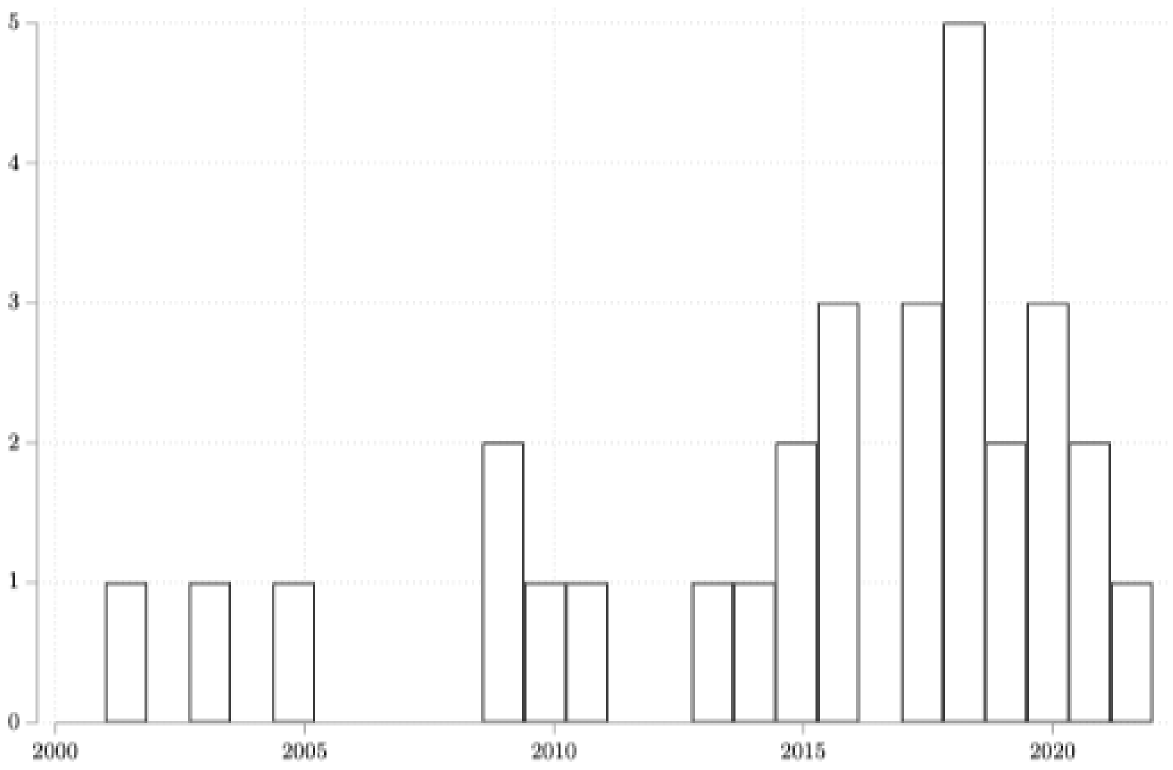
- We start with a set of known included studies (seed studies).
- We then find all connected studies to each seed study and evaluate.
- Newly included studies then become seed studies
- The process continues until there are no more potential included students.

Examples of excluded papers

- Van der Klaw (2008) studies the effect of Title I on student outcomes.
  - *“eligibility does not necessarily lead to a statistically significant increase in average per pupil expenditures.”*
- Husted and Kenny (2000) cannot ascribe their variation to any particular policy.
  - *“Our preferred resource equalization measure. . . equals the change in resource inequality since 1972 relative to the predicted change (that is, the unexplained change in inequality). A fall in this variable reflects either the adoption of state policies that have reduced districts’ ability to determine how much to spend in their district or an otherwise unmeasured drop in spending inequality”*
- Hoxby (2001) does run a IV model relating school spending to dropout rates.
  - No explicit first stage F-statistic for this model and it is not obvious that it is strong.

Distribution of Publication/Draft Dates: Included Papers

Figure A.2: Count of Included Studies per Year



Summary of Data

Study	Study ID	Outcome	Spending Type	Estimation Strategy	Raw Estimate ($\hat{\theta}_j$)	SE of $\hat{\theta}_j$	Bayes Estimate ($\tilde{\theta}_j$)
Abott Kogan Lavertu Peskowitz (2020)	1	High school graduation	Operational	RD	0.0847	0.0876	0.0596
Abott Kogan Lavertu Peskowitz (2020)	1	Test scores	Operational	RD	0.1158	0.0667	0.0439
Baron (2022)	2	College enrollment	Operational	RD	0.1869	0.0767	0.0681
Baron (2022)	2	Test scores	Operational	RD	0.1790	0.1305	0.0388
Baron (2022)	3	Test scores	Capital	RD	-0.1579	0.0979	0.0216
Brunner Hyman Ju (2020)	4	Test scores	Any	ES DiD	0.0531	0.0173	0.0469
Candelaria Shores (2019)	5	High school graduation	Any	ES DiD	0.0511	0.0133	0.0529
Carlson Lavertu (2018)	6	Test scores	Any	RD	0.0902	0.0475	0.0461
Cascio Gordon Reber (2013)	7	High school dropout	Any	ES	0.5546	0.2056	0.0638
Cellini Ferreira Rothstein (2010)	8	Test scores	Capital	RD	0.1773	0.0829	0.0458
Chaudhary (2009)	9	Test scores	Any	IV	0.0179	0.0424	0.0294
Clark (2003)	10	Test scores	Any	ES DiD	0.0148	0.0116	0.0181
Conlin Thompson (2017)	11	Test proficiency rates	Capital	ES	0.0063	0.0047	0.0072
Gigliotti Sorensen (2018)	12	Test scores	Any	IV	0.0424	0.0098	0.0413
Goncalves (2015)	13	Test proficiency rates	Capital	ES	-0.0017	0.0197	0.0117
Guryan (2001)	14	Test scores	Any	IV	0.0281	0.0689	0.0329
Hong Zimmer (2016)	15	Test proficiency rates	Capital	RD	0.0911	0.0512	0.0448
Hyman (2017)	16	College enrollment	Any	IV	0.0552	0.0257	0.0568
Jackson Johnson Persico (2016)	17	High school graduation	Any	ES DiD	0.0798	0.0163	0.0723
Jackson Wigger Xiong (2021)	18	College enrollment	Any	IV	0.0380	0.0133	0.0432
Jackson Wigger Xiong (2021)	18	Test scores	Any	IV	0.0499	0.0196	0.0438
Johnson (2015)	19	High school graduation	Any	ES DiD	0.1438	0.0753	0.0650
Kogan Lavertu Peskowitz (2017)	20	Test scores	Any	RD	0.0190	0.0127	0.0219
Kreisman Steinberg (2019)	21	High school graduation	Any	IV	0.0279	0.0146	0.0369
Kreisman Steinberg (2019)	21	Test scores	Any	IV	0.0779	0.0237	0.0572
Lafortune Rothstein Schanzenbach (2018)	22	Test scores	Any	ES DiD	0.0164	0.0133	0.0201
Lafortune Schonholzer (2021)	23	Test scores	Capital	ES DiD	0.0504	0.0223	0.0430
Lee Polachek (2018)	24	High school dropout	Any	RD	0.0640	0.0141	0.0623
Martorell Stange McFarlin (2016)	25	Test scores	Capital	RD	0.0254	0.0226	0.0290
Miller (2018)	26	High school graduation	Any	IV	0.0662	0.0169	0.0632
Miller (2018)	26	Test scores	Any	IV	0.0515	0.0137	0.0474
Neilson Zimmerman (2014)	27	Test scores	Capital	ES DiD	0.0314	0.0236	0.0324
Papke (2008)	28	Test proficiency rates	Any	IV	0.0817	0.0121	0.0728
Rauscher (2020)	29	Test scores	Capital	RD	0.0070	0.0041	0.0076
Rauscher (2020)	30	Test scores	Operational	DiD	0.0161	0.0271	0.0254
Roy (2011)	31	Test scores	Any	IV	0.3804	0.1563	0.0424
Weinstein Stiefel Schwartz Chalico (2009)	32	High school graduation	Any	RD	0.1595	0.1698	0.0597
Weinstein Stiefel Schwartz Chalico (2009)	32	Test scores	Any	RD	-0.0541	0.0368	0.0054

- The data cover multiple estimation strategies (IV, RD, DiD), time periods (1965 through 2015), populations (low and high income), and geographies (South, north, etc.) and urbanity (rural/urban).

Metanalysis for Quasi-experimental Studies

Constructing Comparable Estimates From Each Paper

- Standardized outcomes (test scores, educational attainment)
  - Accounts for differences in testing, and reporting (*albeit imperfectly*)
    - Test scores are divided by sd. Proficiency rates are divided by $\sqrt{\hat{p}(1-\hat{p})}$.
    - College-going, high school grad, dropout ➔ Each is divided by $\sqrt{\hat{p}(1-\hat{p})}$.
- Population average treatment effects
  - Not just the population with the largest effects (often emphasized)
- Equalize (when possible)
  - Years of exposure (duration)
  - Size of spending change (dosage in 2018 CPI-adjusted dollars)
- **Each study-outcome provides an estimate of the effect of a $1000 change in per-pupil spending, sustained over 4 years.**

Two very different abstracts

Abstract

The Kentucky Education Reform Act (KERA), implemented in 1990, is one of the most ambitious and influential education reform policies ever attempted by any state. KERA's main components include a new funding system to correct large financial disparities between school districts, curriculum revision and standardization, and increased school and district accountability. In this paper I explore KERA's effects on school spending and student achievement. I find that KERA did successfully equalize per-pupil expenditures across rich and poor districts and that this equalization has been sustained throughout the decade since KERA's passage. KERA's effects on student achievement have been more modest. Black students in Kentucky have experienced modest test score gains since KERA's implementation, but the scores of white students have remained unchanged relative to their peers in surrounding states. I also find no evidence that KERA narrowed the gap in test scores between rich and poor districts. Instrumental variables estimates suggest the increased spending induced by KERA did not improve test scores.

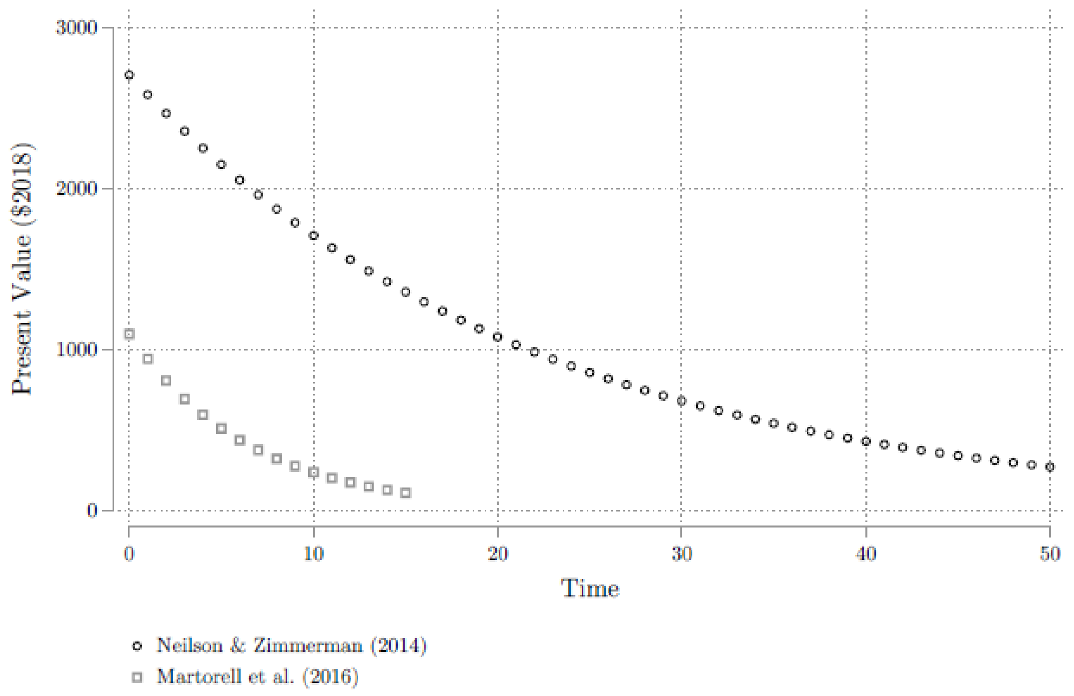
We study the impact of post-1990 school finance reforms, during the so-called "adequacy" era, on absolute and relative spending and achievement in low-income school districts. Using an event study research design that exploits the apparent randomness of reform timing, we show that reforms lead to sharp, immediate, and sustained increases in spending in low-income school districts. Using representative samples from the National Assessment of Educational Progress, we find that reforms cause increases in the achievement of students in these districts, phasing in gradually over the years following the reform. The implied effect of school resources on educational achievement is large. (*JEL* H75, I21, I22, I24, I28)

It turns out that both these studies have very similar effects on average.

