MAER-Net 2026 PhD Workshop
Chemnitz, 23 September 2026

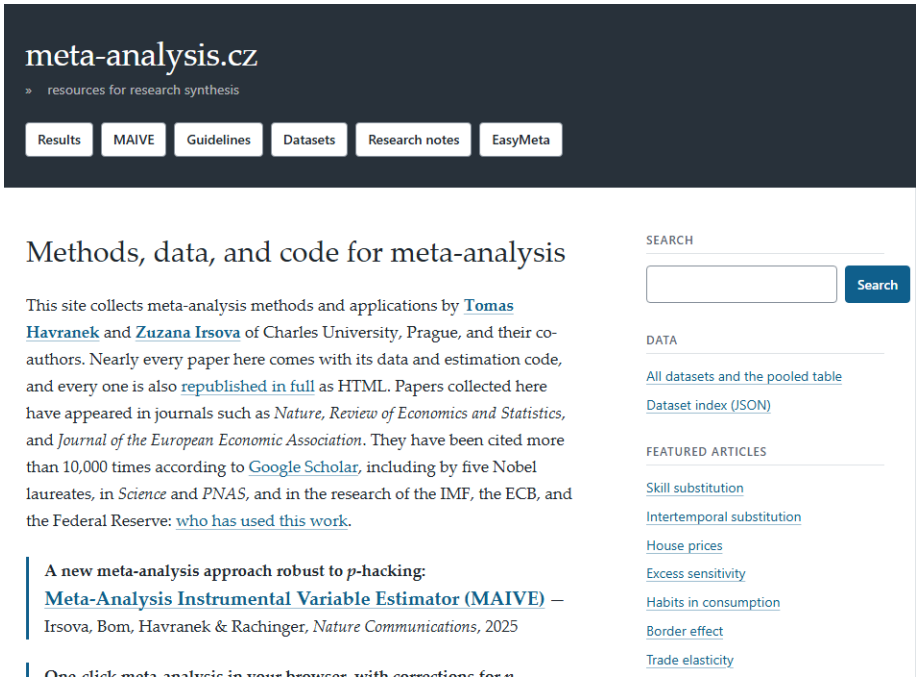
EasyMeta.org and AI in Meta-Analysis

Tomas Havranek

Charles University, Prague

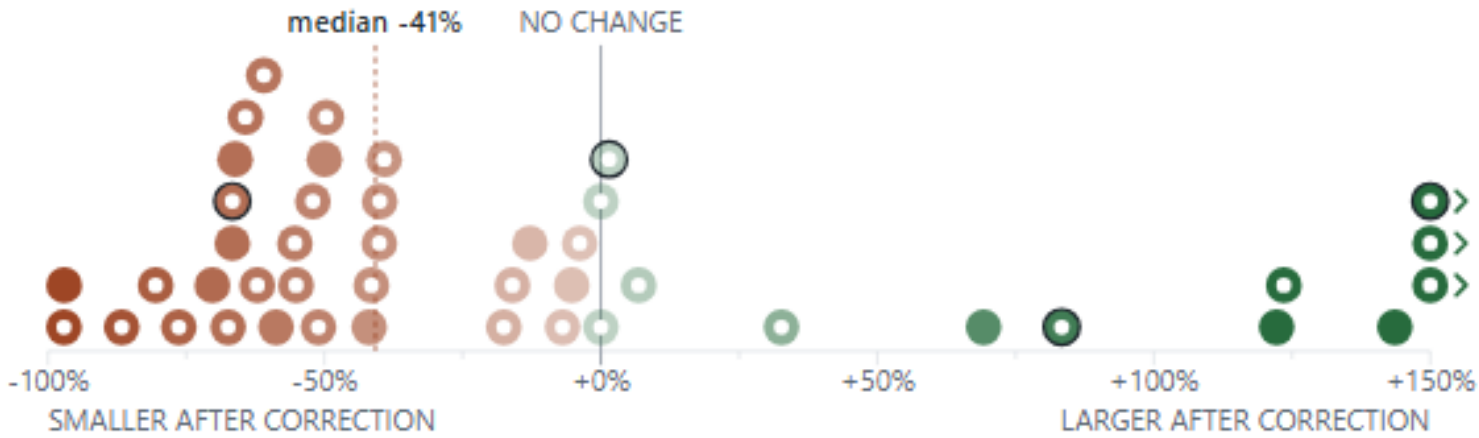
Meta-Research Innovation Center at Stanford

meta-analysis.cz



- 71 papers in full, most with data and code
- 46 datasets; one pooled table of about 50,000 estimates
- R replication for 37 papers
- A page written for AI assistants

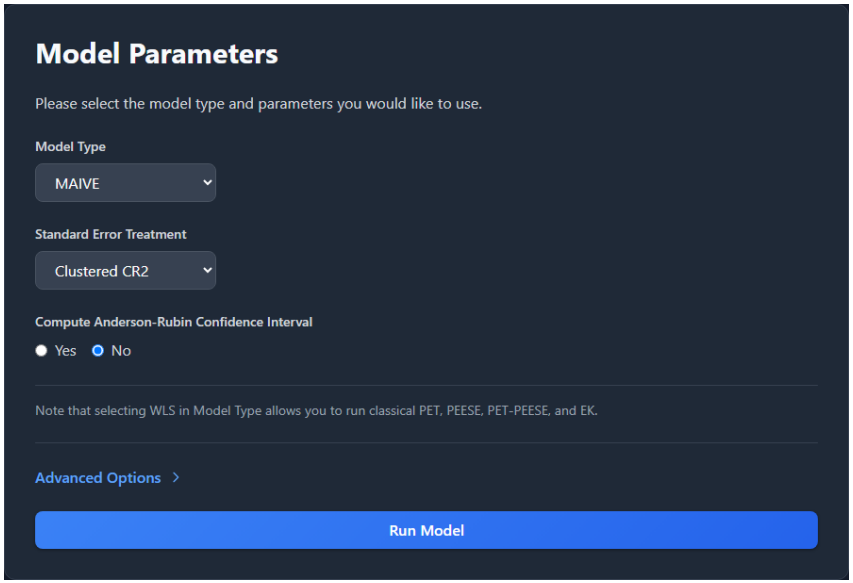
What correction does to 44 of our results



31 moved toward zero, 11 away from it, 2 did not move.

Median revision: −41%.

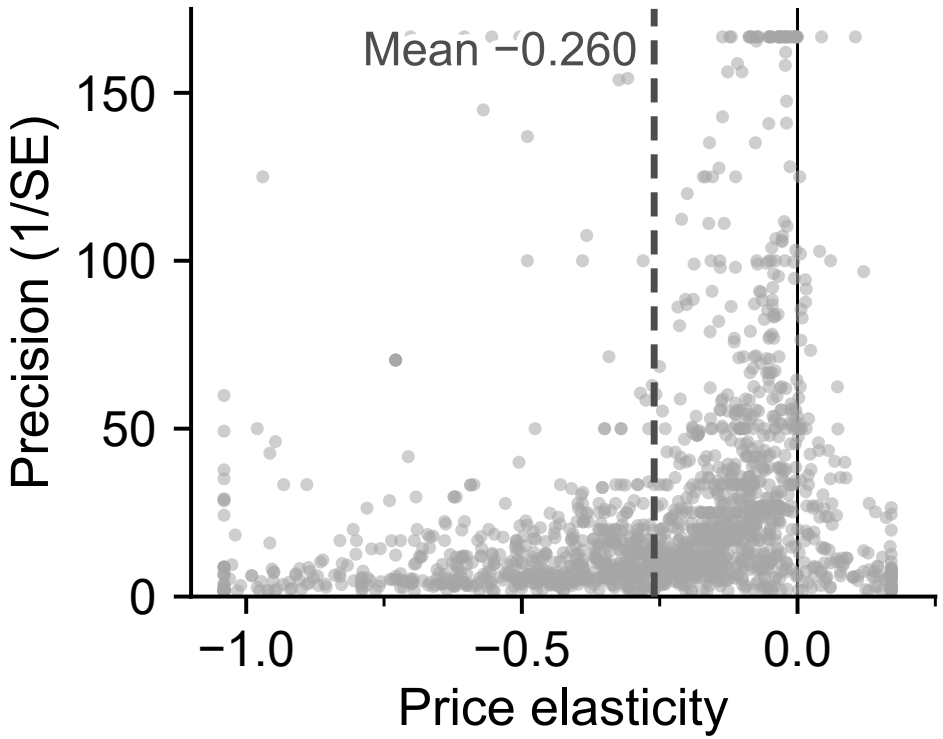
easymeta.org



- Upload effect, SE, sample size, study ID
- PET-PEESE, EK, MAIVE, WAIVE, RTMA, RDT
- Clustering, winsorizing, weights, Anderson-Rubin intervals
- A reproducibility package to download

Built by Petr Cala. Also runs through an API.

Demo data: electricity demand



Short-run price elasticities: 1,647 estimates from 226 studies, winsorized 2.5%

Kudela, Havranek, Irsova, Kudelova & Sikl (2026), meta-analysis.cz/electricity

Try it

1. Download from meta-analysis.cz/chemnitz
2. Upload at easymeta.org
3. Model WLS, winsorize 2.5%, Run

One click per estimator

	Elasticity	SE
Simple mean (winsorized)	−0.260	
PET	−0.105	0.033
PEESE	−0.127	0.027
EK	−0.105	0.033
PEESE, not clustered	−0.127	0.013
PEESE, winsorize 1%	−0.118	0.029
PEESE, equal weights	−0.226	0.019

PET-PEESE picks PEESE. EK finds no kink (too much heterogeneity), so it equals PET. The paper, with a richer model: −0.16.

The same run from Claude

Pasted into Claude Code (excerpt)

```
Use the EasyMeta API, not your own estimation. [...]
Send every row unchanged to
POST https://api.maive.eu/v1/run-model [...]
{"recipe":"PET-PEESE","parameters":{"winsorize":2.5},
 "data":[...]} and the same with "recipe":"EK". [...]
Never report a number the API did not return.
```

	App	API, through Claude
PET-PEESE	−0.127	−0.127104860252972
EK	−0.105	−0.10525106160908

Matching digits show only that the code ran as asked.

