Lecture 8: *p*-hacking

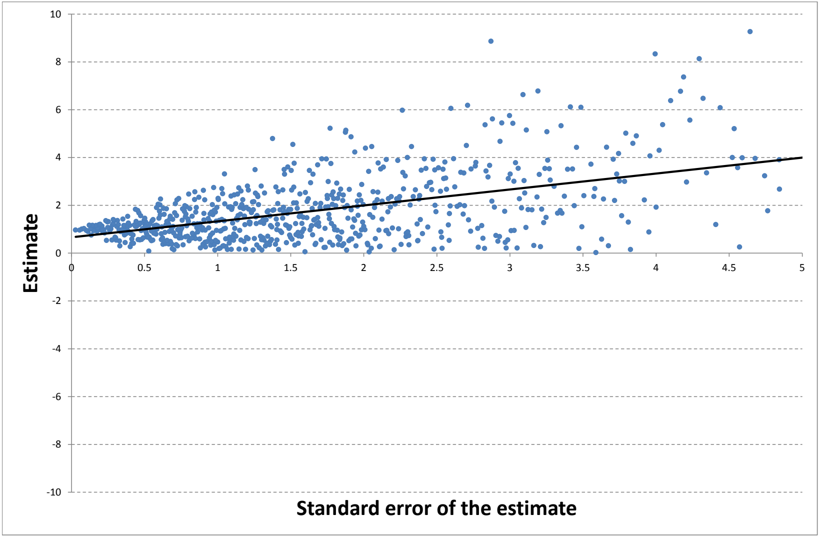
Tomas Havranek

Charles University, CEPR, METRICS

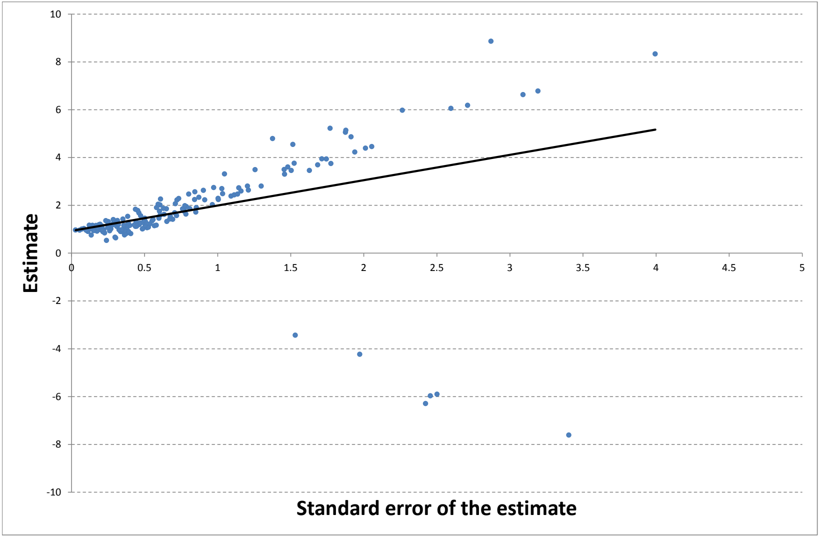
Research Synthesis in Economics and Finance,
University of Canterbury

March 11, 2025

Recall publication bias



Recall publication bias



Key difference

- **Publication bias**: all estimates are individually unbiased
- The entire literature is biased because estimates are selectively reported
- If we know publication probabilities, we can recover the true effect (selection models)
- **P-hacking:** individual estimates can be biased
- Classical selection models fail
- Funnel-based models can work (with adjustments)

True model:

$$Earnings_j = \gamma Education_j + \delta Ability_j + v_j,$$

Ability not observed.

Primary studies:

1. ignore ability $\rightarrow \hat{\gamma}$ too large, SE($\hat{\gamma}$) too small.
2. include a proxy $\rightarrow \hat{\gamma}$ smaller, SE($\hat{\gamma}$) larger.
3. quasi-experiment $\rightarrow \hat{\gamma}$ even smaller, SE($\hat{\gamma}$) even larger.

