Weighted Adjustment Instrumental Variable Estimator

Tomáš Havránek

Charles University • Meta-Research Innovation Center at Stanford

Zuzana Iršová

Charles University

Pedro Bom

University of Deusto

Heiko Rachinger

Univ. of the Balearic Islands

SRSM 2026 Annual Meeting • Chania, Greece

Reported precision is a choice

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	4.82	4.82	4.82	4.82	4.82
	(2.19)	(2.00)	(1.88)	(1.38)	(0.79)	(0.76)
Sample	5,861	5,861	5,861	5,861	5,861	5,861

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

Meta-analysis of the class size effect

Available at meta-analysis.cz/class. Spoiler: effect size tiny.

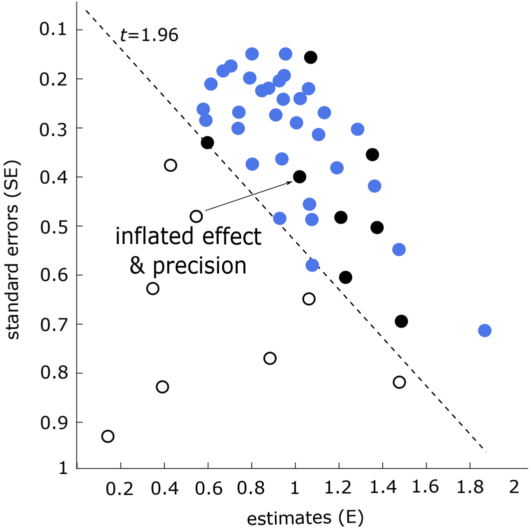
Changing controls changes precision

Table: Regression of test scores on class size

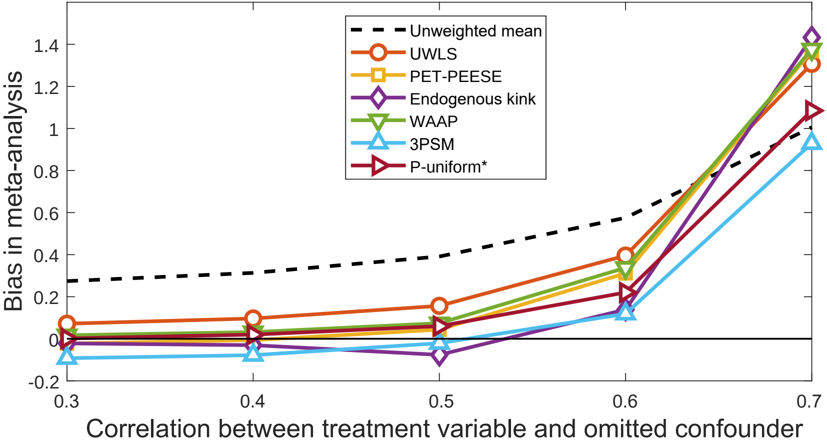
	(1)	(2)	(3)	(4)
Small class (treatment)	4.82	5.37	5.36	5.37
	(2.19)	(1.26)	(1.21)	(1.19)
White/Asian			8.35	8.44
			(1.35)	(1.36)
Girl			4.48	4.39
			(0.63)	(0.63)
Free lunch			-13.15	-13.07
			(0.77)	(0.77)
White teacher				-0.57
				(2.1)
Teacher experience				0.26
				(0.10)
Master's degree				-0.51
				(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.

P-hacking possible on both effect and precision



Cures worse than the disease?



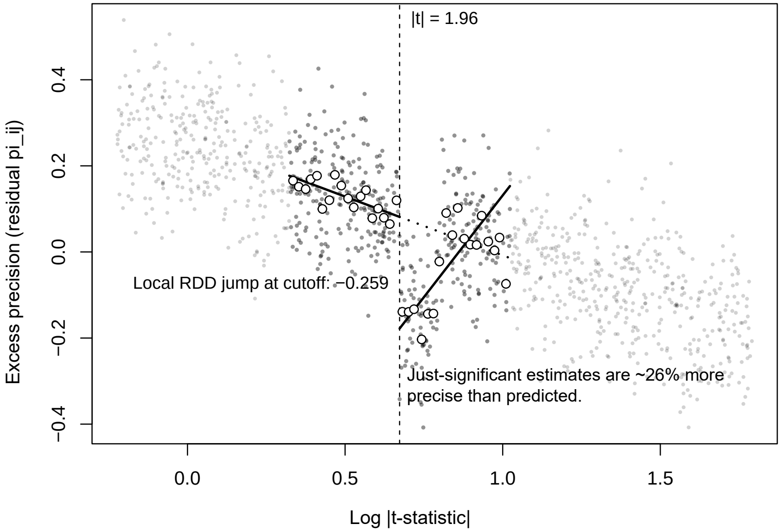
Diagnostic: excess precision

Residual from meta-regression:

$$\log \mathrm{SE}(\hat{E}_i) = \alpha_0 + \alpha_1 \log N_i + \underbrace{\pi_i}_{<0:\ \text{excess precision}}$$

- Negative $\pi_i \rightarrow$ reported SE too small relative to N $\rightarrow$ may indicate p-hacking.
- Interpretation: $\pi_i = -0.25$ means SE is roughly one quarter smaller than expected.