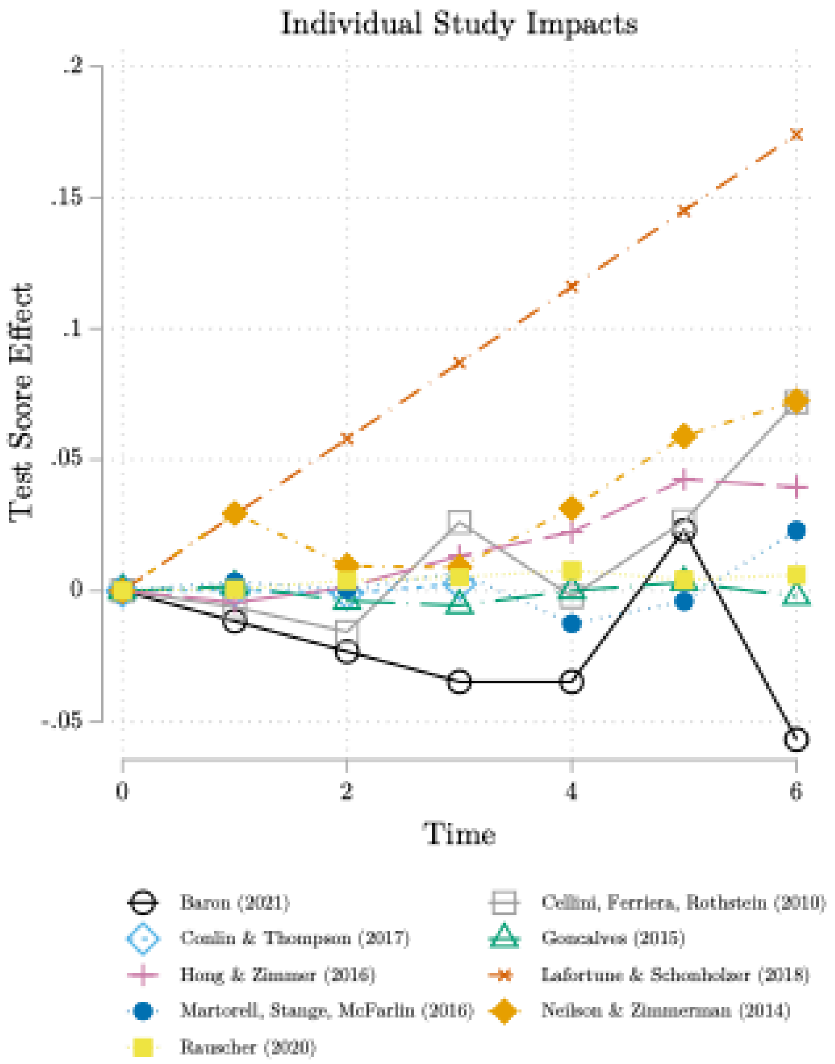
- The first paper (Clark 2003) uncovers very noisy positive effects in the IV models.
- The second paper (LaFortune et. al. 2018) large effects for low-income groups that when averaged with imprecise negative effect for high-income groups yield modest effects overall.

Making Capital and Non-Capital Spending Types Comparable

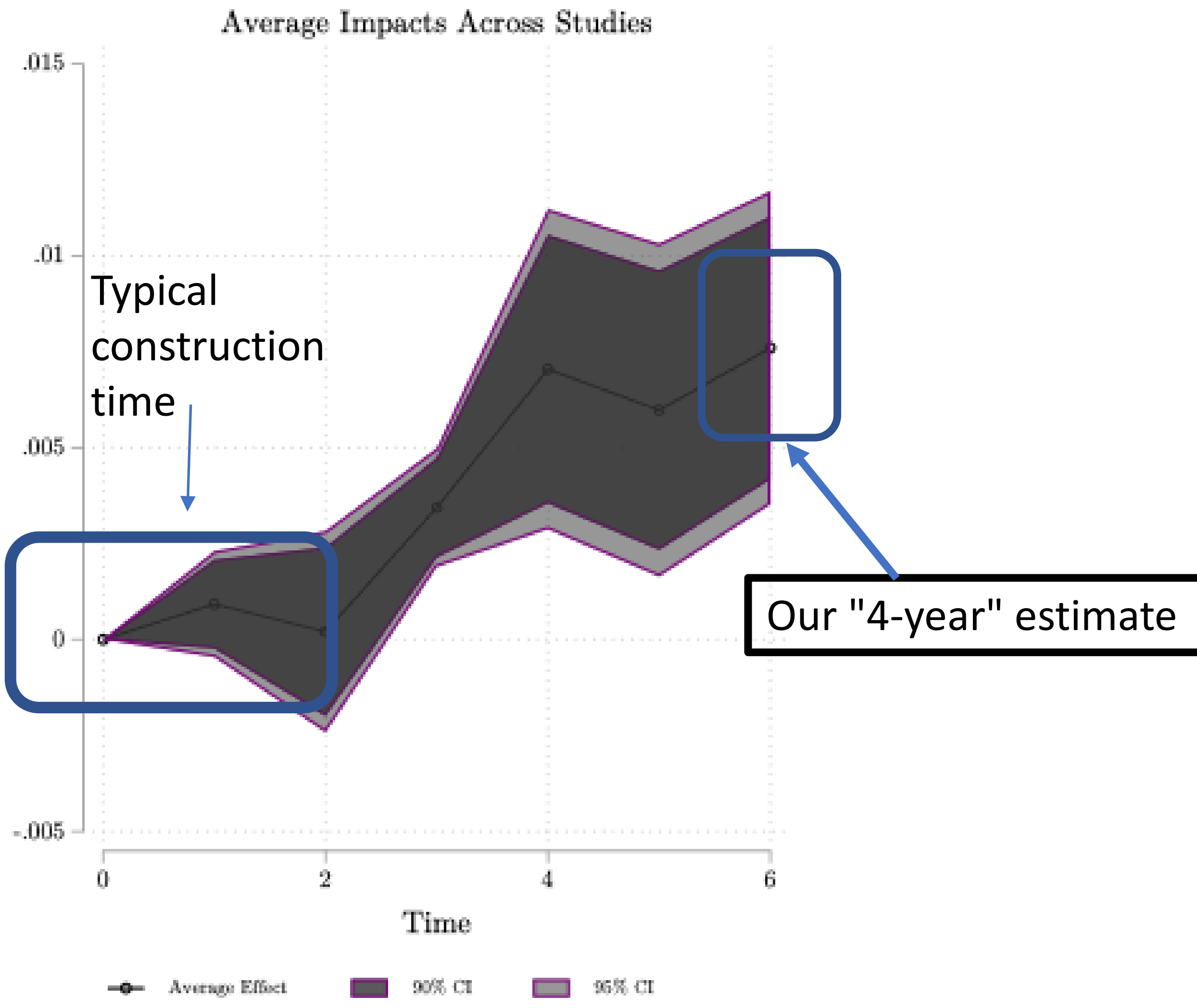
- Spread large capital expenses over the life of the asset.
- New buildings are depreciated at 4.7% and non-building projects are depreciated at 16.5%.
  - ➔Buildings and other have 10 percent of their value remaining after 50 and 15 years, respectively.
- Account for construction time (2 years).



Capital Spending Effects Over Time



Note: Lafortune & Schonholzer (2018) reports linear effects over time.



For both figures, time is year relative to construction or spending increase.
Effect is the standardized effect, relative to year 0.

Random Effects Meta-Analysis Setup I

$$\hat{\theta}_j \sim N(\theta_j + \sigma_j^2)$$

Estimate from study j

True effect for study *j*

Within-study sampling variance

$$\theta_j \sim N(\boldsymbol{\Theta}, \tau^2)$$

True effect for study j

Pooled average effect

Variance of between-study heterogeneity

$$\hat{\theta}_j \sim N(\boldsymbol{\Theta}, \sigma_j^2 + \tau^2)$$

Estimates deviate from the grand mean due to sampling variability and true heterogeneity.

Random Effects Meta-Analysis Setup II

The optimal precision weighted average is

$$\widehat{\boldsymbol{\Theta}}_{pw} = \frac{\sum w_j \hat{\theta}_j}{\sum w_j}$$