Full prompt: meta-analysis.cz/chemnitz

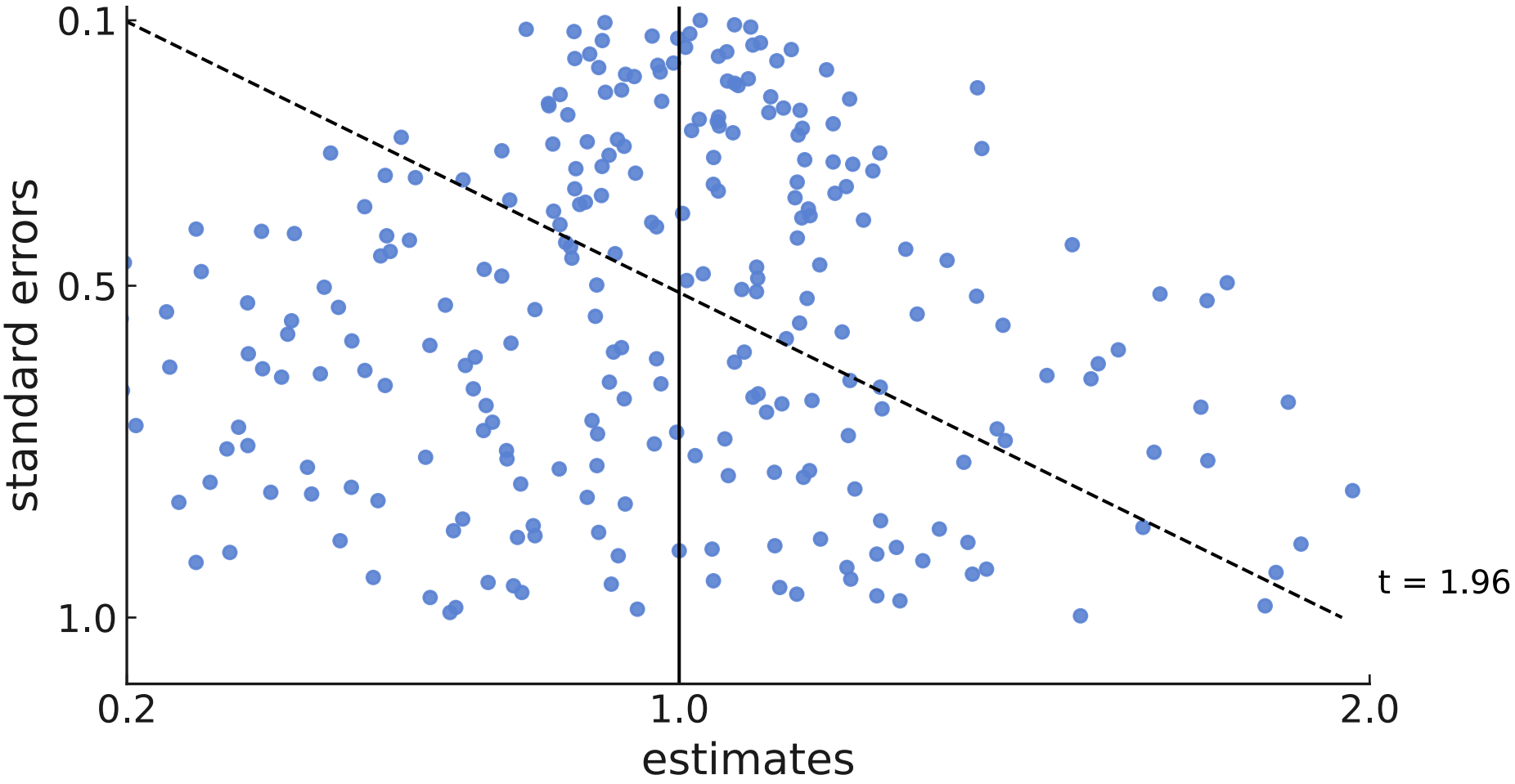
MAIVE/WAIVE

MAIVE Meta-Analysis Instrumental Variable Estimator

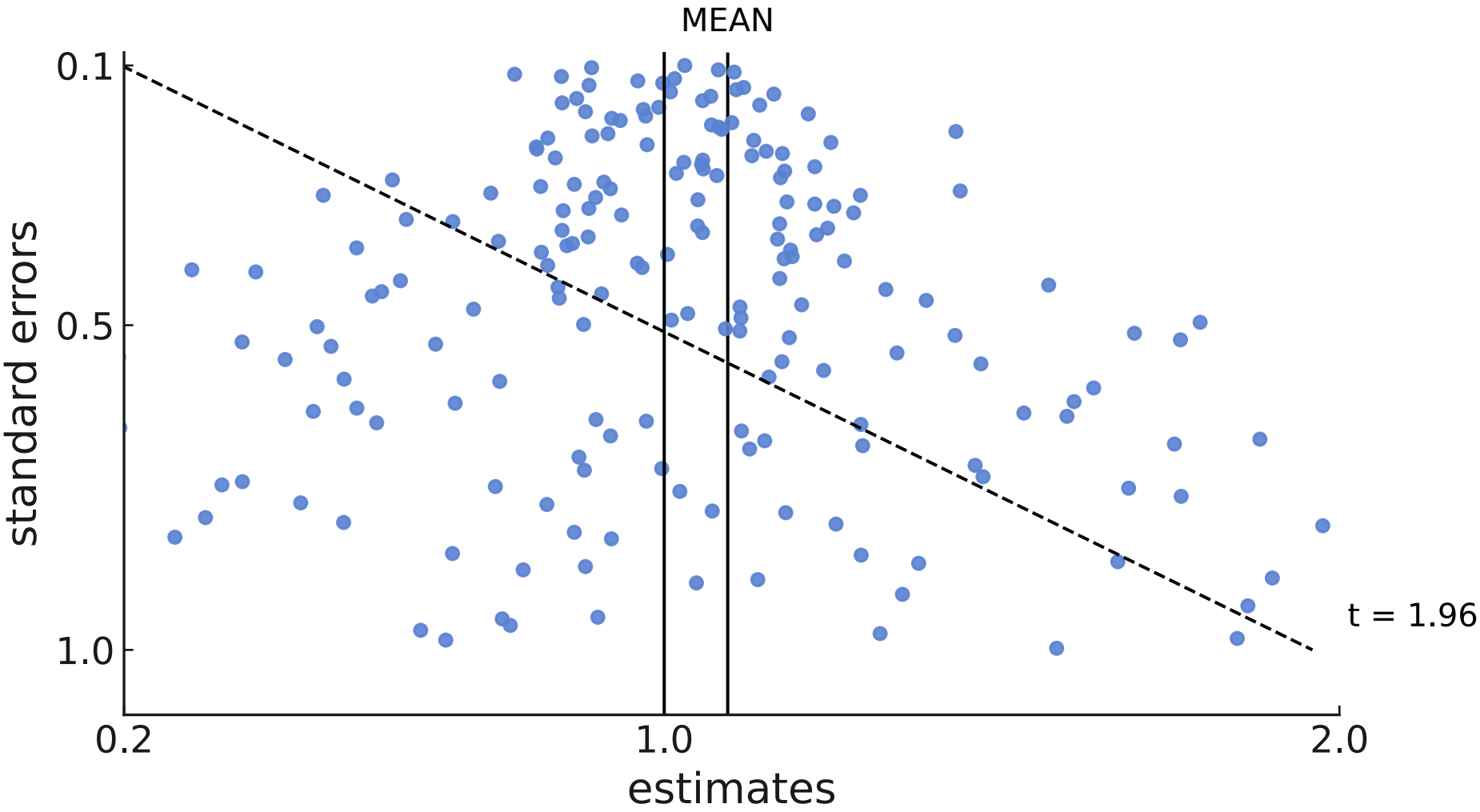
WAIVE Weighted Adjustment Instrumental Variable Estimator