Example: education premium

Omitted variable:

$$Earnings_j = \gamma Education_j + w_j,$$

Ability not observed.

Primary studies:

1. ignore ability $\rightarrow \hat{\gamma}$ too large, SE($\hat{\gamma}$) too small.
2. include a proxy $\rightarrow \hat{\gamma}$ smaller, SE($\hat{\gamma}$) larger.
3. quasi-experiment $\rightarrow \hat{\gamma}$ even smaller, SE($\hat{\gamma}$) even larger.

Proxy:

$$Earnings_j = \gamma Education_j + \omega IQ_j + x_j,$$

Ability not observed.

Primary studies:

1. ignore ability $\rightarrow \hat{\gamma}$ too large, SE($\hat{\gamma}$) too small.
2. include a proxy $\rightarrow \hat{\gamma}$ smaller, SE($\hat{\gamma}$) larger.
3. quasi-experiment $\rightarrow \hat{\gamma}$ even smaller, SE($\hat{\gamma}$) even larger.

Instrument:

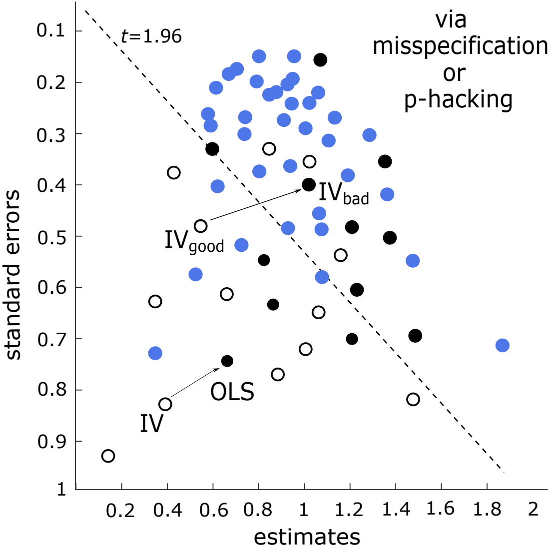
$$Earnings_j = \gamma Education_j + z_j, \quad Distance_j$$

Ability not observed.

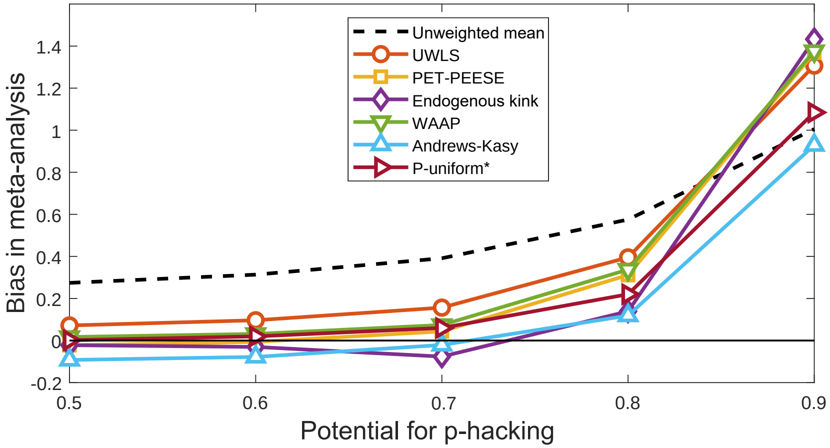
Primary studies:

1. ignore ability $\rightarrow \hat{\gamma}$ too large, SE($\hat{\gamma}$) too small.
2. include a proxy $\rightarrow \hat{\gamma}$ smaller, SE($\hat{\gamma}$) larger.
3. quasi-experiment $\rightarrow \hat{\gamma}$ even smaller, SE($\hat{\gamma}$) even larger.