where $$w_j = \frac{1}{\sigma_j^2 + \tau^2}$$

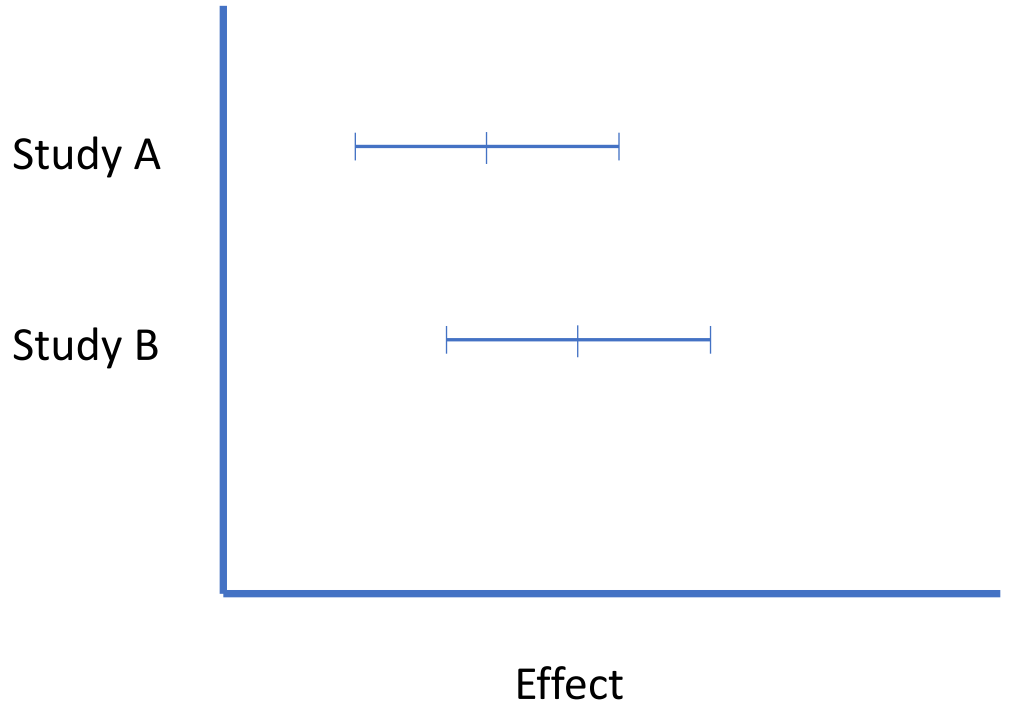
Approximated with se_j^2

To be estimated

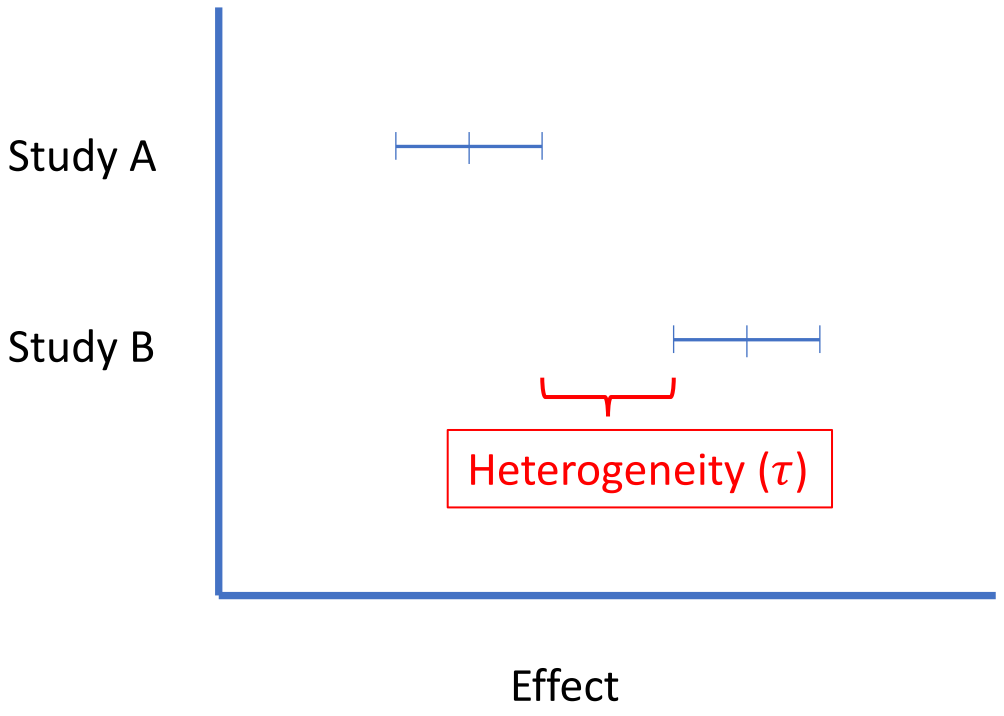
There are many ways to estimate τ^2. Intuitively, τ^2 is identified based on the variability that cannot be explained by sampling errors.

Meta-Analytic Methods: Measuring Heterogeneity

Cross-Study Variability Driven Entirely by Within-Study (Sampling) Variability



Cross-Study Variability Driven by <u>some</u> Across-Study Heterogeneity



We use each study's SE to measure within-study variability, and estimate τ with what is unexplained by within-study variability

Meta-Analytic Methods: Reporting Intervals

- We use random effects meta-analysis to compute:
  - A precision weighted pooled average ➔ $\widehat{\Theta}_{pw}$
  - The standard error of the pooled average ➔ $se_{\widehat{\Theta}_{pw}}$
  - The variance of the *true* heterogeneity ➔ $\hat{\tau}^2$

- Reported distribution of impacts
  - Confidence Interval (where the pooled average will be)

    $$CI = \widehat{\Theta}_{pw} \pm t^* \times se_{\widehat{\Theta}_{pw}}$$

  - Prediction Interval (where future individual effects will be)

    $$PI = \widehat{\Theta}_{pw} \pm t^* \times \sqrt{se^2_{\widehat{\Theta}_{pw}} + \hat{\tau}^2}$$

$\boxed{\text{Relevant for Policy}}$

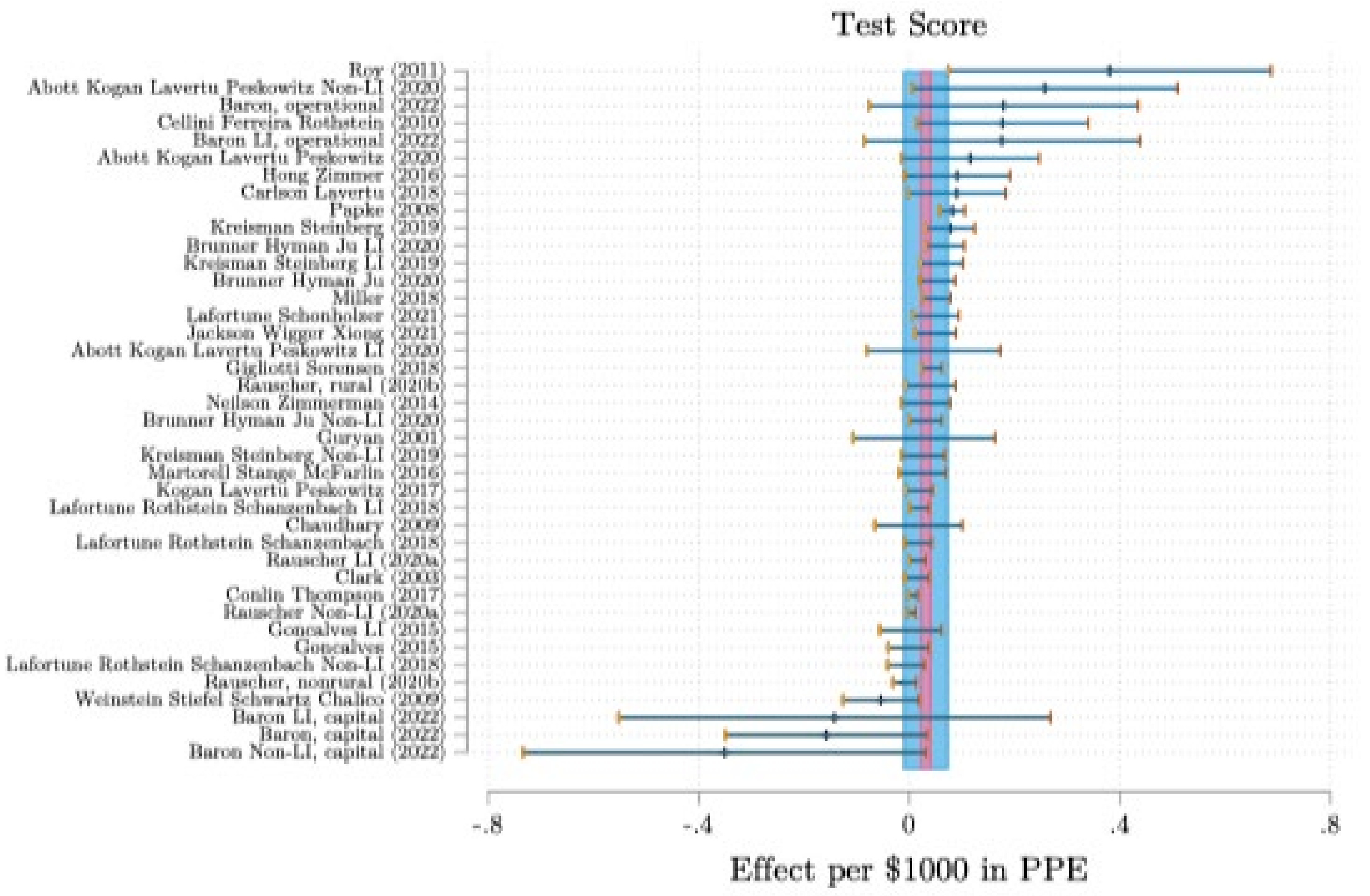
Empirical Bayes Estimate for Individual Studies

- Random effect meta-analysis allows for Bayes estimates of the effect from individual studies.
  - The motivating intuition is that estimates from other studies can provide information about any specific study's true effect.
- The logic is analogous to shrinkage estimates for teacher or school or hospital effects.
  - Noisy estimates are shrunk towards the grand mean yielding the Best Linear Unbiased Prediction (BLUP).
- Under the distributional assumptions, we get that $\hat{\theta}_j \sim N(\boldsymbol{\Theta}, \sigma_j^2 + \tau^2)$. It follows that ....

$$E\left(\theta_j | \widehat{\theta_j}, \sigma_j, \tau\right) = \hat{\theta}_j(1 - \boldsymbol{B_j}) + \boldsymbol{B_j}\hat{\boldsymbol{\Theta}} \quad , \quad \text{where} \boldsymbol{B_j} = \frac{\sigma_j^2}{\sigma_j^2 + \tau^2}$$

  - To form the BLUP ($\tilde{\theta}_j$), we replace $\boldsymbol{B_j}$ with its empirical analog $\frac{se_j^2}{se_j^2 + \hat{\tau}^2}$.

The Distribution of Marginal Test Score Effects



- Take a very large noisy estimate like Roy (2011). The estimate is 0.38, but the Bayes estimate is 0.038.
  - If it was as precise as the smaller estimates, the shrinkage estimate would be about 0.2.
- The most precise estimates are those that are very close to the pooled average.