Residual discontinuity test



Weighted Adjustment Instrumental Variable Estimator

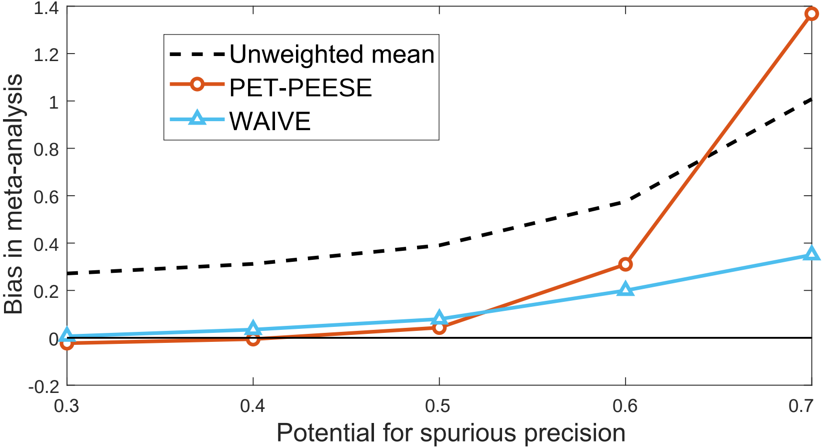
WAIVE combines two adjustments:

1. Replace reported SEs by the component predicted from sample size (instrument).
2. Downweight spuriously precise estimates near significance cutoffs. Exponential tilt:

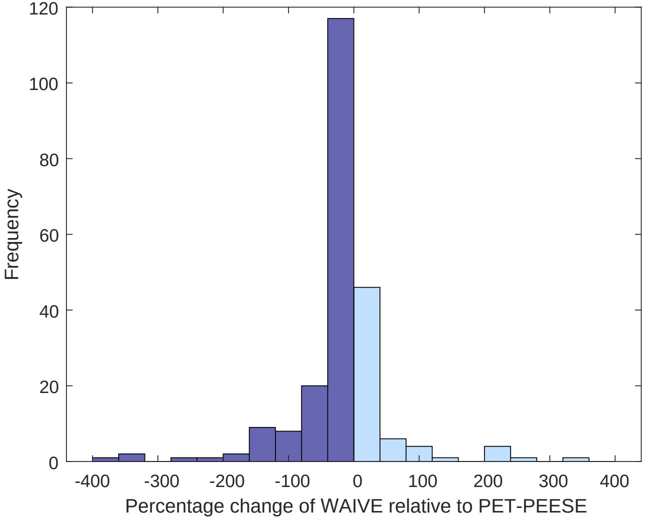
$$\omega_i = \exp[-D_i \max(0, -\pi_i)],$$

- D_i is high when $|t_i|$ is close to 1.96.
- Weight $\omega_i \in (0, 1]$; only excess precision penalized.

WAIVE alleviates the bias in simulations



WAIVE shrinks PET-PEESE in applications



Practical takeaway

Bottom line

MAIVE (*Nature Comms* 2025) instruments the standard error; **WAIVE** adds a downweighting layer for studies with excess precision near significance cutoffs.

When to use

- Can methods or *p*-hacking influence both $\hat{E}$ and SE?
- Yes $\rightarrow$ MAIVE / WAIVE help (EasyMeta.org).
- No $\rightarrow$ MAIVE / WAIVE may be too conservative; standard methods suffice.

Project website

meta-analysis.cz/maive

EasyMeta.org! Papers on instrumental meta:

- Irsova Z., P. Bom, T. Havranek, & H. Rachinger (2025): Spurious Precision in Meta-Analysis of Observational Research.
  Nature Communications 16: 8454.
- Irsova Z., T. Havranek, X. Bortnikova, & F. Bartos (2026): Meta-Analysis of Field Studies on Beauty and Professional Success.
  Nature Human Behaviour, R&R.
- Cala P., T. Havranek, Z. Irsova, M. Luskova, J. Matousek & J. Novak (2026): Financial Incentives and Performance: A Meta-Analysis of Experiments in Economics.
  Journal of Political Economy Microeconomics, forthcoming.

Technical appendix

Backup slides for discussion.

Where to look for more

- **Method paper.** Irsova, Bom, Havranek & Rachinger (2025), *Nature Communications* 16:8454.
- **Tools and replication.** easymeta.org · meta-analysis.cz/maive · MAIVE package on CRAN

Excess precision: log first-stage residual

First stage (log form; extension over published MAIVE)

$$\log(\mathrm{SE}_i) \;=\; \alpha + \beta \,\log(N_i) + \gamma' X_i + \pi_i.$$

- $\pi_i < 0 \;\Rightarrow\;$ reported SE smaller than sample size predicts.
- X_i: pre-specified design features; *not* researcher-chosen inference tweaks.
- Interpretation: $\pi_i = -0.10 \Rightarrow$ SE $\approx$ 10% below prediction (exact: $1 - e^{-0.10} = 9.5\%$).

Foundational first stage from MAIVE (Nature Comms 2025); log version used going forward.