Some estimates spuriously large & precise



All estimators biased upwards



Example: Changing controls changes precision

Table: Regression of test scores on class size

	(1)	(2)	(3)	(4)
Small class (treatment)	4.82	5.37	5.36	5.37
	(2.19)	(1.26)	(1.21)	(1.19)
White/Asian			8.35	8.44
			(1.35)	(1.36)
Girl			4.48	4.39
			(0.63)	(0.63)
Free lunch			-13.15	-13.07
			(0.77)	(0.77)
White teacher				-0.57
				(2.1)
Teacher experience				0.26
				(0.10)
Master's degree				-0.51
				(1.06)
School intercepts	No	Yes	Yes	Yes
Sample	5,861	5,861	5,861	5,861

Notes: Adapted from Krueger (1999). Dependent variable: test score percentile.
Standard errors in parentheses.

Changing clustering changes precision

Table: Regression of test scores on class size

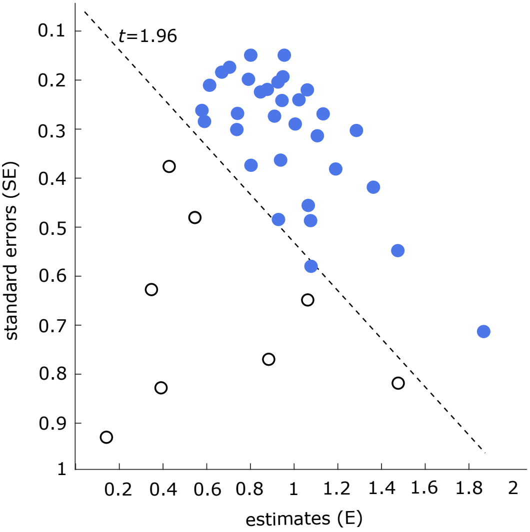
	Krueger (1999)	Replications using different computations of SE				
	(1)	(2) Bootstrap of class clusters	(3) Class clusters	(4) School clusters	(5) Huber-White SE	(6) Plain vanilla SE
Small class	4.82	4.71	4.71	4.71	4.71	4.71
	(2.19)	(2.00)	(1.88)	(1.38)	(0.79)	(0.76)
Sample	5,861	5,743	5,743	5,743	5,743	5,743

Notes: Dependent variable: test score percentile. Standard errors (SE) in parentheses.

Meta-analysis of the class size effect

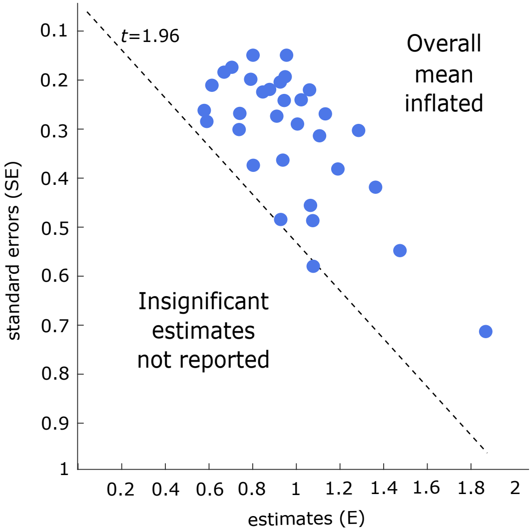
Available at meta-analysis.cz/class. (forthcoming in *JOLE*)

Funnel plot

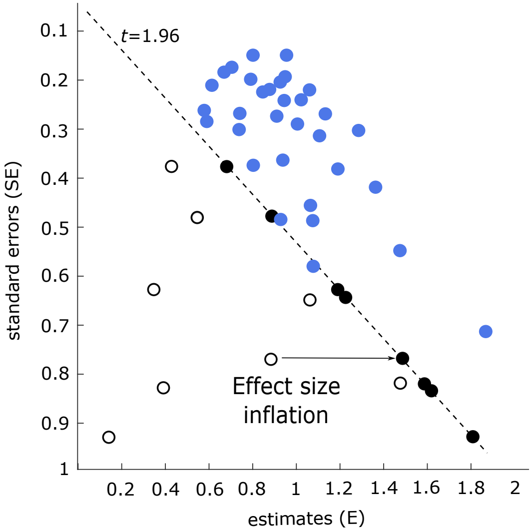


t=1.96
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
standard errors (SE)
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
estimates (E)