Some examples of Raw Estimates vs. BLUPs

Table 2: Summary of Studies

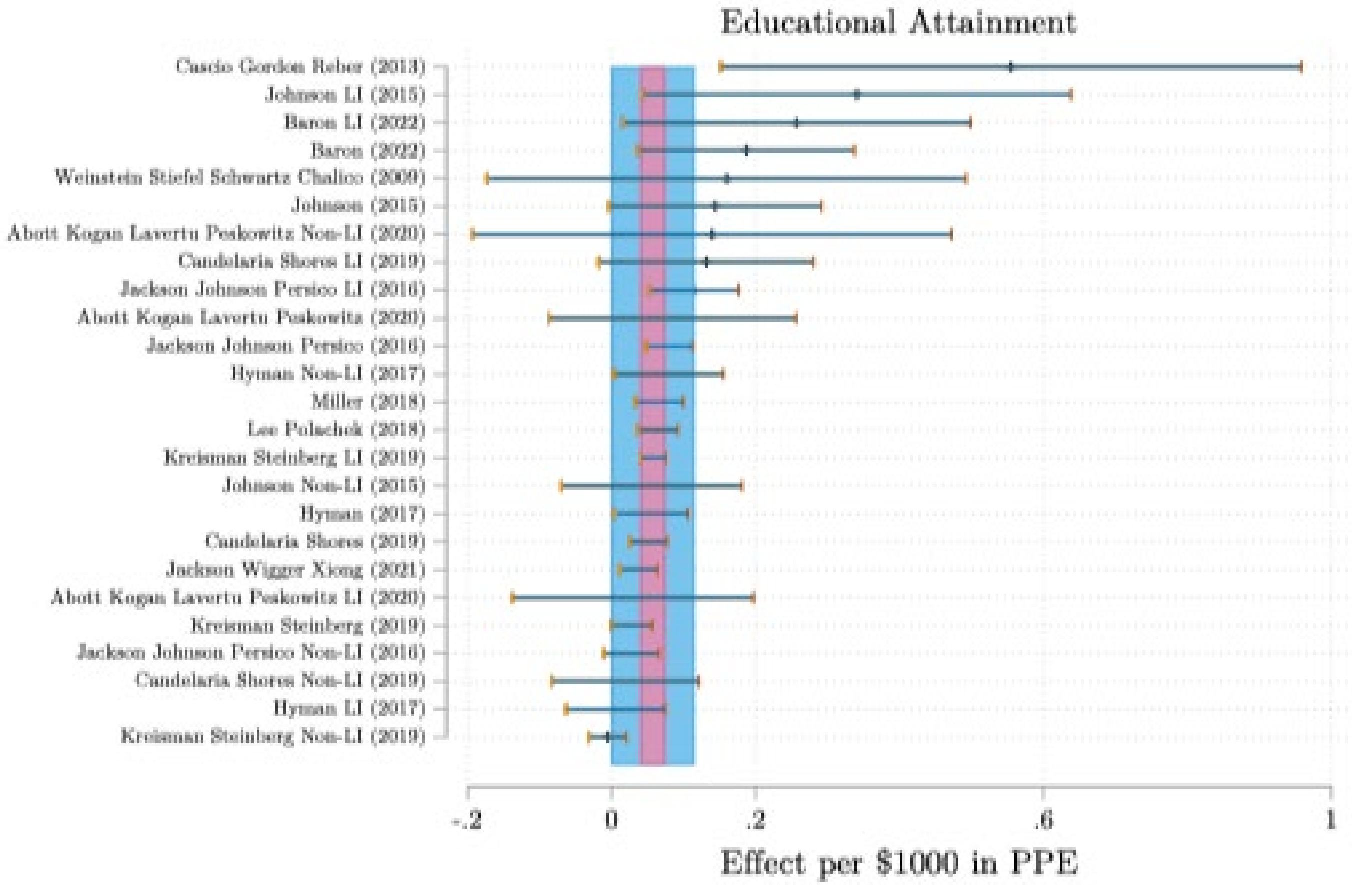
Study	Study ID	Outcome	Spending Type	Estimation Strategy	Raw Estimate ($\hat{\theta}_j$)	SE of $\hat{\theta}_j$ (se_j)	Bayes Estimate ($\tilde{\theta}_j$)
Abott Kogan Lavertu Peskowitz (2020)	1	High school graduation	Operational	RD	0.0847	0.0876	0.0596
Abott Kogan Lavertu Peskowitz (2020)	1	Test scores	Operational	RD	0.1158	0.0667	0.0439
Baron (2022)	2	College enrollment	Operational	RD	0.1869	0.0767	0.0681
Baron (2022)	2	Test scores	Operational	RD	0.1790	0.1305	0.0388
Baron (2022)	3	Test scores	Capital	RD	-0.1579	0.0979	0.0216
Brunner Hyman Ju (2020)	4	Test scores	Any	ES DiD	0.0531	0.0173	0.0469
Candelaria Shores (2019)	5	High school graduation	Any	ES DiD	0.0511	0.0133	0.0529
Carlson Lavertu (2018)	6	Test scores	Any	RD	0.0902	0.0475	0.0461
Cascio Gordon Reber (2013)	7	High school dropout	Any	ES	0.5546	0.2056	0.0638
Cellini Ferreira Rothstein (2010)	8	Test scores	Capital	RD	0.1773	0.0829	0.0458
Rauscher (2020)	29	Test scores	Capital	RD	0.0070	0.0041	0.0076
Rauscher (2020)	30	Test scores	Operational	DiD	0.0161	0.0271	0.0254
Roy (2011)	31	Test scores	Any	IV	0.3804	0.1563	0.0424
Weinstein Stiefel Schwartz Chalico (2009)	32	High school graduation	Any	RD	0.1595	0.1698	0.0597
Weinstein Stiefel Schwartz Chalico (2009)	32	Test scores	Any	RD	-0.0541	0.0368	0.0054

Regression Estimates

	(1)
Overall	0.0316***
	(0.00559)
Capital	
Low-Income	
Non-Low-Income	
Capital	
(SE)	
LI - Non-LI	
(SE)	
Observations	40
Clusters	22
τ	0.0207
$\hat{\sigma_\tau}$	(0.004)
Average 80% PI	[0.004,0.059]
Average 90% PI	[-0.004,0.067]
Average 95% PI	[-0.011,0.074]

- The pooled estimate of the effect of increasing per-pupil school spending by $1000 over 4 years is 0.0316σ.
- One rejects that the pooled average is zero and the 0.001 level.
- The estimate of heterogeneity is 0.021.
  - Positive effects over 90 percent of the time.
  - Prediction intervals are much tighter (and informative) than raw estimates would suggest.

Distribution of Effects on Educational Attainment



(1) Notice the wide range of estimates (raw estimates are over dispersed).
(2) Notice that the CI for the average overlaps the CI for almost all the studies.
(3) Studies are much more consistent than the raw estimates would indicate.
(4) The 95% PI is narrow and lies entirely above zero!!!

Regression Estimates

	Educational Attainment
	(5)
Overall	0.0574***
	(0.00786)
Capital	
Low-Income	0.0217
	(0.0215)
Non-Low-Income	-0.0336*
	(0.0193)
Capital	
(SE)	
LI - Non-LI	0.055**
(SE)	(0.028)
Observations	25
Clusters	11
τ	0.0219
$\hat{\sigma}_\tau$	(0.007)
Average 80% PI	[0.018,0.097]
Average 90% PI	[0.007,0.108]
Average 95% PI	[-0.003,0.118]

For high-school graduation the average is 2.05 percentage points.
For college-going, it is 2.81 percentage points.

High school completion impacts between 0.07 and 3.99 percentage points 95 percent of the time.

College going impacts between 0.9 percentage points and 5.51 percentage points 95 percent of the time.

Making Predictions Using This model

- To make policy predictions, the model relies on the notion that the true effects are normally distributed around the grand mean.
- Under normality, true effects are distributed $\theta_j \sim N(\boldsymbol{\Theta}, \tau^2)$.
- We do not observe $\boldsymbol{\Theta}$ but have a noisy estimate if it. As such, taking τ as given, we predict the distribution of true effects using

$$\theta_j \sim N(\widehat{\boldsymbol{\Theta}}_{pw}, se_{\widehat{\boldsymbol{\Theta}}_{pw}}{}^2 + \tau^2).$$

- If normality approximately holds, these prediction can be very informative.
- BUT.....There is no theoretical reason for normality to hold, so we test this assumption empirically.

Assessing Normality I

- Wang and Lee 2020 develop a simple test of normality based on the Shapiro-Wilk test of appropriately standardized effects.

$$\hat{\theta_j^S} = \frac{\hat{\theta}_j - \hat{\Theta}_{(-j)}}{(\hat{\tau}^2 + se_j^2 + se^2_{\hat{\theta}_{(-j)}})^{(1/2)}}$$

- One fails to reject normality for both outcomes.

Figure A.15: Educational Attainment Unweighted (L) and Weighted (R)

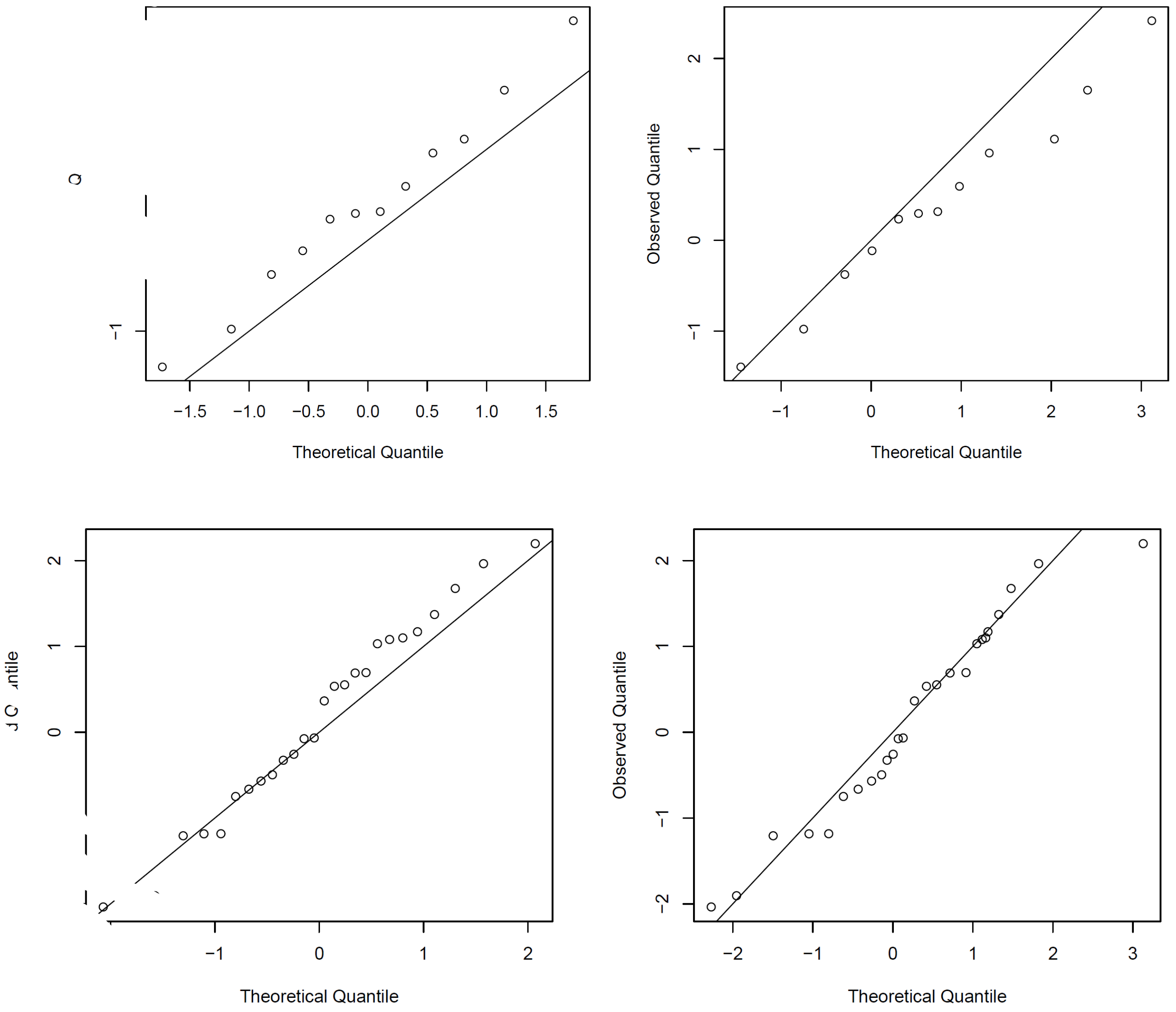
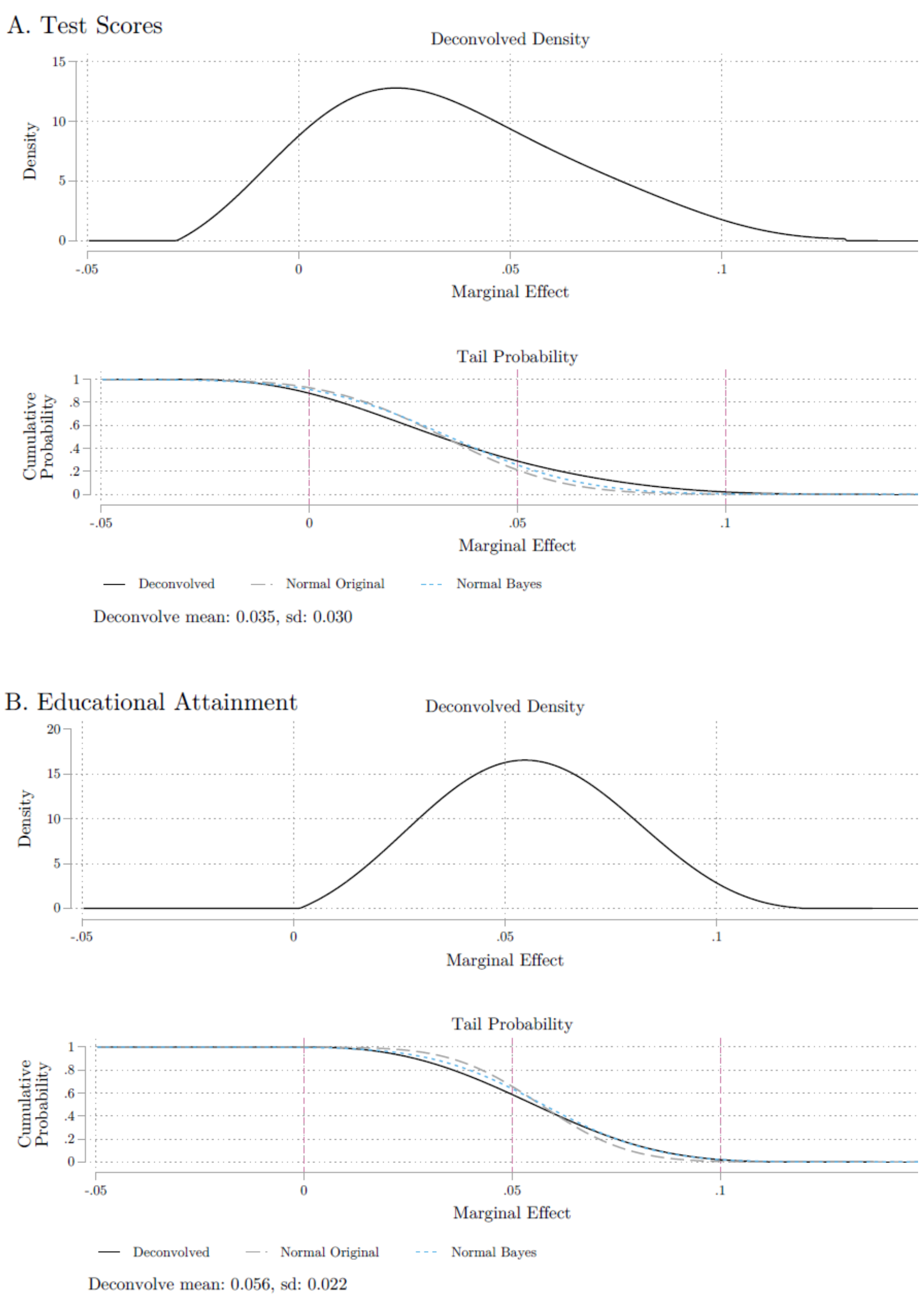