Joint work with Pedro Bom, Petr Cala, Zuzana Irsova and Heiko Rachinger

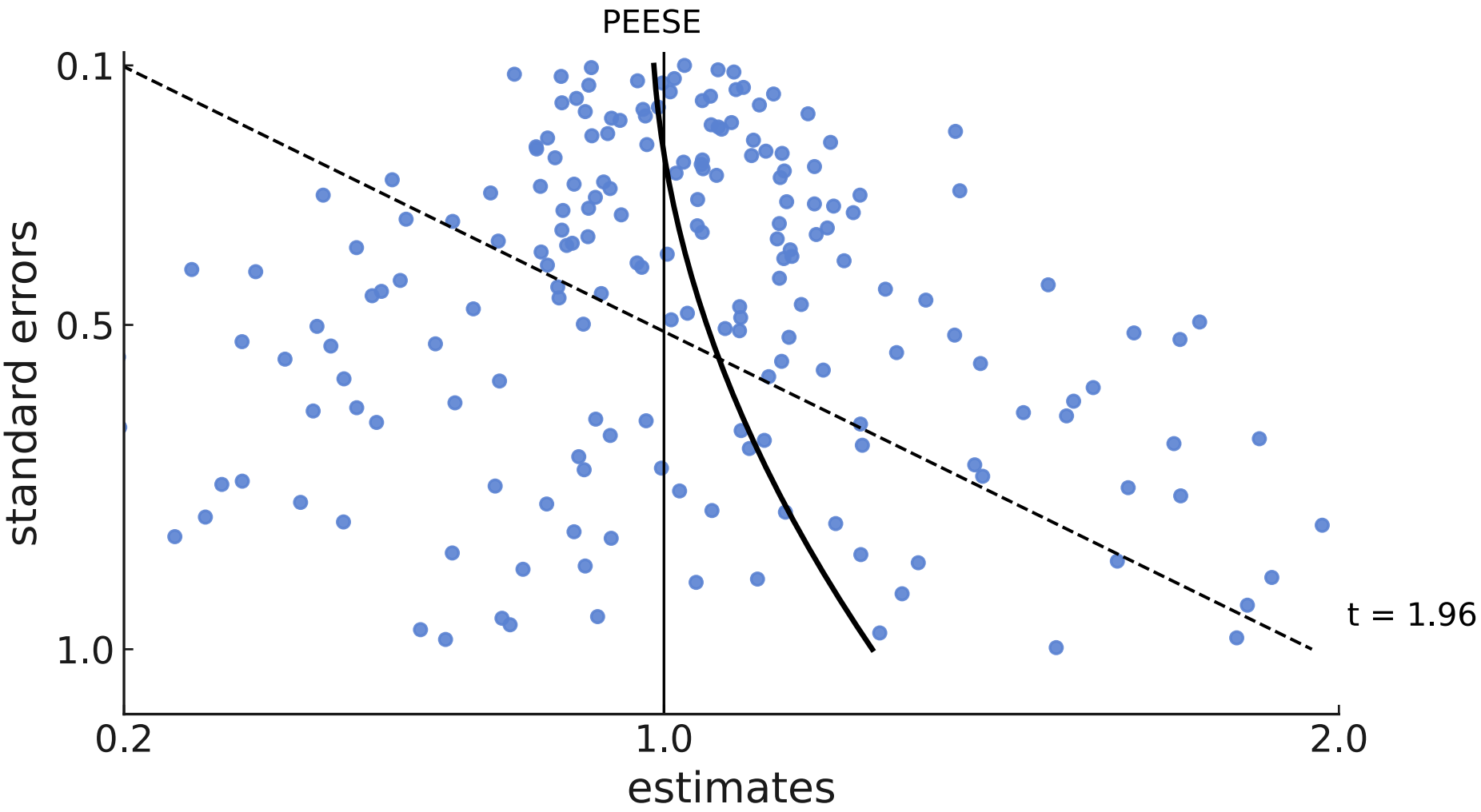
No selection, no bias



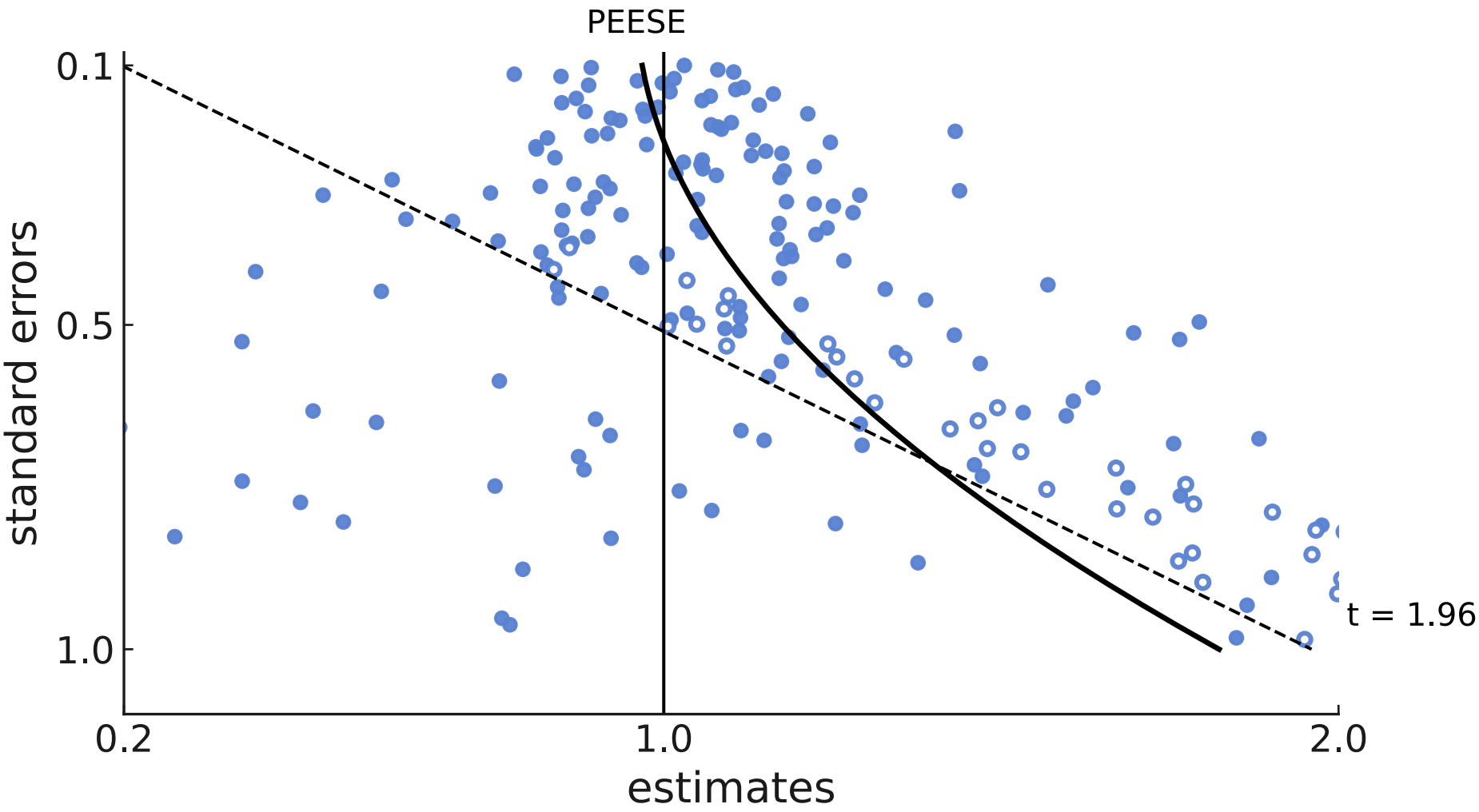
Publication bias: mean too large



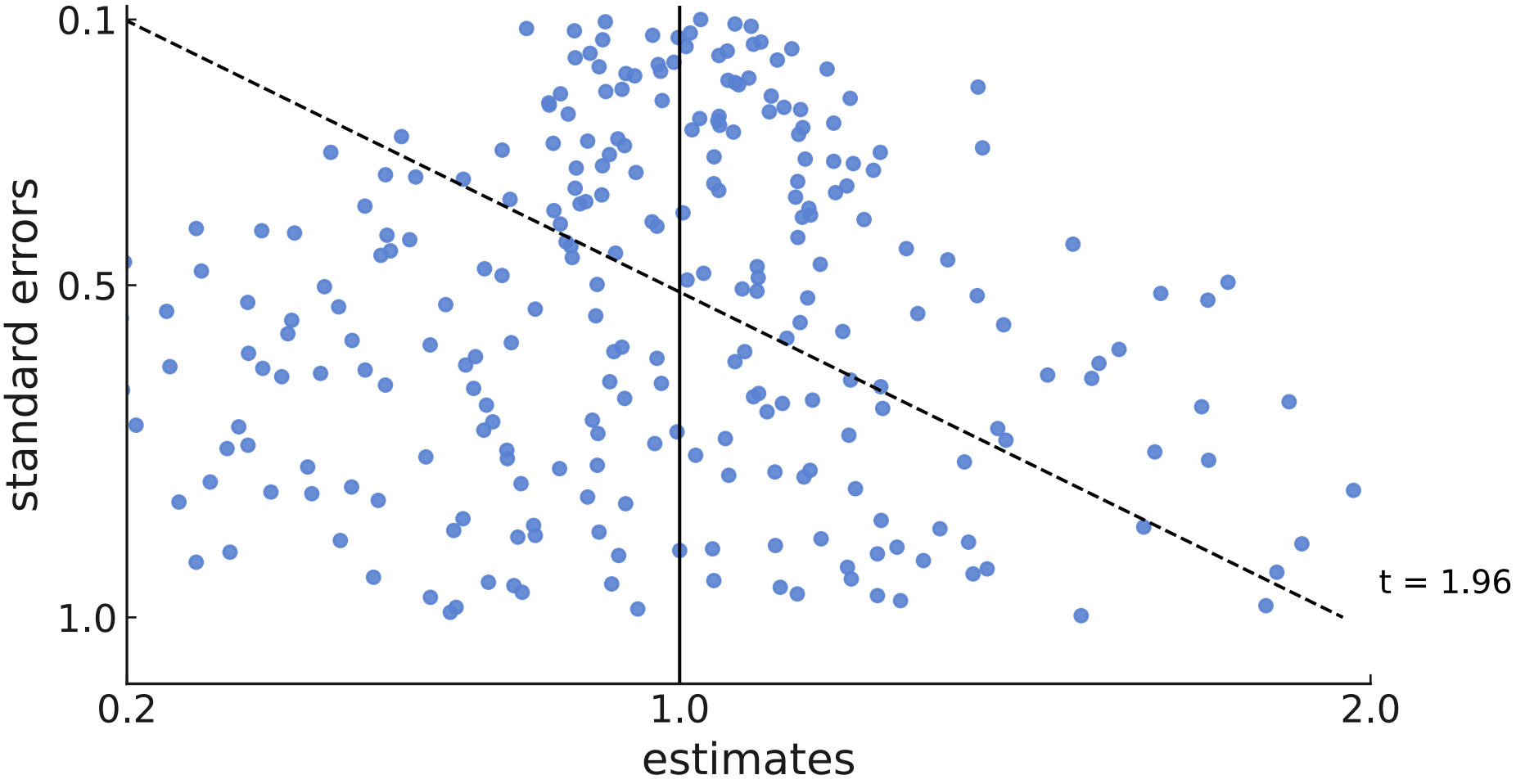
Publication bias: PEESE works



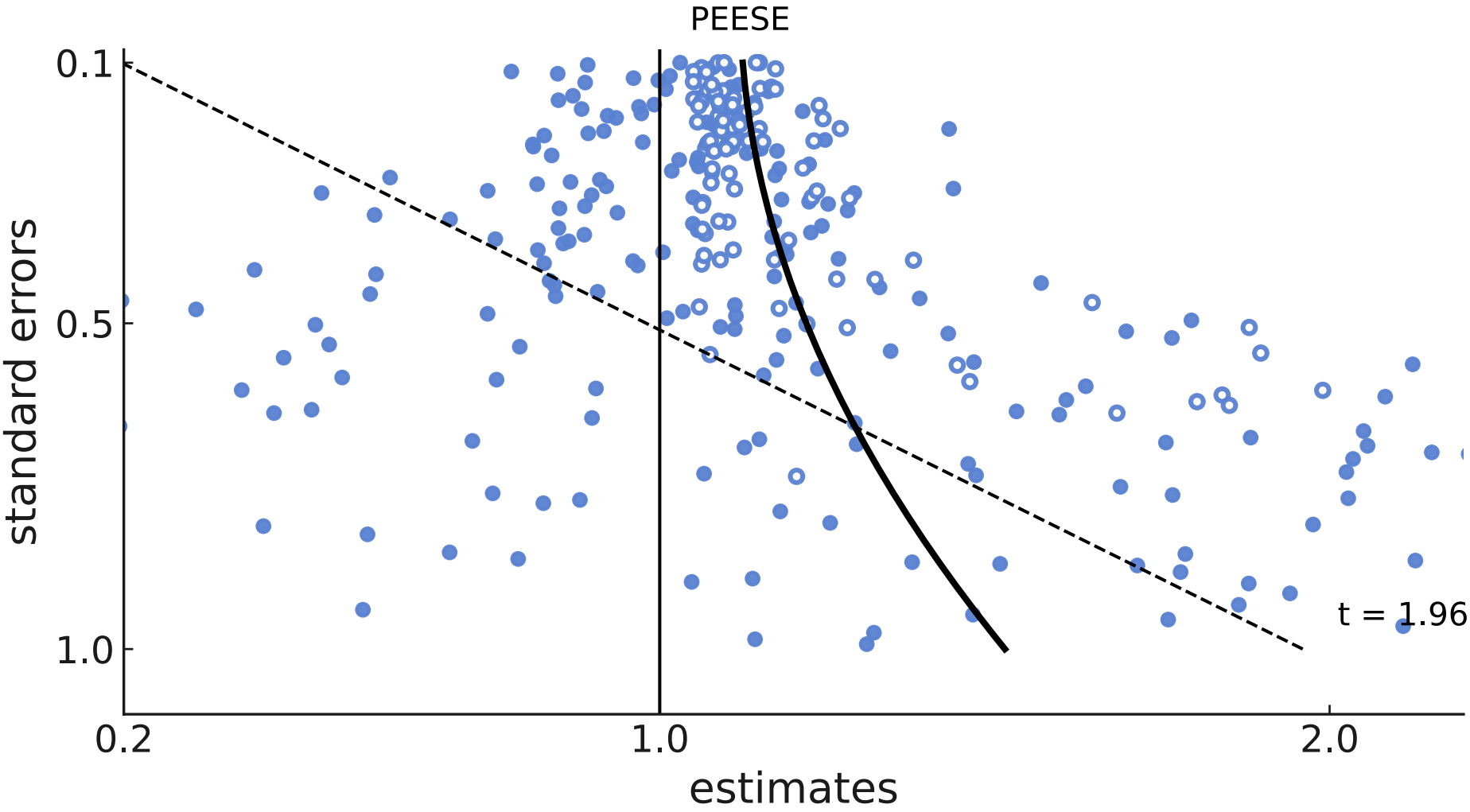
p-hacking on estimates: PEESE works



Back to no selection

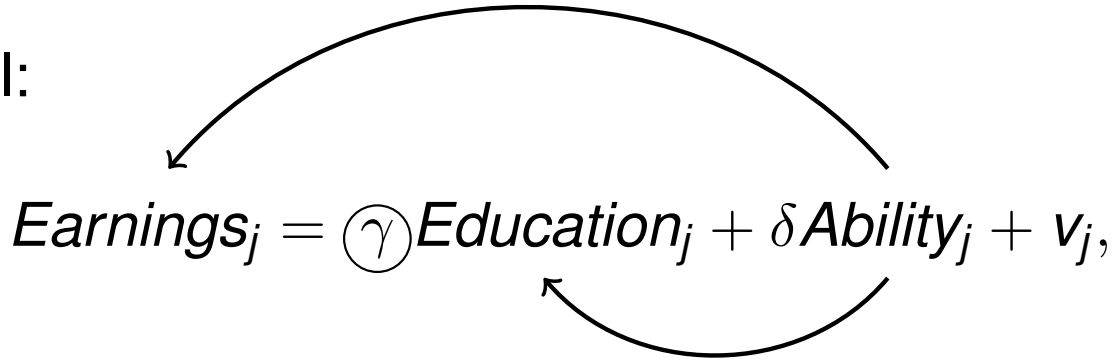


Aggressive *p*-hacking: PEESE in trouble



Could this happen? Meta of education premium

True model:



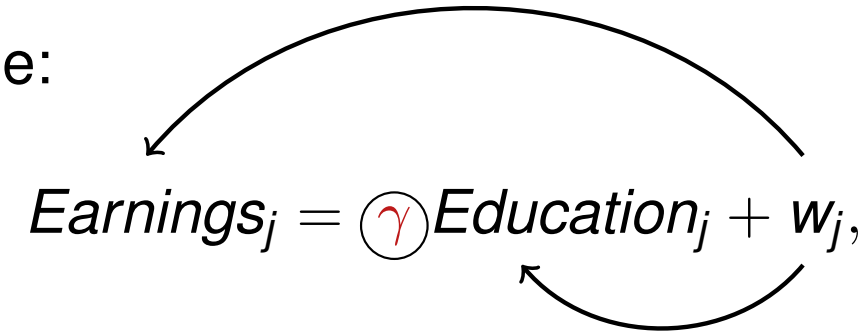
Ability not observed.

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.

Could this happen? Meta of education premium

Omitted variable:



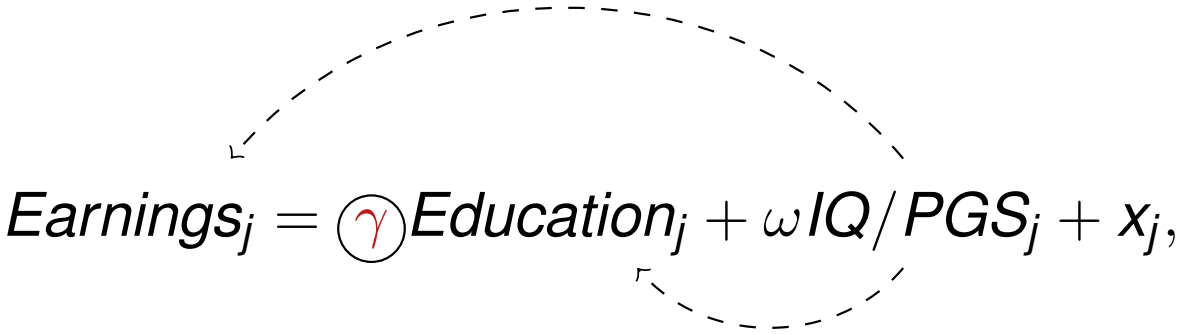
Ability not observed.