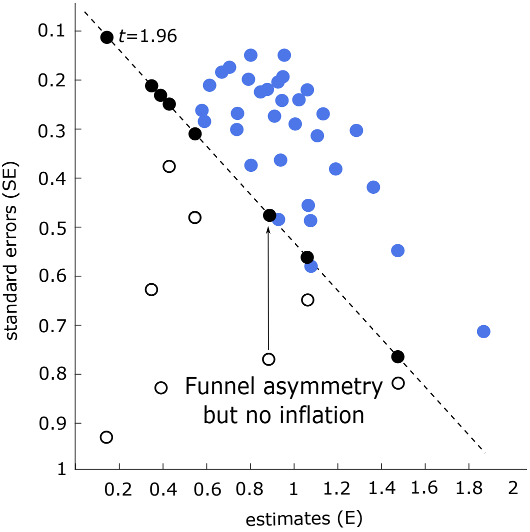
Publication bias



Conventional p-hacking



Spurious precision



PEESE:

$$\hat{E}_i = \overset{\text{Methods or p-hacking}}{E_0} + \beta SE(\hat{E})_i^2 + u_i,$$

Methods or p-hacking

$\text{corr}(SE, u) \neq 0 \Rightarrow \hat{\beta}$ and $\hat{E}_0$ biased.

Natural solution: N instrumenting $SE(\hat{E})_i^2 \rightarrow$ MAIVE.

Options for the meta-analyst

1. Use only quasi-experimental studies.
2. Include controls (dummies for OLS, DID, ...).

   But in observational research
   we never know the true model!

3. Remove "bad" variation from SE $\rightarrow$ MAIVE.

Meta-analysis instrumental variable estimator

MAIVE intuition:

Definition of SE

$$\hat{E}_i = E_0 + \beta SE(\hat{E})_i^2 + u_i. \quad 1/N_i$$

- MAIVE + PEESE = classical IV.
- Can add controls in the 2nd stage.
- Plug MAIVE-adjusted SEs into other estimators.

Meta-analysis instrumental variable estimator

MAIVE first stage:

$$SE(\hat{E})_i^2 = \alpha_0 + \alpha_1\,(1/N_i) + \pi_i. \qquad \text{hacking or misspecifications}$$

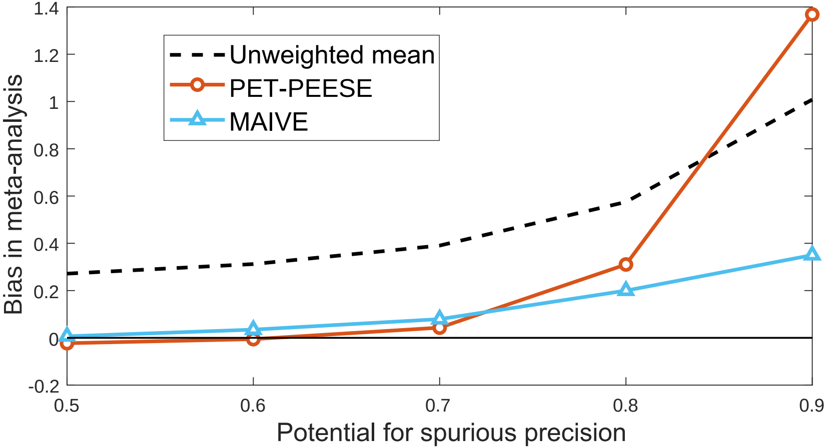
- MAIVE + PEESE = classical IV.
- Can add controls in the 2nd stage.
- Plug MAIVE-adjusted SEs into other estimators.