Figure A.16: Deconvolved Density and Effect Size Probabilities



Assessing Normality II

- We do not observe the distribution of true θ_j but only the that of $\hat{\theta}_j$. The distribution of $\hat{\theta}_j$ may be misleading about the distribution of true effects θ_j.

$$\widehat{\theta_j} \sim N(\theta_j + \sigma_j^2)$$
$$\theta_j \sim g(\mathbf{\Theta}, \tau^2)$$

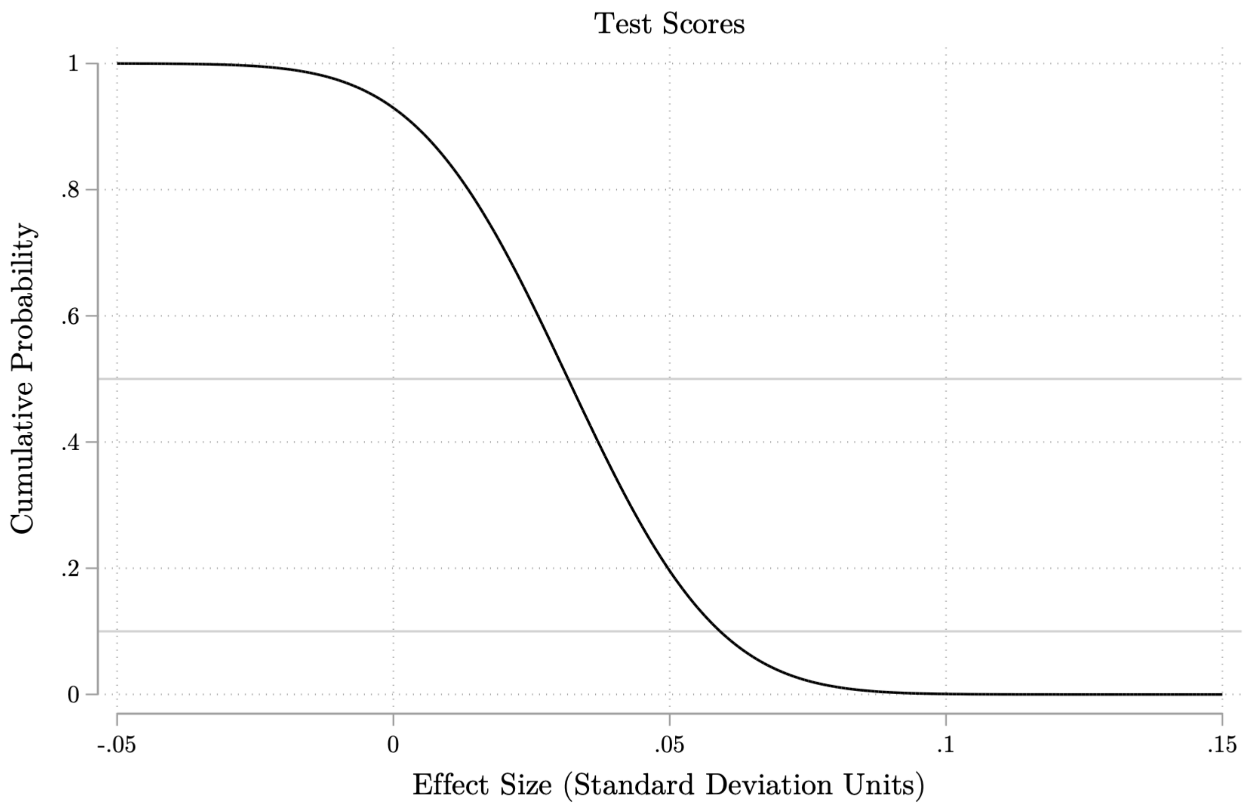
- Use a deconvolution kernel approach following Delaigle et al. (2008) to fit and approximate $g(\mathbf{\Theta}, \tau^2)$ using a Fourier transform.
- Compare the deconvolved density to that predicted by the normal distribution. ➔ <u>They are similar.</u>

Policy Predictions for Test Scores

Recall: $\theta_j \sim N(\boldsymbol{\Theta}, \tau^2) \Rightarrow N(\widehat{\boldsymbol{\Theta}}, se_{\widehat{\boldsymbol{\Theta}}}^2 + \tau^2)$

Looking overall, one would observe:

- positive test score impacts 92 percent of the time.
- Impacts above 0.03 half the time.
- Impacts above 0.05 twenty percent of the time.
- Almost never see true effects above 0.08 (despite some raw estimates in that range).

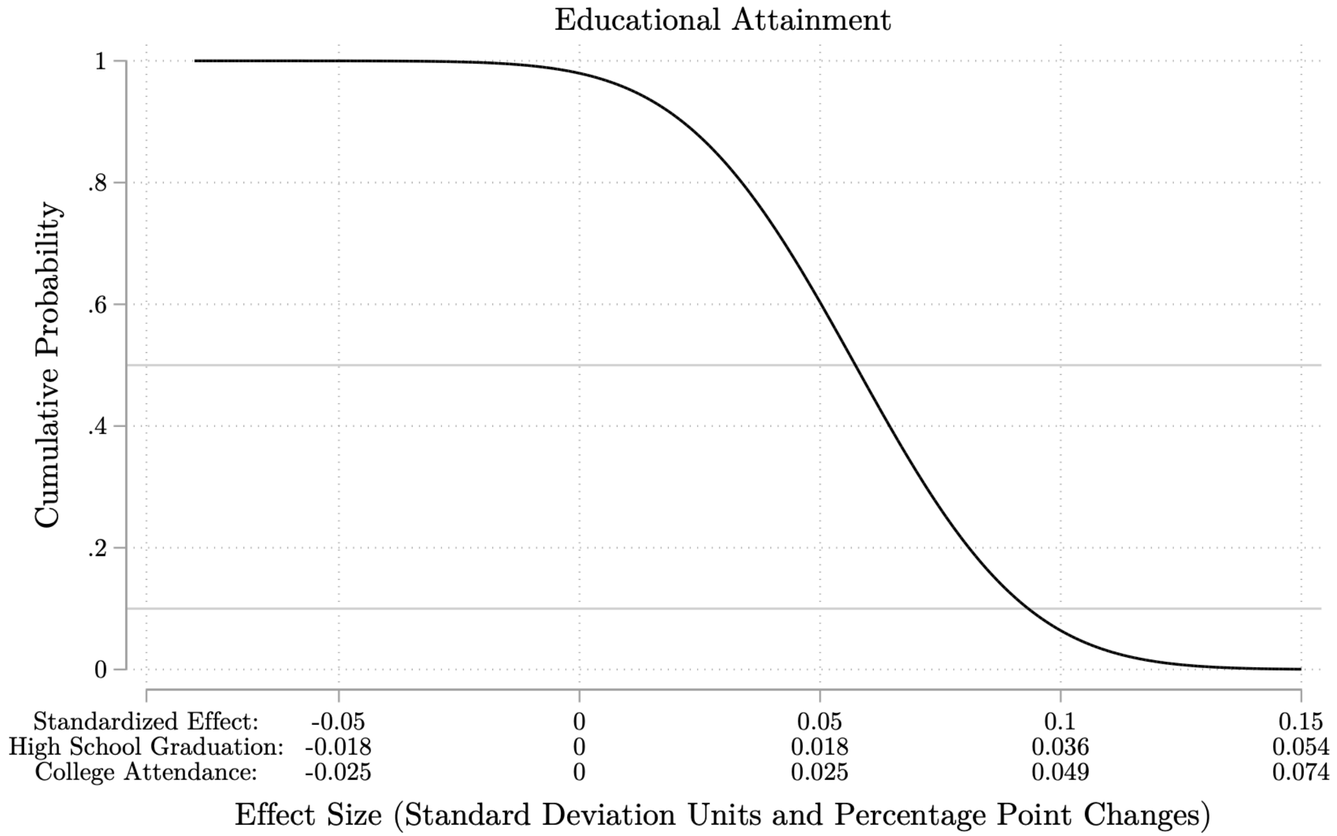


Policy Predictions for Educational Attainment

Looking overall, one would observe positive educational attainment impacts over 97 percent of the time.

A policy that increases spending by $1000 for four years will increase high school completion by about 2.5 and college-going by about 3.5 percentage-points 30 percent of the time.

And that same policy will increase high school completion by about 3.2 and college-going by about 4.5 percentage-points just over 10 percent of the time.



What explains the effect heterogeneity?

- Estimate random effects meta-regression with observable predictors of policy differences.