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.

Could this happen? Meta of education premium

Proxy:



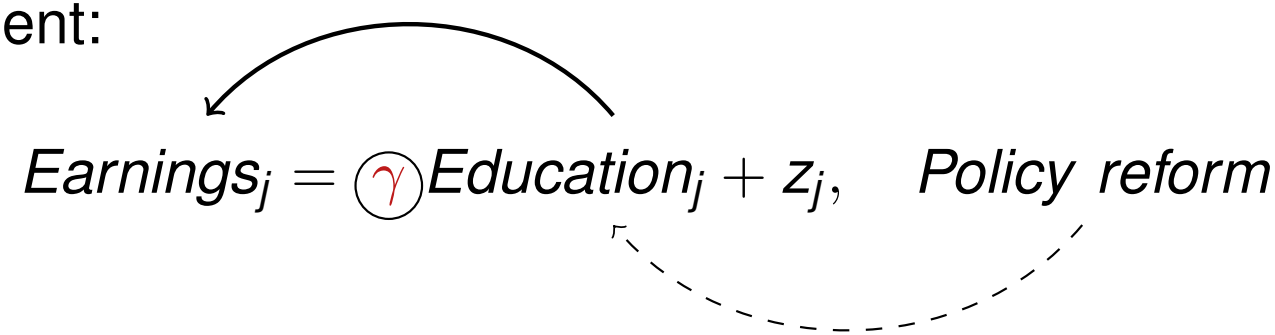
Ability not observed.

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.

Could this happen? Meta of education premium

Instrument:

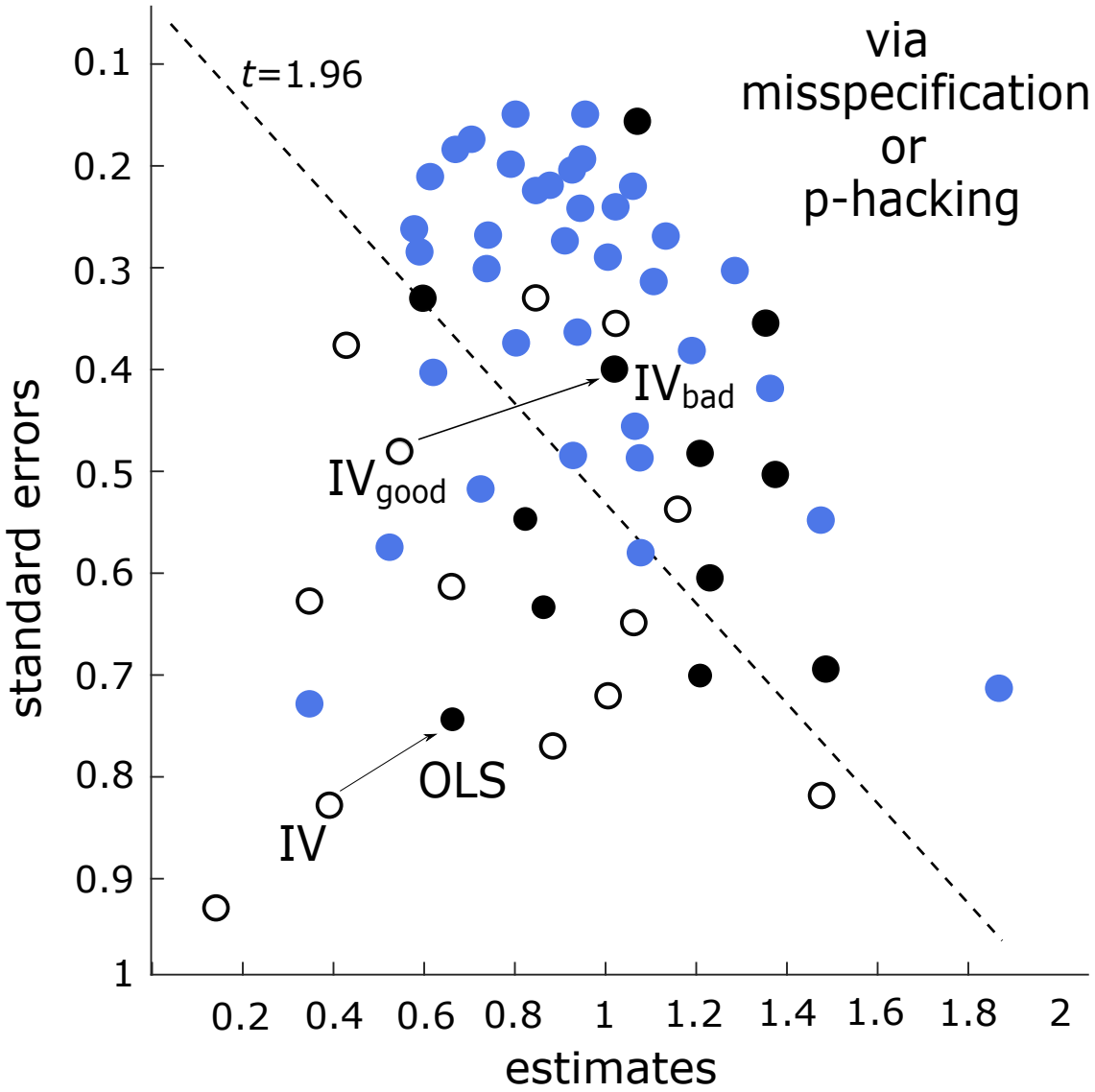


Ability not observed.

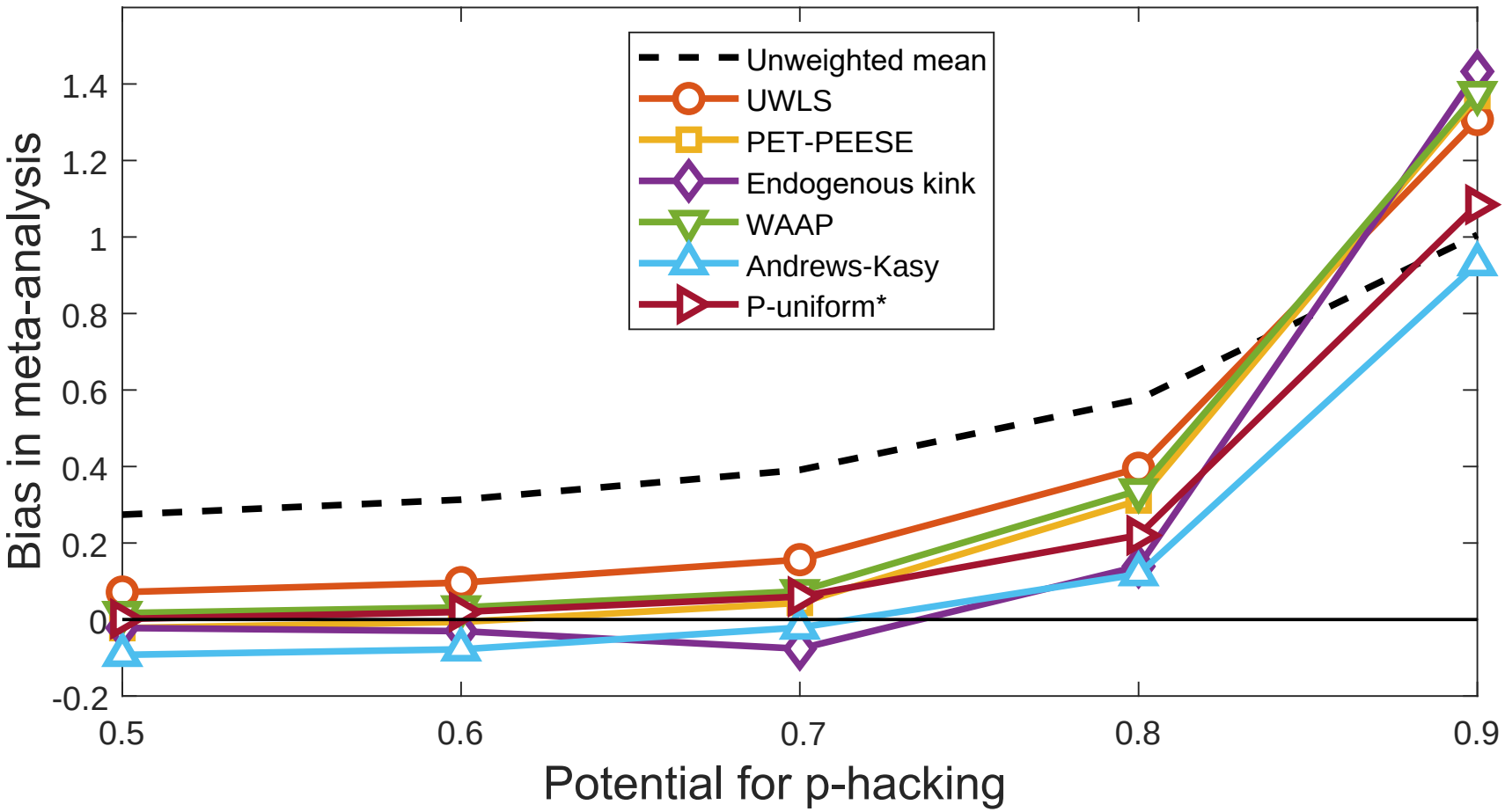
Primary studies:

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

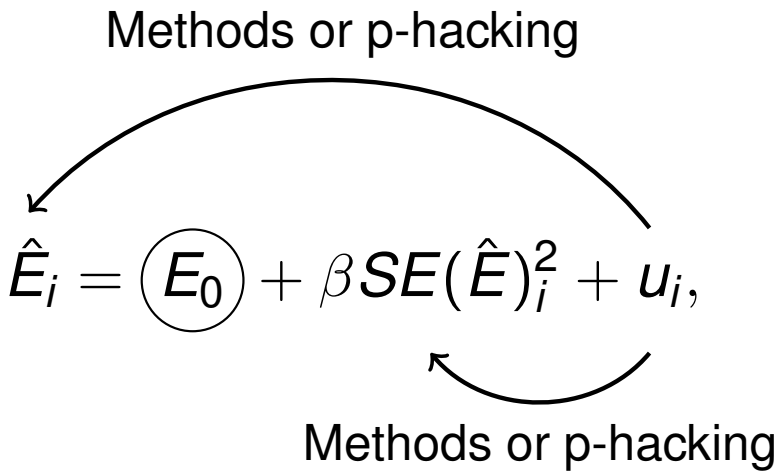


All meta estimators biased upwards



Why? Key meta assumption broken

Traditional meta:



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

Natural solution: 1/N instrumenting $SE(\hat{E})_i^2 \rightarrow$ MAIVE.

Meta-analysis instrumental variable estimator

MAIVE intuition:

Definition of SE

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

- Can add controls.
- Plug MAIVE-adjusted SEs into other estimators.
- In practice it's better to use logs in the first stage.

Meta-analysis instrumental variable estimator

MAIVE first stage:

$$SE(\hat{E})_i^2 = \alpha_0 + \alpha_1 (1/N_i) + \pi_i.$$

hacking or misspecifications

- Can add controls.
- Plug MAIVE-adjusted SEs into other estimators.
- In practice it's better to use logs in the first stage.