Downweighting: D_i on $\log|t|$

Triangular danger-zone kernel (Chania operational default)

$$D_i \;=\; \max\left\{0,\; 1-\frac{\left|\log|t_i|-\log(1.96)\right|}{h}\right\}, \quad t_i=\hat{\beta}_i/\mathrm{SE}_i.$$

- Compact support: $D_i = 0$ outside bandwidth h; same scale as RDT.

WAIVE weight

$$\omega_i \;=\; \exp[-D_i \max(0,-\pi_i)] \;\in\; (0,1].$$

Penalizes only estimates with excess precision *and* $|t|$ near the cutoff.

RDT: residual discontinuity at the significance cutoff

Test of continuity in $E[\pi_i \mid r_i]$ at $r = c$

$$
\begin{aligned}
r_i &= \log|t_i|, \qquad c = \log(1.96), \\
\tau_{\mathrm{RDT}} &= \lim_{r \downarrow c} E[\pi_i \mid r_i = r] - \lim_{r \uparrow c} E[\pi_i \mid r_i = r].
\end{aligned}
$$

- Null: $E[\pi_i \mid r_i]$ continuous at c.
- $\tau_{\mathrm{RDT}} < 0$: just-significant SEs systematically below prediction.
- $\tau = -0.10$: SEs $\approx$ 10% smaller than predicted ($1 - e^{-0.10} = 9.5\%$ exact).

RDT: estimation and falsification battery

Estimation

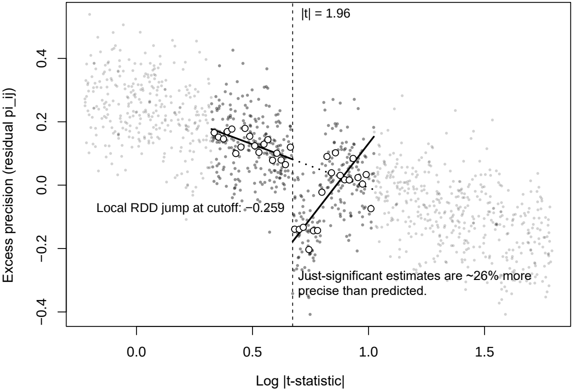
- Outcome: $\hat{\pi}_i$
- Running variable: $\log |t_i|$
- Local linear, both sides
- Triangular kernel
- MSE-optimal bandwidth (Calonico et al. 2014)
- Cluster bootstrap by study

Falsification battery

- Donut hole around $|t| = 1.96$
- Placebo cutoffs ($|t| = 1.5, 2.5$)
- Density tests (McCrary; Cattaneo et al.)
- Bandwidth sensitivity
- Mechanism check: residualize $\log |\hat{\beta}|$ as placebo outcome

No global polynomial; identification is intentionally local.

RDT: the picture



Local jump $\hat{\tau}_{\mathrm{RDT}} \approx -0.259$: substantial concentration of excess precision just above $|t| = 1.96$.

Simulation: data-generating process

Primary studies

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + u, \qquad X_2 = \psi X_1 + (1-\psi)\,\varepsilon.$$

- True effect $\alpha_1 = 1$ (bias on the same scale).
- X_2: true control; replacing it with weaker proxies creates OVB, stronger as ψ rises.
- $M = 80$ studies; sample sizes calibrated to published meta-analyses.
- 50% potential p-hackers; search over proxy controls until $t > 1.96$ or $H = 50$ attempts.

Simulation design: MAIVE (Nature Comms 2025); tools at easymeta.org.

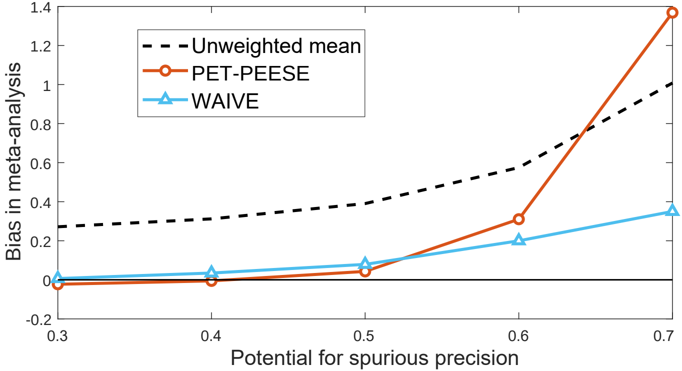
Simulation: what is being selected?

Specification search moves *both* estimate and precision

Changing the conditioning set perturbs $\hat{\beta}_i$ and SE_i simultaneously; the accepted specification may have a larger estimate, a smaller SE, or both.

- Estimate channel. Accepted spec yields larger $|\hat{\beta}_i|$.
- Precision channel. Accepted spec yields smaller SE_i — the MAIVE/WAIVE target.
- Conventional funnel-based corrections treat SE_i as exogenous and cannot see the precision channel.

Simulation: WAIVE alleviates the bias



WAIVE bias stays close to zero as the potential for spurious precision rises; PET-PEESE diverges.

Experimental WAIVE layer; simulation design from MAIVE (Nature Comms 2025).