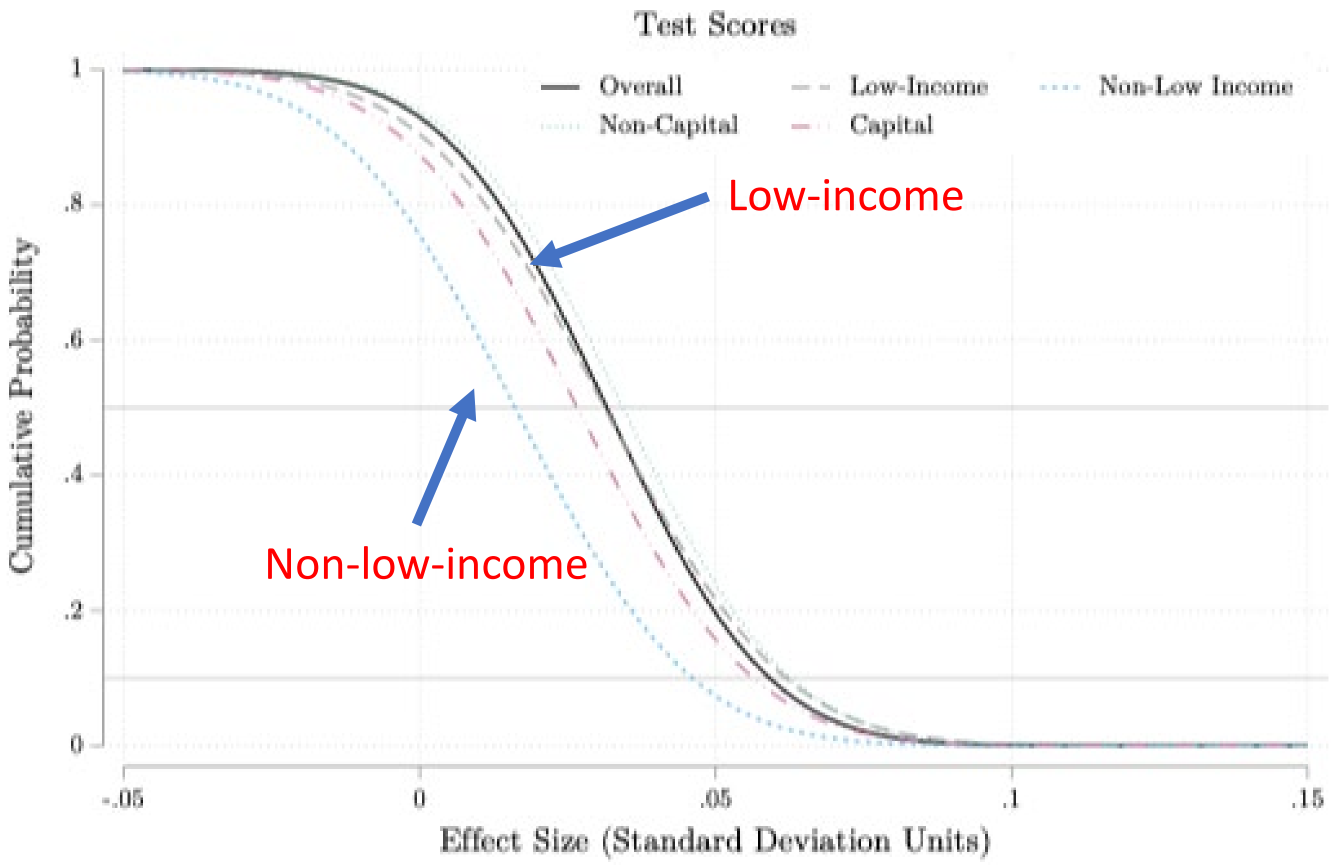
- By spending type (Capital and Non-capital)
- By student population served
  - Income level
  - Geography

Similar Effect of Capital and Non-Capital Spending

- Consistent with output maximization, *on the margin*, the marginal dollar spent on capital yields similar effects as those on non-capital.
- Because the PDV per-pupil dollar amounts are typically small, most individual studies are underpowered to detect effects of capital spending.

	Test Scores		
	(1)	(2)	(3)
Overall	0.0316***	0.0343***	0.0359***
	(0.00559)	(0.00681)	(0.00747)
Capital		-0.00769	-0.00533
		(0.0119)	(0.0112)
Low-Income			-0.00470
			(0.0107)
Non-Low-Income			-0.0198*
			(0.0102)
Capital		0.027***	0.031***
(se)		(0.010)	(0.009)
LI - Non-LI			0.015***
(se)			(0.008)
Observations	40	40	40
Clusters	22	22	22
τ	0.0207	0.0211	0.0210
$\hat{\sigma_\tau}$	(0.004)	(0.004)	(0.005)
Average 80% PI	[0.004,0.059]	[0.003,0.065]	[0.006,0.066]
Average 90% PI	[-0.004,0.067]	[-0.006,0.074]	[-0.003,0.075]
Average 95% PI	[-0.011,0.074]	[-0.013,0.082]	[-0.010,0.082]

Policy Predictions for Test Scores by Income



The CDF for capital is similar to that for non-capital spending.

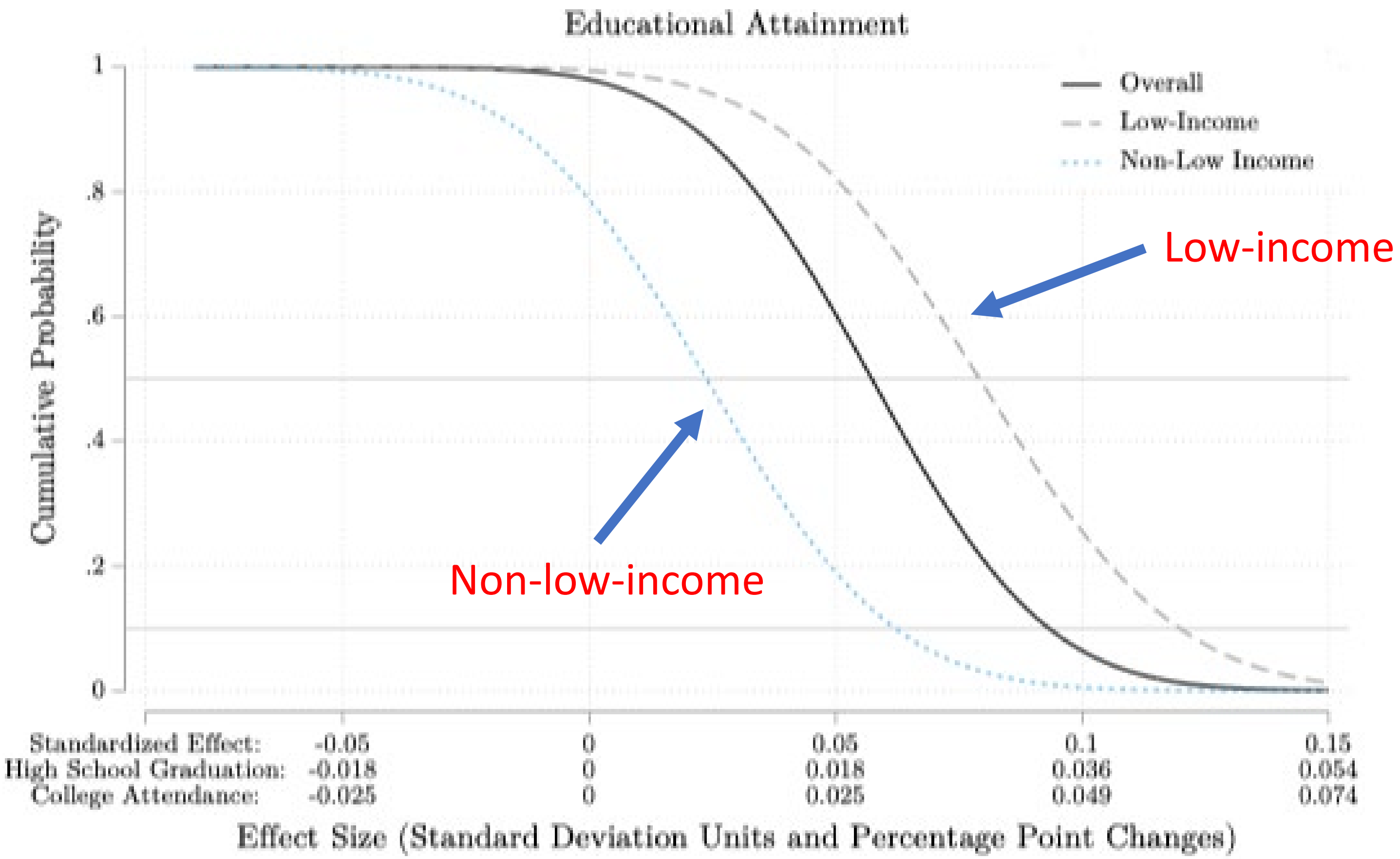
<u>The CDFs are quite different by income level.</u>

For non-low-income groups (and low income), one would observe positive test score impacts 70 (and 90) percent of the time.

Impacts above 0.04 only thirteen percent of the time (and over one third) the time.

Impacts above 0.062 almost never.

Policy Predictions for Educational Attainment by Income

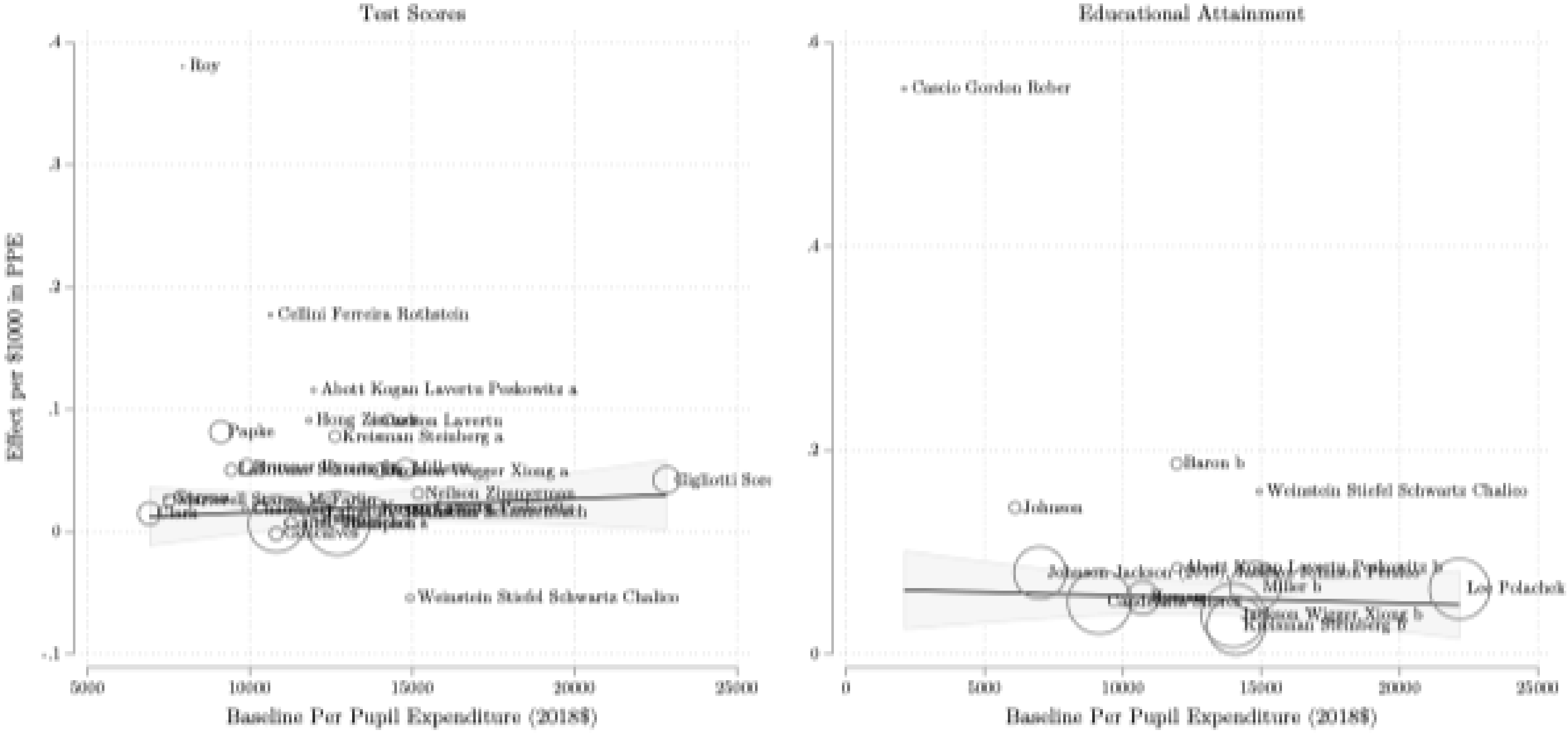


A $1000 increase for four years would improve educational attainment for low-income and not-low-income groups 99 and 79 percent of the time, respectively.

Such a policy would increase college-going by 2pp among low-income groups over 90 percent of the time, compared to under 30 percent of the time for higher-income groups.

Large effects above 5pp would occur with probability 0.2 for low-income and almost never for not-low-income groups.