Meta-analysis instrumental variable estimator

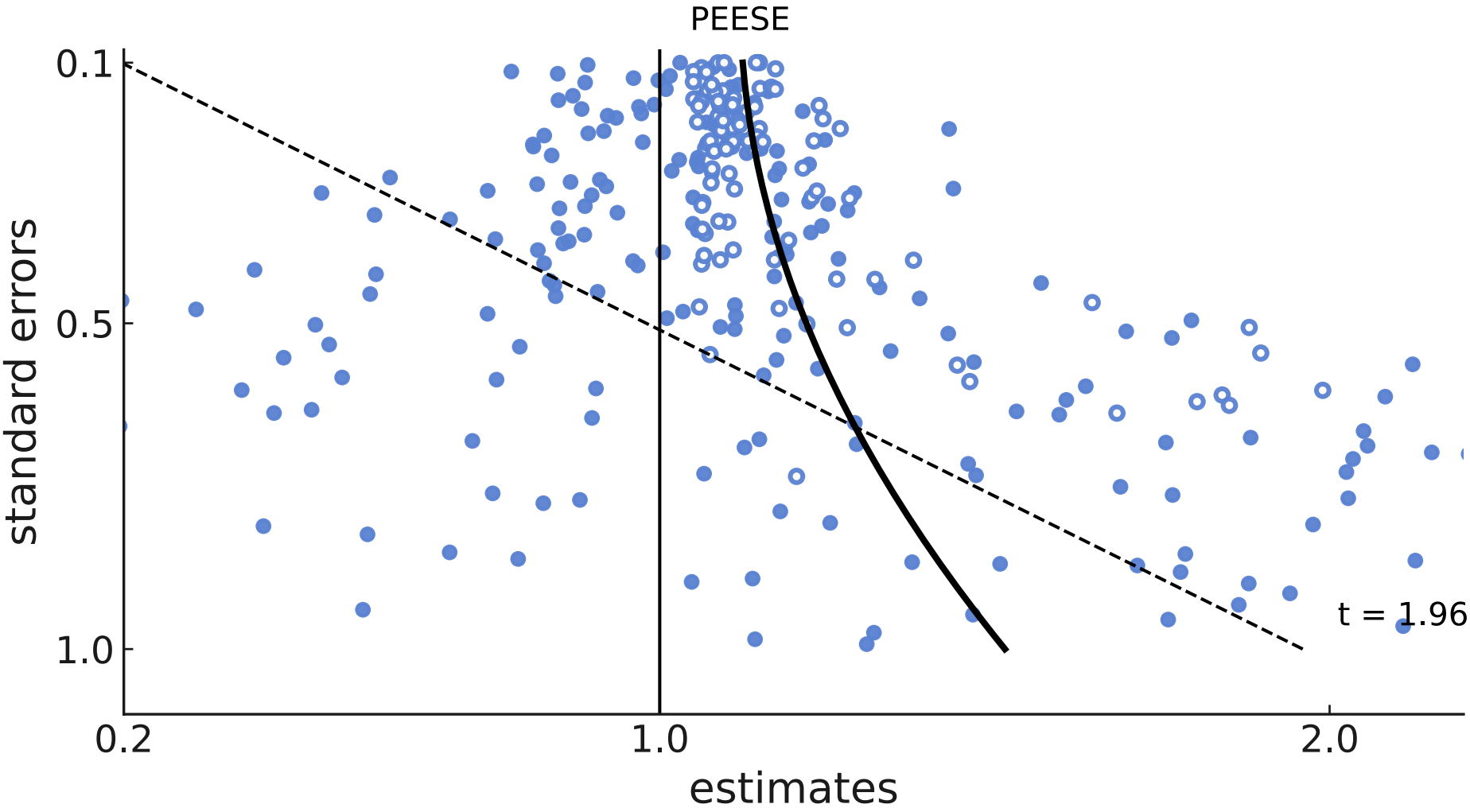
MAIVE adjustment:

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

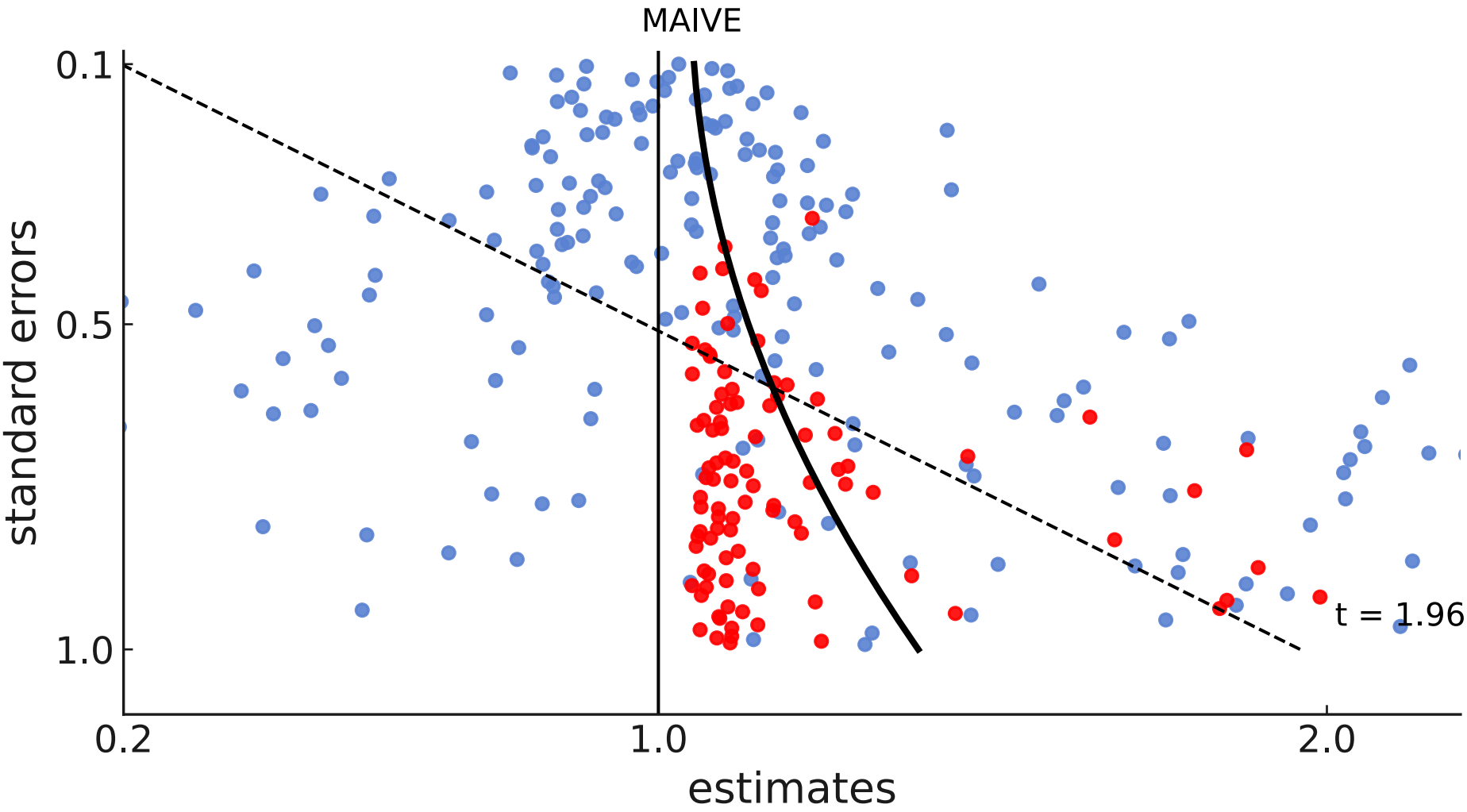
~~hacking~~ or ~~misspecifications~~

- Can add controls.
- Plug MAIVE-adjusted SEs into other estimators.
- In practice it's better to use logs in the first stage.

Aggressive *p*-hacking: PEESE in trouble



Meta-Analysis Instrumental Variable Estimator



MAIVE published last year

nature communications

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature communications › articles › article

Article | Open access | Published: 26 September 2025

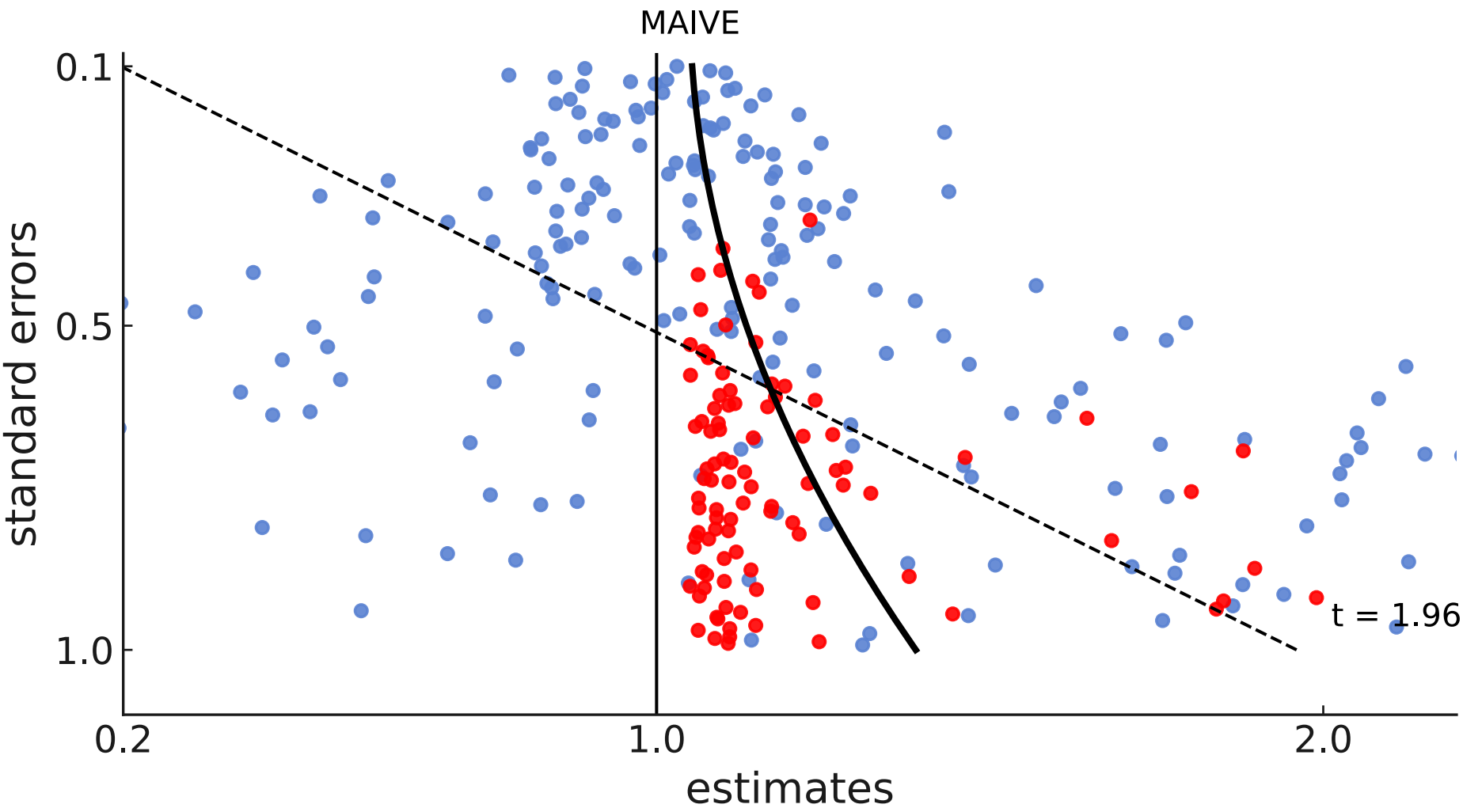
Spurious precision in meta-analysis of observational research