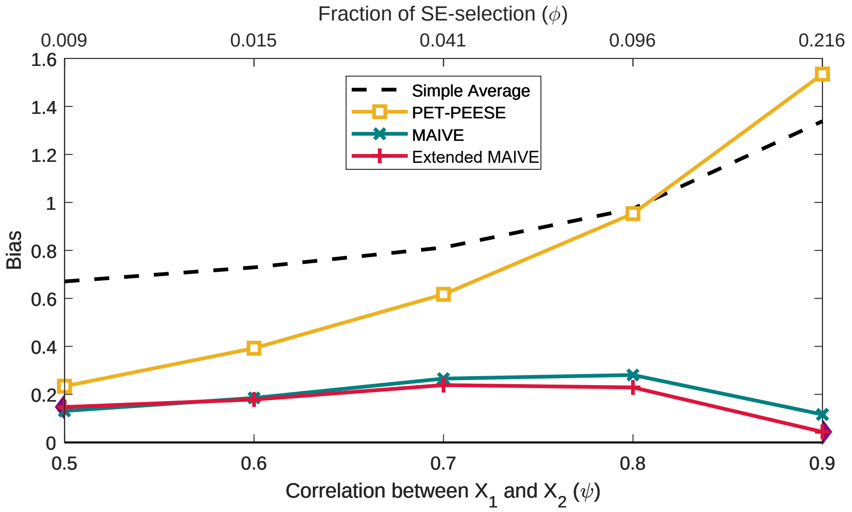
MAIVE adjustment:

$$SE(\hat{E})^2_{adj,i} = \hat{\alpha}_0 + \hat{\alpha}_1 \,(1/N_i) + \cancel{\pi_i}. \qquad \text{~~hacking~~ or ~~misspecifications~~}$$

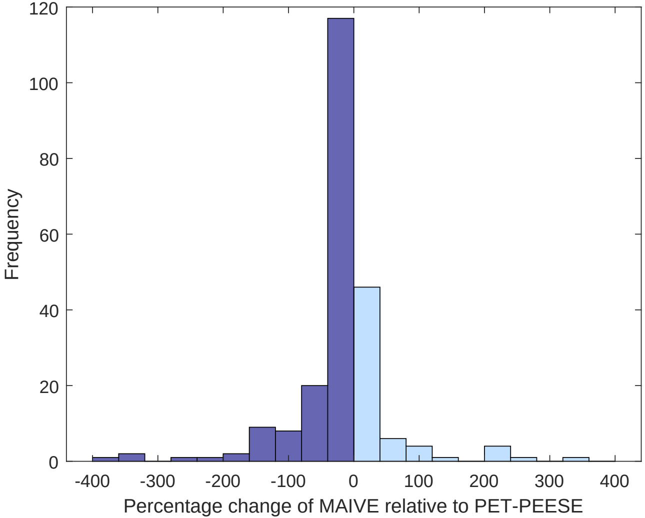
- MAIVE + PEESE = classical IV.
- Can add controls in the 2nd stage.
- Plug MAIVE-adjusted SEs into other estimators.

MAIVE alleviates the bias




Fraction of SE-selection (ϕ)
0.009
0.015
0.041
0.096
0.216
Simple Average
PET-PEESE
MAIVE
Extended MAIVE
Bias
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.5
0.6
0.7
0.8
0.9
Correlation between X_1 and X_2 (ψ)

MAIVE reduces PET-PEESE in 70% of the cases



Main consideration

- Can methods or p-hacking influence both $\hat{E}$ and SE?
- Yes $\rightarrow$ MAIVE helps.
- No $\rightarrow$ MAIVE doesn't hurt much.

Project Website

meta-analysis.cz/maive

Bonus: Maya Mathur's p-hacking correction

- Right-truntcated meta-analysis (RTMA, Mathur 2024, RSM)
- Assumption 1: insignificant estimates are NOT p-hacked
- Assumption 2: estimates are normally distributed
- Assumption 3: there are enough insignificant estimates published that we can recover the underlying distribution
- Bayesian techniques used for better convergence