We Find No Evidence of Diminishing Returns

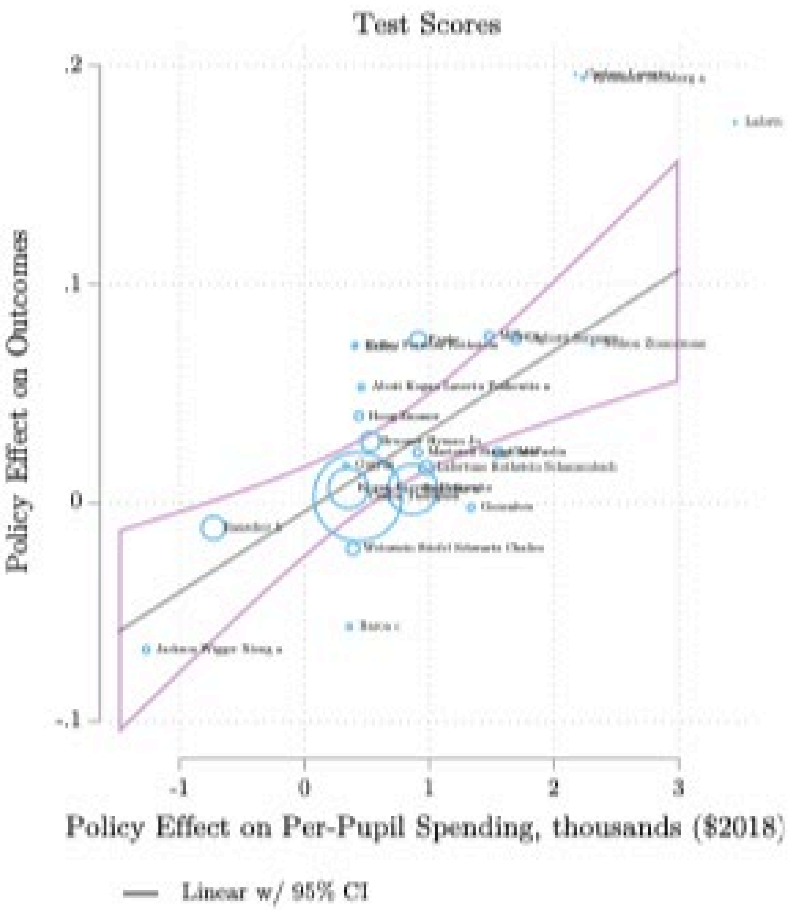


Bias in Individual Studies

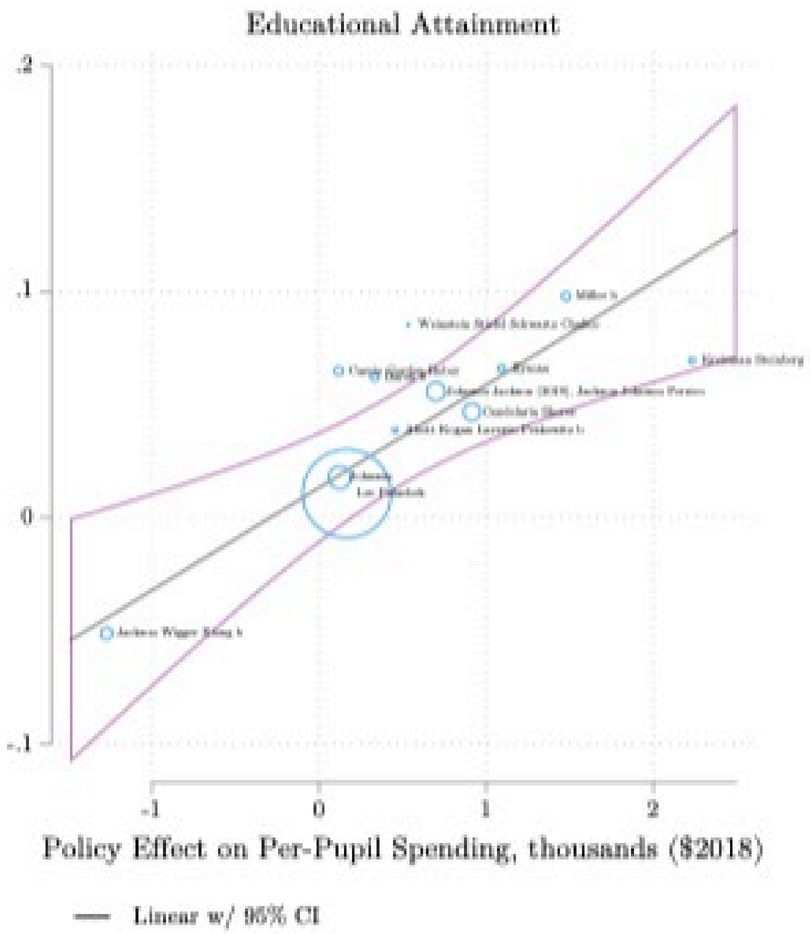
- If the bias in individual studies is random (as in not all biased up or down), then the pooled average will be unbiased.
- If all studies tended to be biased in the same direction, the within-study spending effect relationship would not go through the origin.
  - We show that it does.
- Bias tends to be larger for weakly-identified studies (Bound et. al. 1995)
  - First stage strength unrelated to marginal effect.
  - No relationship between the marginal effect and the size of the spending change.
- Effects are similar for policies that involve voluntary adoption versus others.
- Effects are similar in well-powered studies (where the auxiliary/placebo/falsification tests are most likely to be valid)

Larger Policies Have Larger Effects

(line goes through origin)



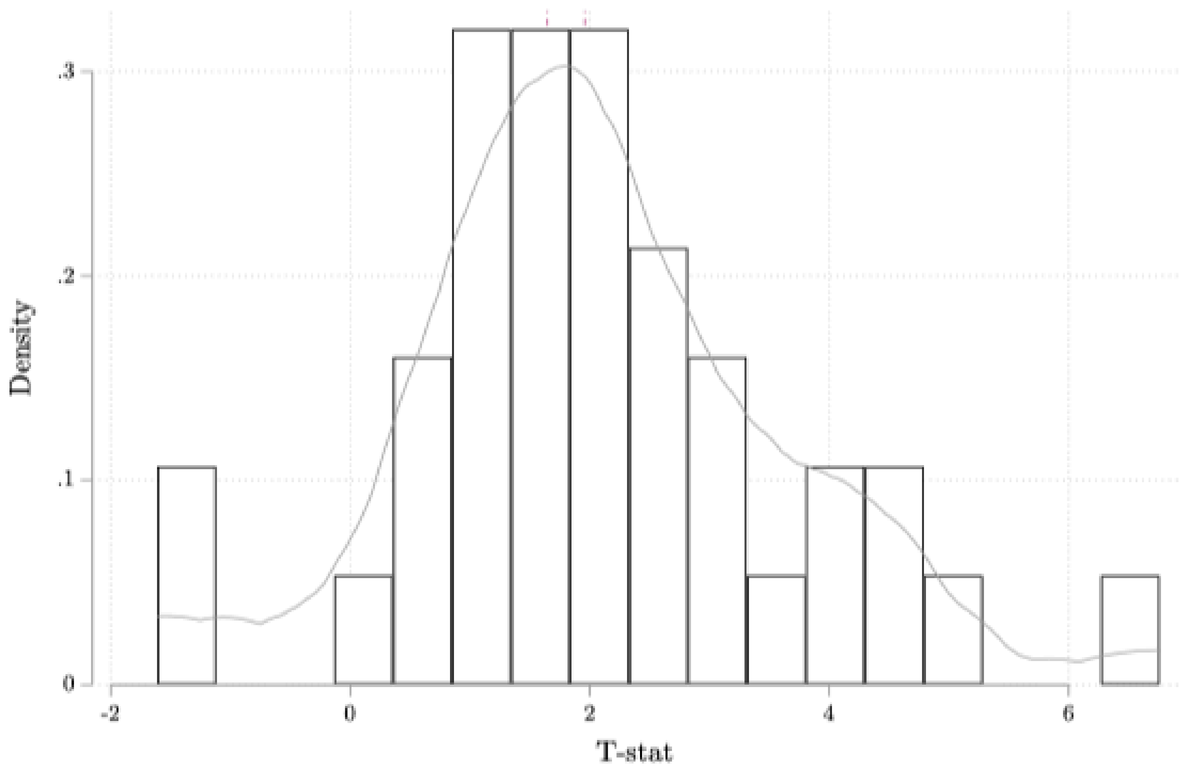
Precision-weighted linear slope: 0.0368 (p-value = 0.004),
constant: -0.0041 (p-value = 0.635)
Instrument overidentification test p-value: 0.4656
Roy (2011) and Chaudhary (2009) not plotted for scale.



Precision-weighted linear slope: 0.0454 (p-value = 0.011)
constant: 0.0133 (p-value = 0.165)
Instrument overidentification test p-value: 0.4813

Publication Bias 1a

Figure A.23: Histogram of *all* effects



<u>We may also worry about bias due to the selective publication of papers.</u>

No evidence of differential publication of statistically significant studies.

Results robust to Andrews and Kasy adjustment.

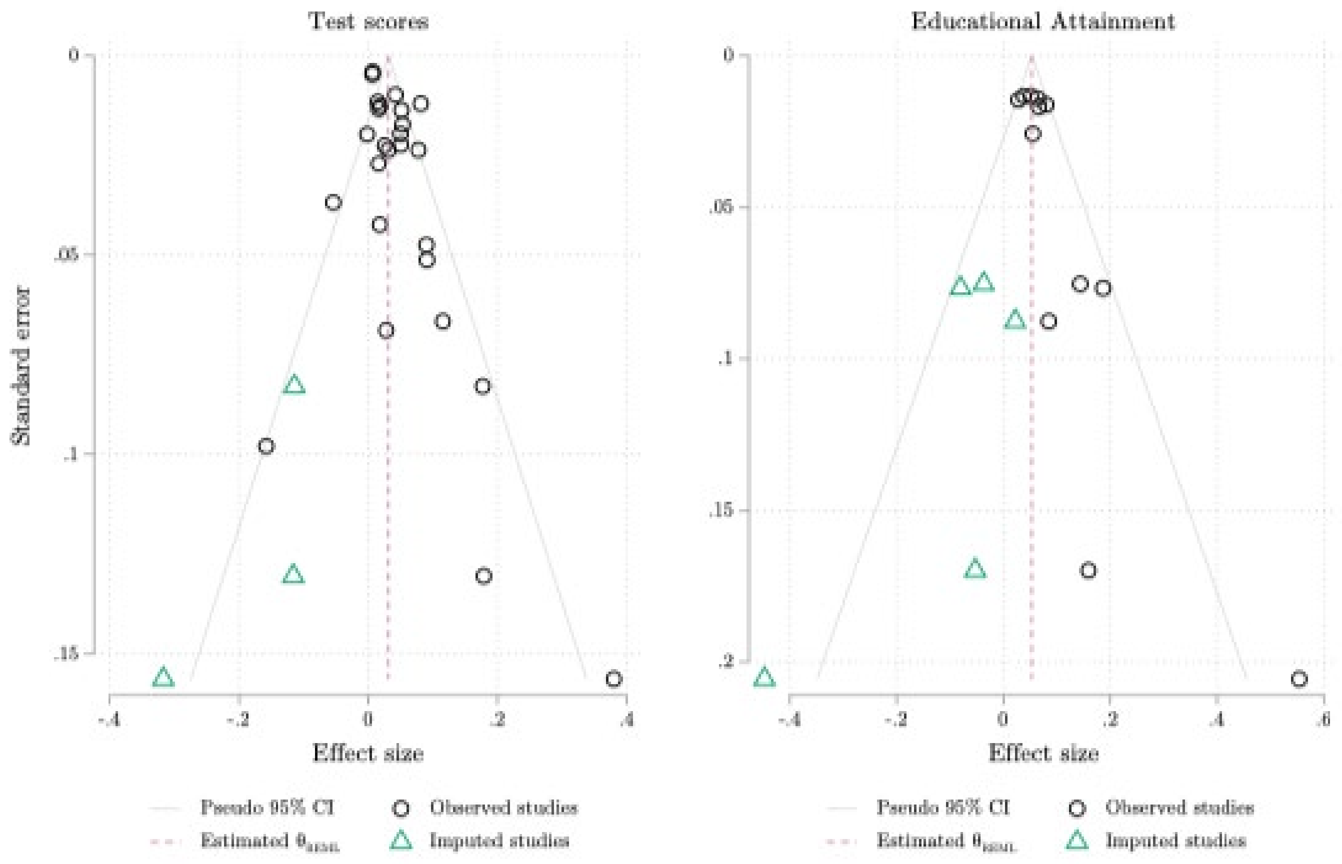
Publication Bias 1b

(Published and unpublished papers are similar)