Zuzana Irsova ✉, Pedro R. D. Bom, Tomas Havranek & Heiko Rachinger

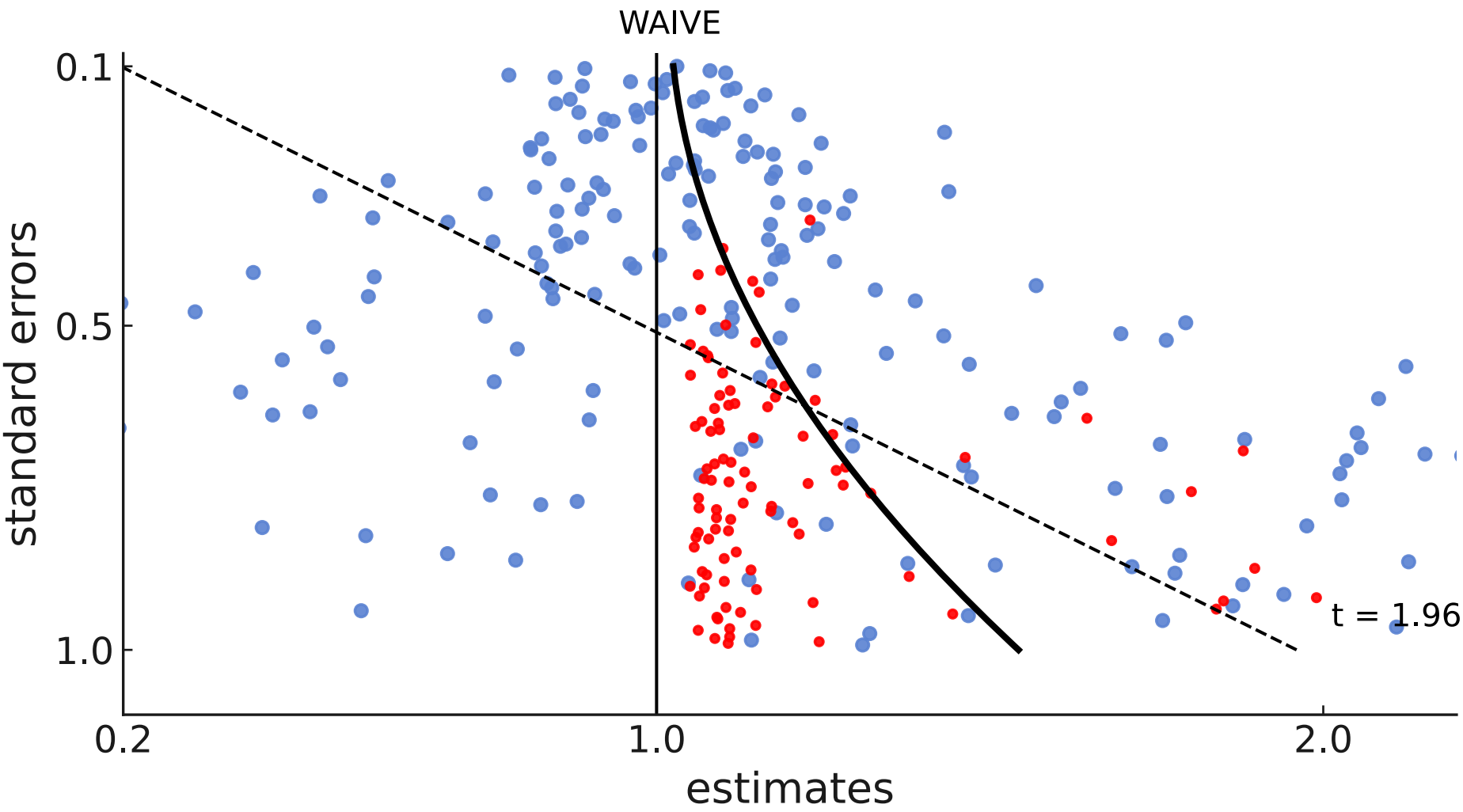
Nature Communications **16**, Article number: 8454 (2025) | Cite this article

1036 Accesses | **15** Altmetric | Metrics

Meta-Analysis Instrumental Variable Estimator



Weighted Adjustment Instrumental Variable Estimator



WAIVE first stage – same as MAIVE

Residual from MAIVE first stage:

$$SE(\hat{E})_i^2 = \alpha_0 + \alpha_1 (1/N_i) + \pi_i.$$

hacking if negative?

- Negative $\pi_i \rightarrow$ reported SE too small relative to N $\rightarrow$ possibly *p*-hacking.
- Analogy: profit shifting in accounting (regress profits on sales, look at residuals).

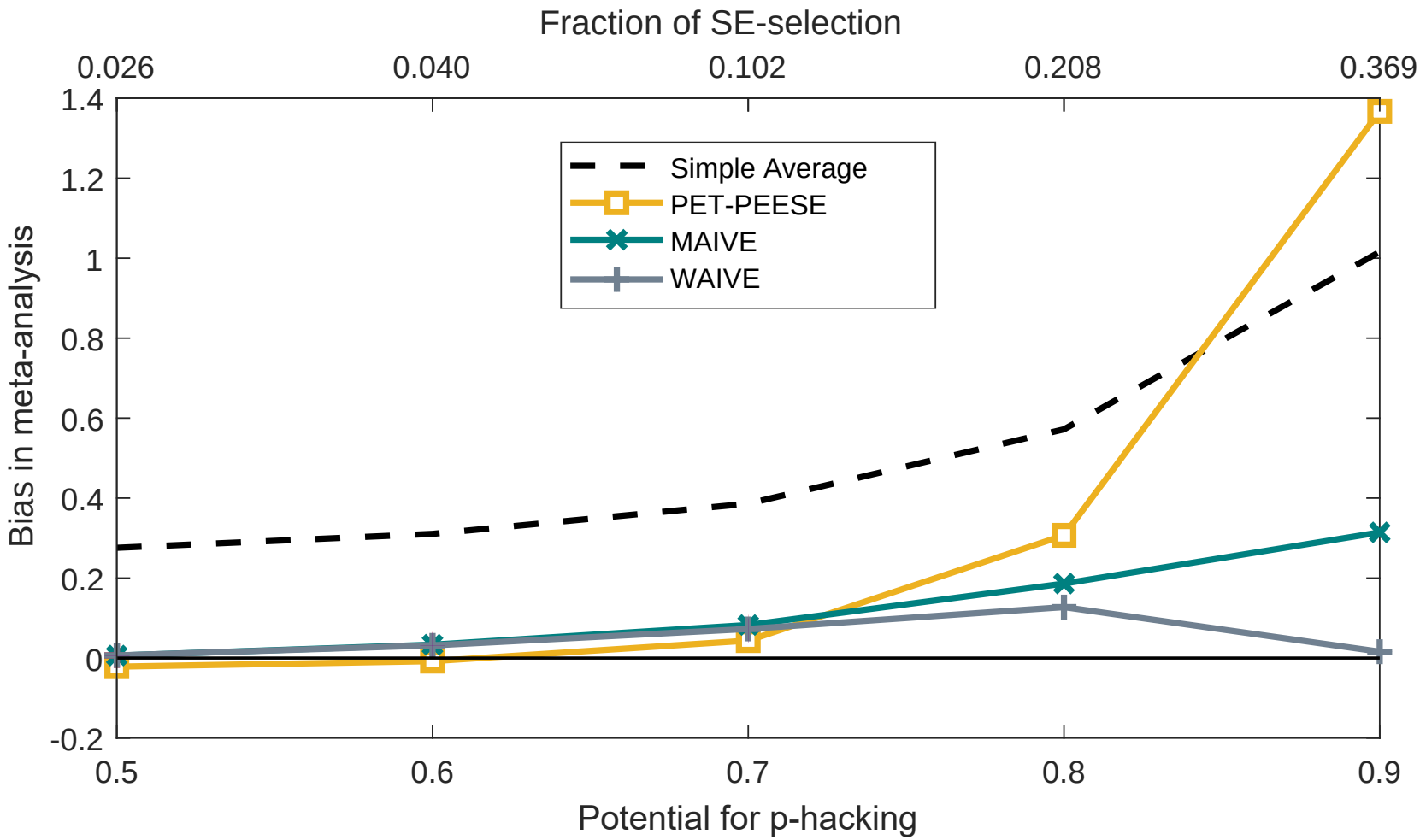
WAIVE weight for the second stage

Exponential tilt:

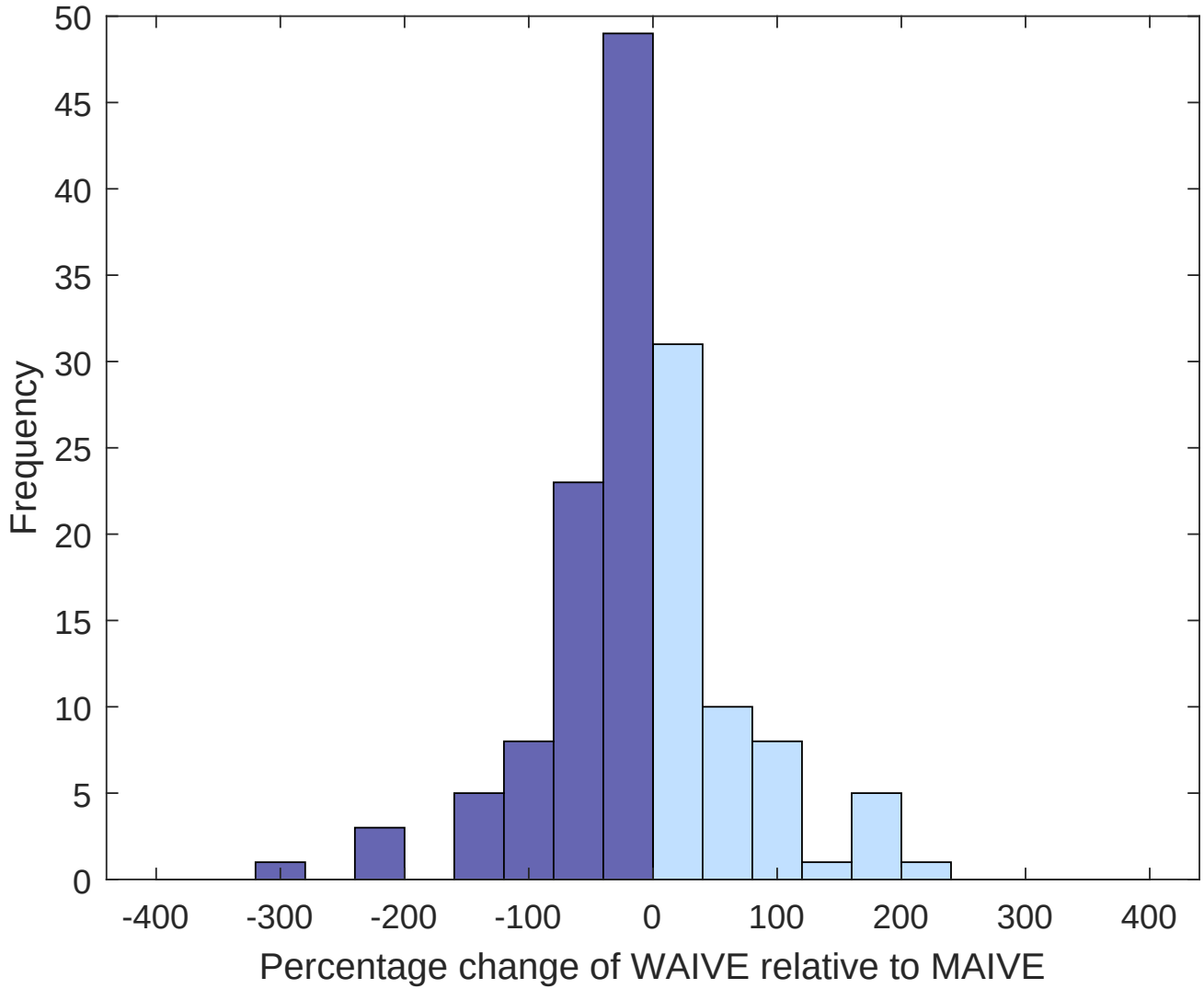
$$\omega_i \;=\; \exp\big[-\;\max(0,\,-\pi_i)\big]$$

- Penalize only negative residuals ($\pi_i < 0$) – downweight spuriously precise estimates.
- Weight $\in (0, 1]$, smooth and proportional decay.
- Then run PEESE of $\hat{E}_i$ on $SE(\hat{E})^2_{adj,i}$ with weights ω_i.

WAIVE alleviates the bias



WAIVE reduces MAIVE in 2/3 of meta-analyses



Practical issues

WAIVE still just an experimental concept.

Main consideration

- Can methods or *p*-hacking affect *SE*?
- Yes $\rightarrow$ MAIVE and WAIVE help (EasyMeta.org).
- No $\rightarrow$ MAIVE and WAIVE don't hurt much (but wider CI).

Project Website

meta-analysis.cz/maive

Papers using MAIVE

Irsova Z., P. Bom, T. Havranek, & H. Rachinger (2025): Spurious Precision in Meta-Analysis of Observational Research.
Nature Communications 16: 8454.

Opatrny M., T. Havranek, Z. Irsova, & M. Scasny (2026): Publication Bias and Model Uncertainty in Measuring the Effect of Class Size on Achievement.
Journal of Labor Economics, forthcoming.

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.

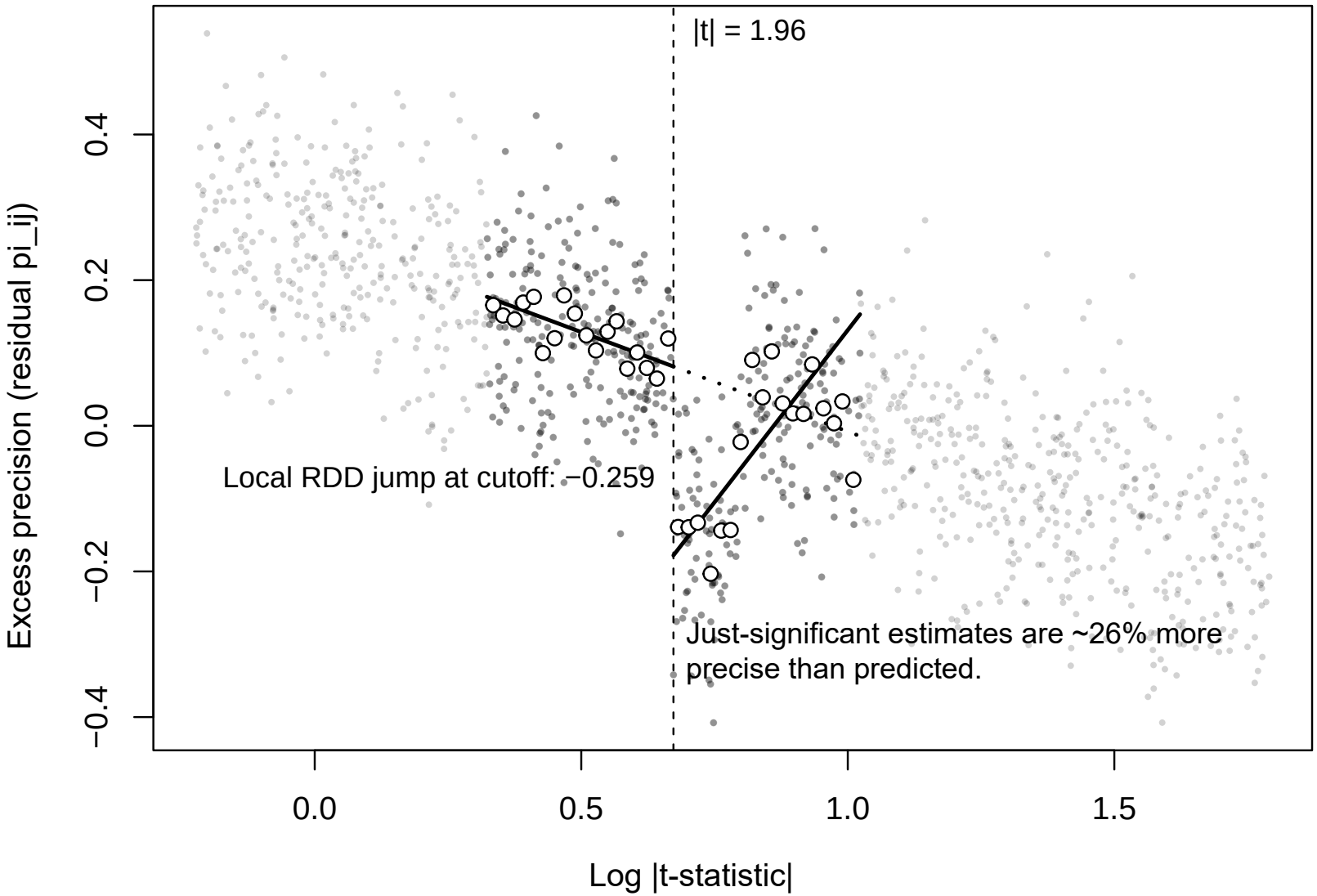
Diagnostic: excess precision

Residual from meta-regression:

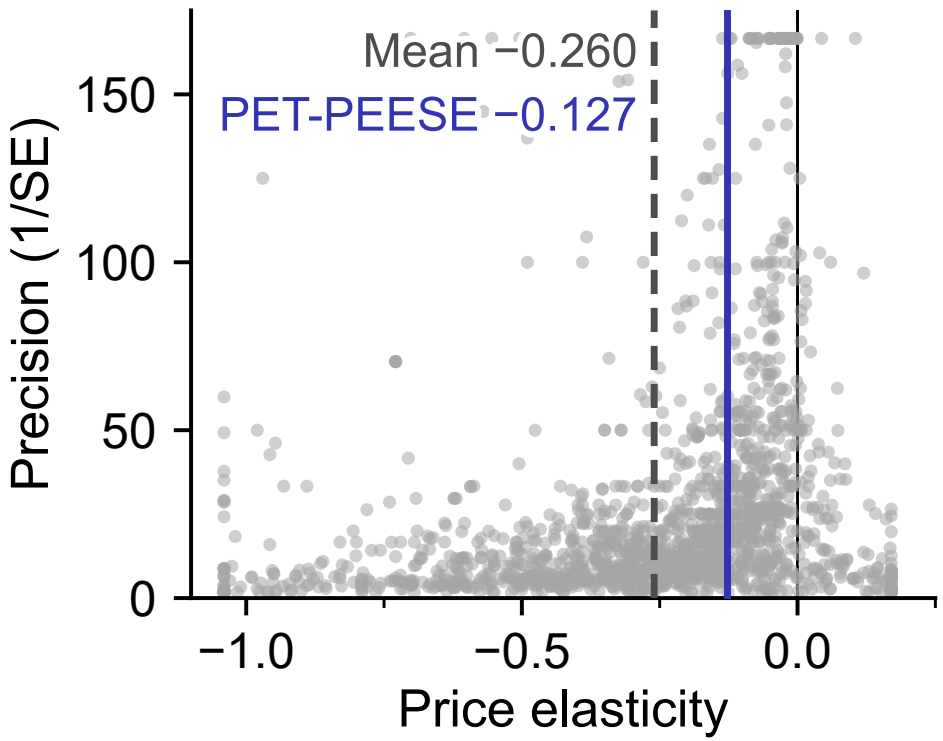
$$\log \mathrm{SE}(\hat{E}_i) = \alpha_0 + \alpha_1 \log N_i + \underbrace{\pi_i}_{<0\text{: excess precision}}$$

- 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.

RDT: residual discontinuity test (simulated)



Your turn: where does MAIVE land?



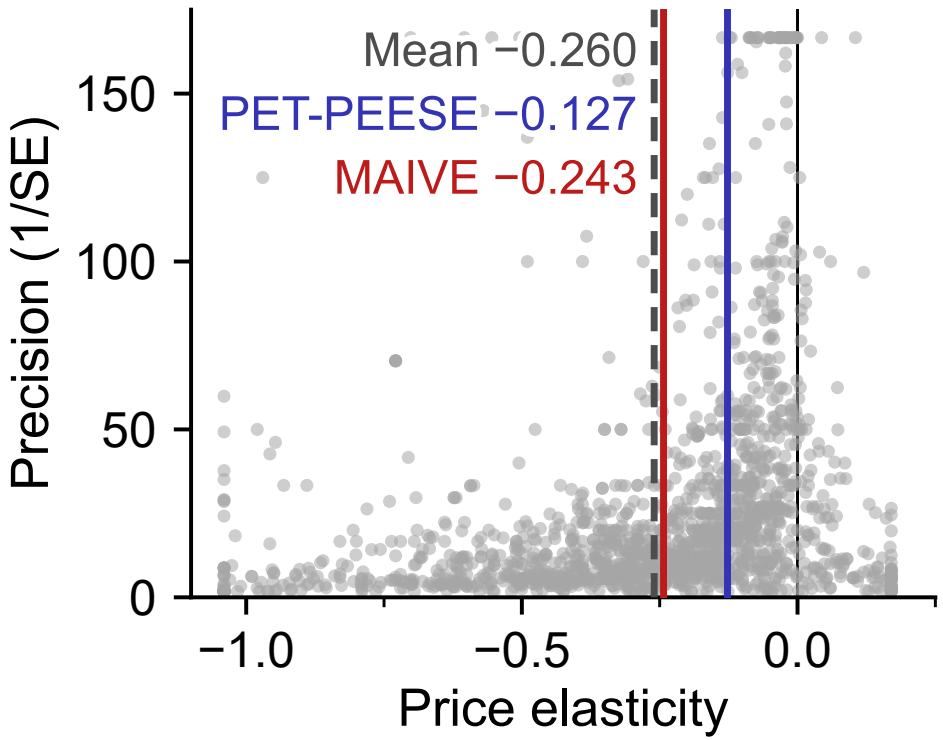
Mean −0.260
PET-PEESE −0.127

MAIVE: closer to zero, or further?

Try it

Start again at easymeta.org, not Back. Model MAIVE; Advanced Options: log first stage Yes, winsorize 2.5%; Run

MAIVE: −0.24



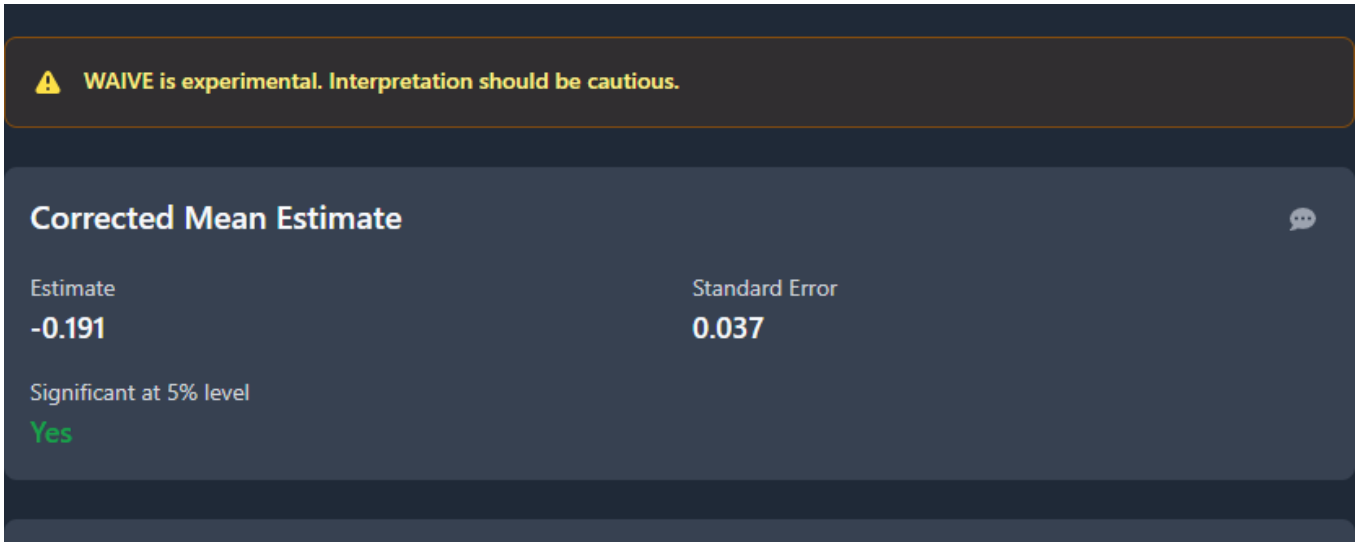
PET-PEESE	−0.127
equal weights	−0.226
MAIVE	−0.243

Log first stage, $F = 25$;
SE 0.038, clustered by study

42 estimates report SEs far below what their sample size predicts. They carry 37% of PET-PEESE's weight and are small in absolute value.