Table A.15: Meta-Regressions w/ Publication Type

	(1) Test Score	(2) Test Score	(3) Educational Attainment	(4) Educational Attainment
Unpublished	-0.0109	-0.00687	0.0137	0.00448
	(0.0123)	(0.0175)	(0.0203)	(0.0308)
Top Field Journal		0.0100		-0.0349
		(0.0170)		(0.0269)
Field Journal		0.00611		0.00274
		(0.0182)		(0.0202)
Average Effect	0.0363***	0.0322**	0.0560***	0.0685***
	(0.00679)	(0.0138)	(0.00998)	(0.0191)
N	40	40	25	25
τ	0.0226	0.0246	0.0294	0.0462
Top Field = Field = Unpublished = 0 (p-val)		0.687		0.352
Unpublished = 0 (p-val)	0.376	0.694	0.500	0.884

Publication Bias 2

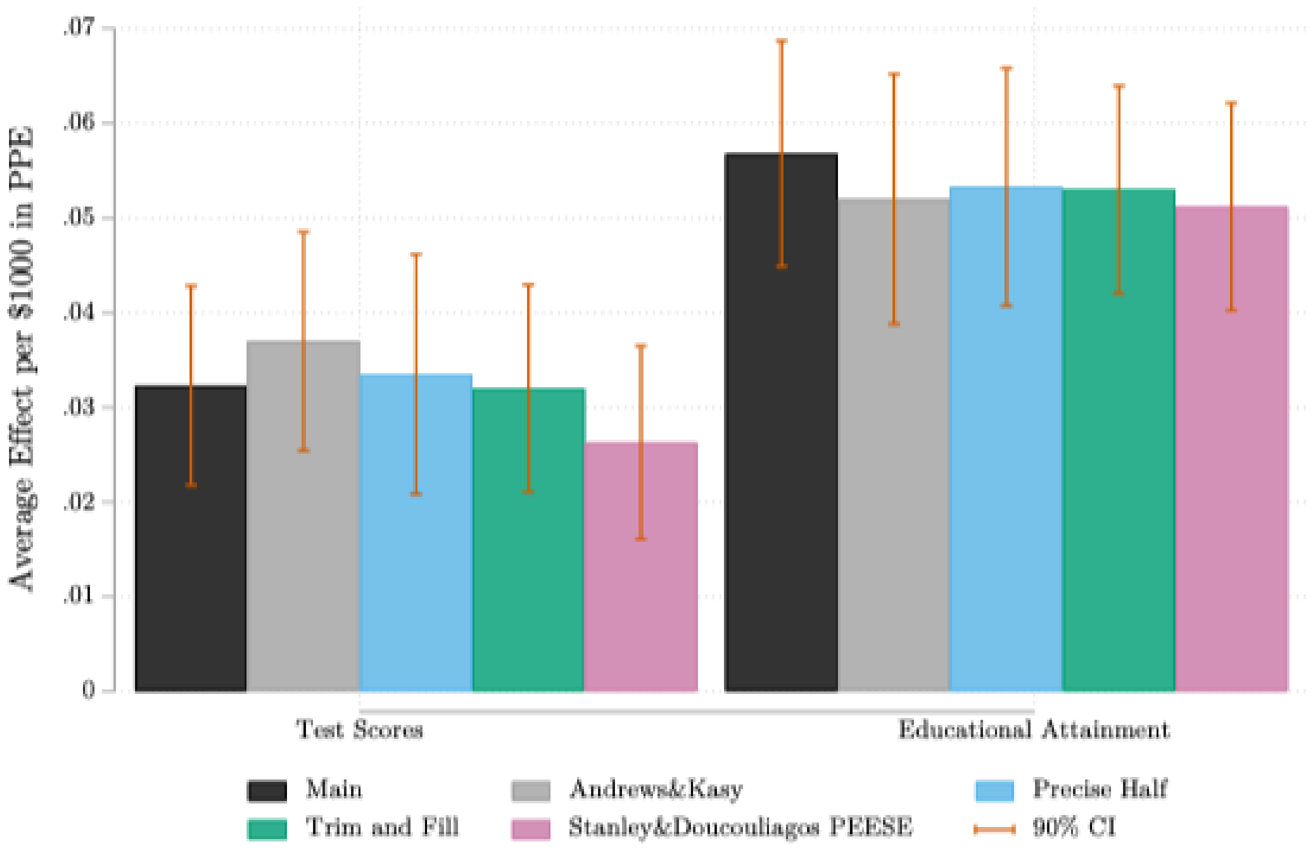


Some evidence of missing imprecise estimates (green)

Recall that estimates are precision weighted.

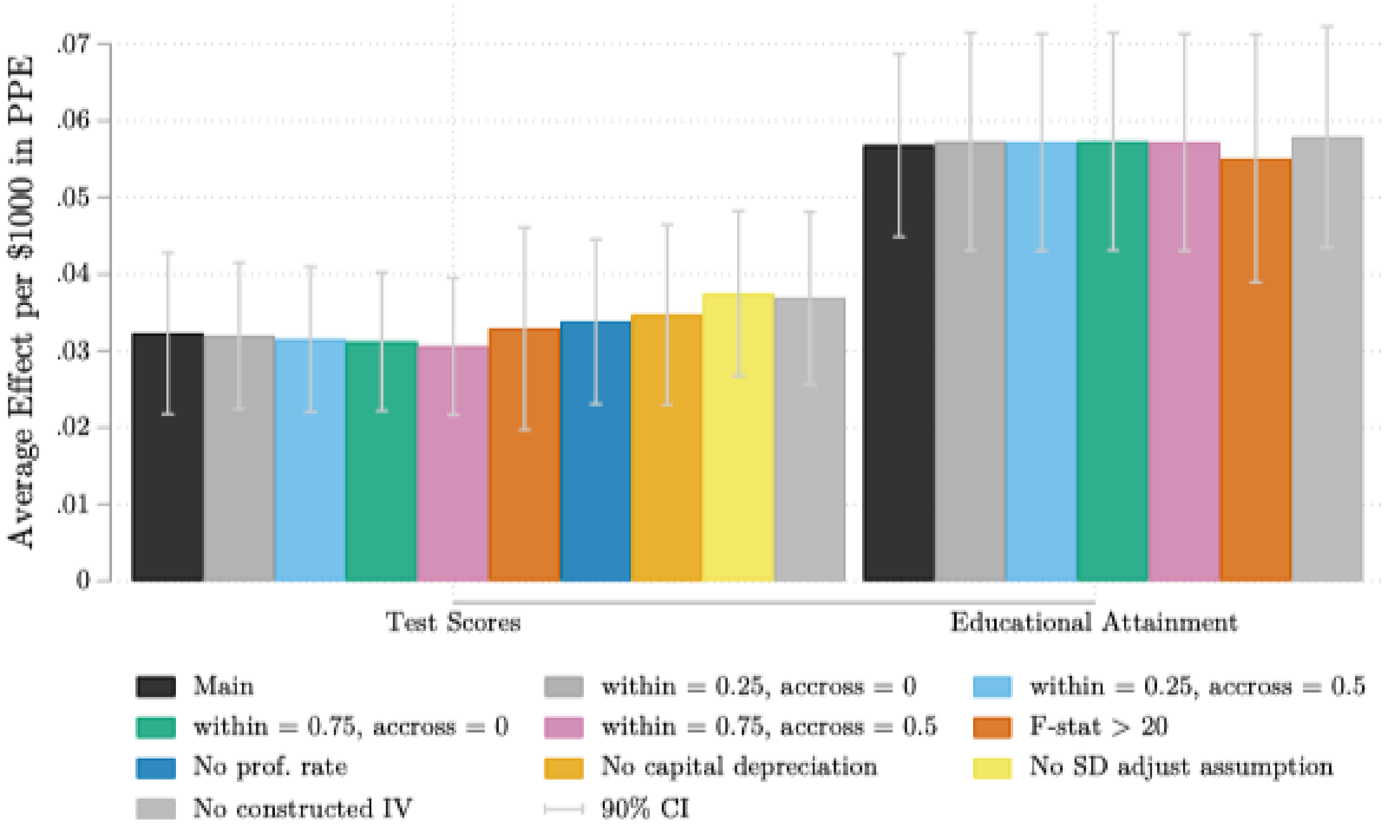
Summary of Adjustments For Publication Bias

Figure A.20: Four Approaches to Publication Bias



Robustness to Modelling Assumptions

Figure A.19: Modelling Assumptions



Each bar represents a precision-weighted average estimate for each outcome type, comparing our main specification to different modelling assumptions.

Looking only at well-identified studies

Table 4: Meta-Analysis, F-stat > 20

	(1) Equal Weight Test Scores	(2) Overall Test Scores	(3) Non-Capital Test Score	(4) Capital Test Score	(5) Equal Weight Ed. Attainment	(6) Overall Ed. Attainment
Average Effect	0.0342	0.0329***	0.0480***	0.0121**	0.0561***	0.0551***
	(0.0263)	(0.00799)	(0.00333)	(0.00609)	(0.00833)	(0.00978)
N	18	18	9	9	8	8
SD of $\hat{\theta}_j$'s	0.142	0.037	0.022	0.025	0.034	0.030
τ	0.0809	0.0197	0.00752	0.00890	0.0181	0.0191

Standard errors in parentheses are adjusted for clustering of related papers.

* $p < .1$, ** $p < .05$, *** $p < .01$

Benchmarking

- The magnitude of effects (when compared to those of other interventions) are always larger for educational attainment than test score effects.
- Project STAR: Reducing class size by seven increased test scores by 0.12σ, and college-going by between 1.8 and 2.7 percentage points (Chetty et al. 2011; Dynarski et al. 2013).
  - Pooled $1000 test score effects are equivalent to reducing class size by 1.8 students, while our college-going impacts are equivalent to reducing class size by between 10 and 7.3 students.
- Teacher Quality: Chetty et al. (2014) find that increasing teacher quality by one standard deviation increases test scores by 0.12σ and college going by 0.82 percentage points.
  - Our $1000 test score impacts on test scores and college going are equivalent to increasing teacher quality by 0.26 and 3.4 standard deviations, respectively.
- Test scores may not measure all the benefit to more educational resources.

Conclusions

- School spending studies are more consistent than might appear at first blush.
- On average, a $1000 increase in school spending (sustained over four years) increases test scores by 0.0316σ, high-school graduation by 2.8 percentage points, and college-going by 2.8 percentage points.
- The average effect are unambiguously positive, but there is nontrivial heterogeneity.
- While there is nontrivial heterogeneity, under some reasonable distributional assumptions, one can make relatively precise policy predictions.
    - Benefits for capital spending are similar to those of other spending types after accounting for timing.
    - Benefits are larger for less-advantaged populations.
- Marginal effects are remarkably stable by baseline spending and geography.
- Meta-analysis can provide useful and important insights.