Levels first stage: $F = 6.9$, and the app warns.

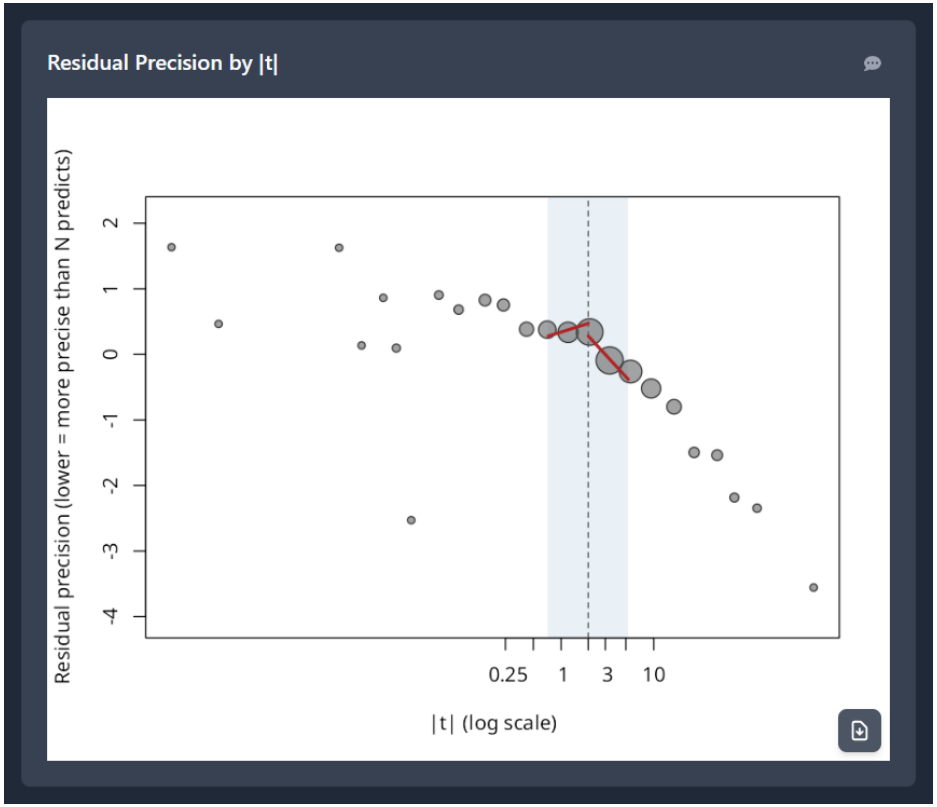
WAIVE, flagged



The app and Claude (through the API): -0.19. With the weights the right way round: -0.28.

Experimental. The package applies WAIVE's weights inverted; reported as MAIVE issue #30.

RDT in EasyMeta



Jump at $|t| = 1.96$: -0.19 ($p = 0.13$). A single literature gives this test little power.

App only: the API does not offer RDT.

MAER-Net guidelines, updated for AI

Cook et al. (2026a, 2026b), *Journal of Economic Surveys*. Thirteen authors, among them Bartos, Stanley, Gechert and Prante.

Human leadership

People direct every stage.

Accountability

AI is never a co-author.

Auditing

Humans recheck a random 10% of what AI screens and codes.

Disclosure

Any AI use that shapes the results is disclosed.

Recommendations; deviations should be disclosed. Free at meta-analysis.cz/guidelines.

New items in the reporting checklist

Stage	New in 2026
Search	AI use disclosed; AI search logs shared
Screening	random audit of 10% (at least 100 records); false-negative rate
Coding	AI may be a coder; humans recode 10% (at least 20 studies); kappa
Data	a dummy for human-reviewed estimates; share coded by AI
Figures	"Figure produced by AI." if not validated
Replication	tool, version, dates, prompts, settings

Cook et al. (2026b) replaces the 2020 guidelines for every meta-analysis, with or without AI.

Our open AI feedback tools

Dec 2025	**Duel**	ChatGPT against Gemini, copy and paste
Mar 2026	**MAD**	four chatbots debate, a fresh chat sums up
May 2026	**mad-research**	Claude Code and Codex run the debate
Jun 2026	**paper-workshop**	referee agents built for your paper

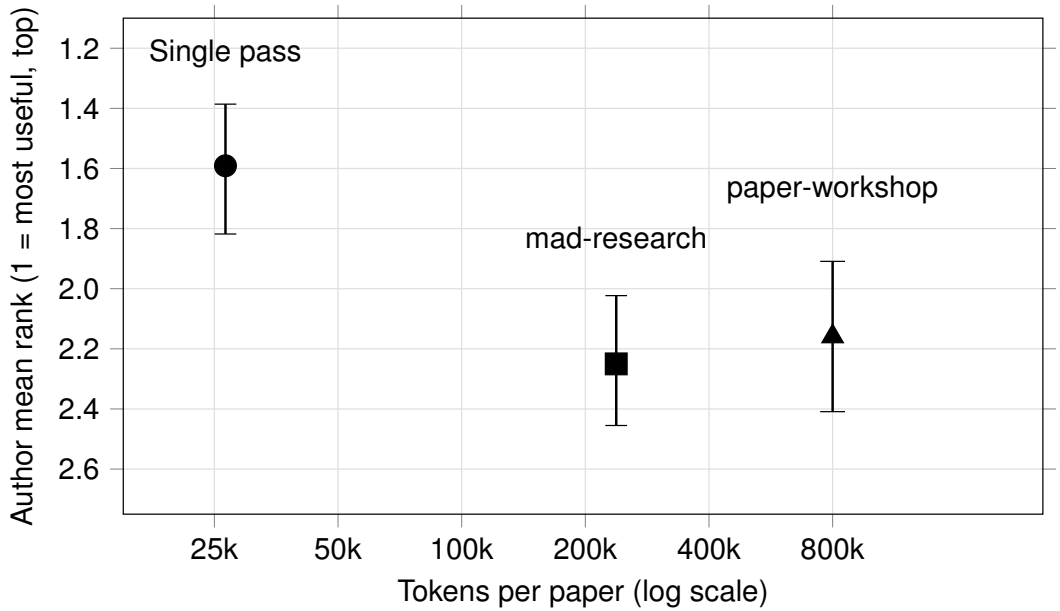
Free on GitHub, github.com/tjhavranek. Cook et al. (2026b) cite the Duel + MAD protocol for audits.

Does debate help? A pre-registered test

- 55 published or accepted meta-analyses in economics: 27 ours, 28 from JoES since 2022
- Three AI reports per paper: one prompt to one model, mad-research, paper-workshop
- Same template and length, tool names removed, random order
- Authors of 44 of them rank the reports by usefulness

Pre-registered on OSF on 22 June 2026, before any report was written.

Probably not, at least for meta-analyses



- Single pass ahead of mad-research on 32 of 44 papers, of paper-workshop on 30 (mean rank 1.59 vs 2.25 and 2.16)
- paper-workshop used about 30 times the tokens of the single pass
- An AI judge (Gemini) would have picked paper-workshop

Havranek and Irsova (2026), arXiv 2607.14713; CEPR DP 21752.

The single pass: this prompt, one call

79-word base prompt, given to all three arms

You are refereeing a meta-analysis submitted to an economics journal. Read the attached paper and write a referee report whose goal is to help the authors improve it. Be specific and constructive, and point to concrete parts of the paper.

Structure the report as:
1. Summary — one short paragraph: what the paper does and its main claim.
2. Major comments — the most important issues, most serious first.
3. Minor comments.
4. Overall assessment.

Keep it concise (about 1,000 words).

Verbatim from the replication package, doi 10.5281/zenodo.21273528.

Practical takeaway

Publication bias and *p*-hacking

Report several corrections side by side.

Reported precision

If studies report sample size, **MAIVE** instruments the standard error with it. **WAIVE** (experimental) goes further and downweights estimates still too precise for their sample size.

AI

Recheck a random 10% of what AI codes (at least 20 studies) and keep the prompts. For feedback, start with one good prompt.

Try it

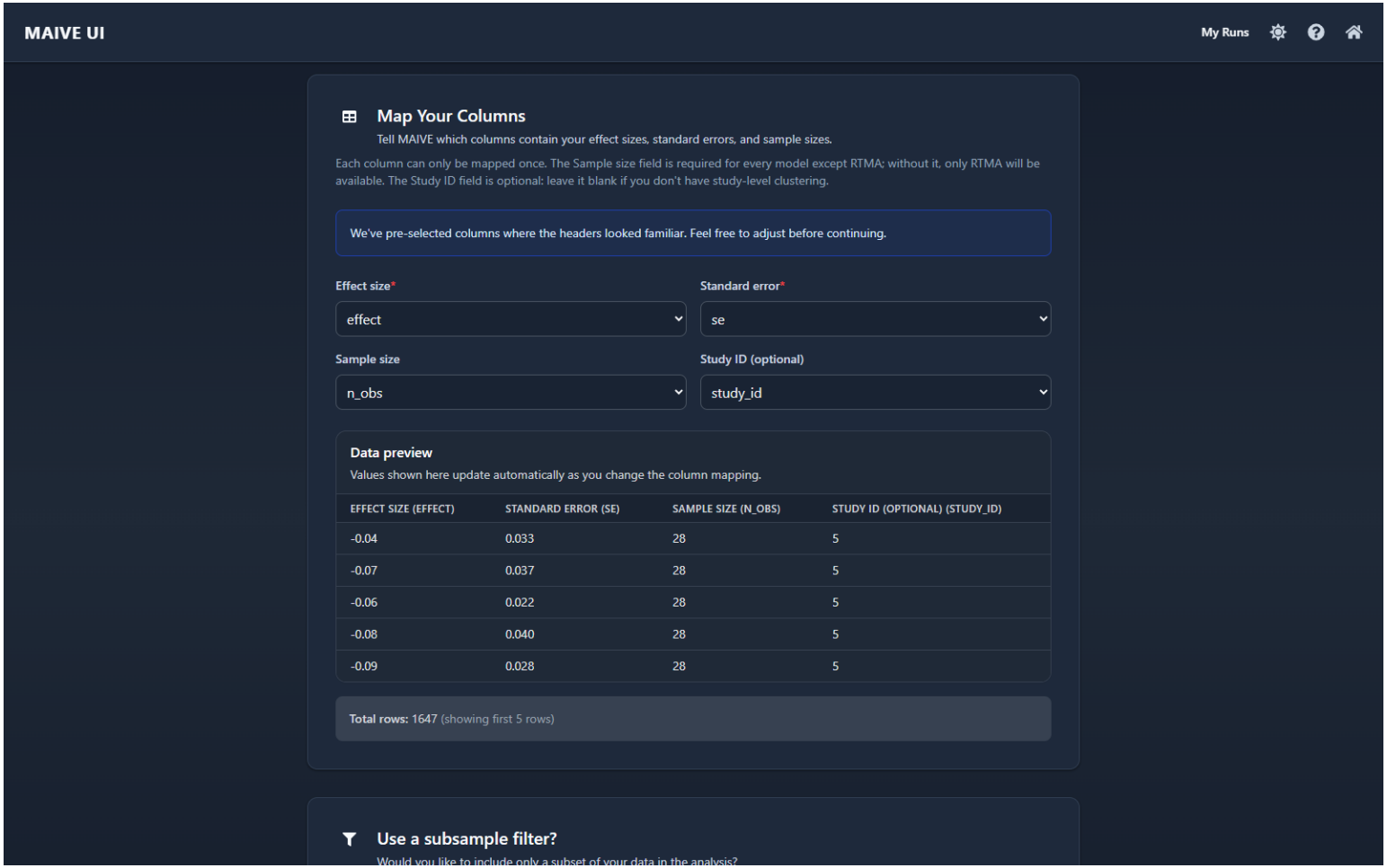
easymeta.org · meta-analysis.cz/chemnitz

Selected recent papers

- Irsova Z., P. Bom, T. Havranek, & H. Rachinger (2025): Spurious Precision in Meta-Analysis of Observational Research.
  Nature Communications 16: 8454.
- Havranek T., & Z. Irsova (2026): Does Multi-Agent Debate Improve AI Feedback on Research Papers?
  arXiv 2607.14713.
- 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.

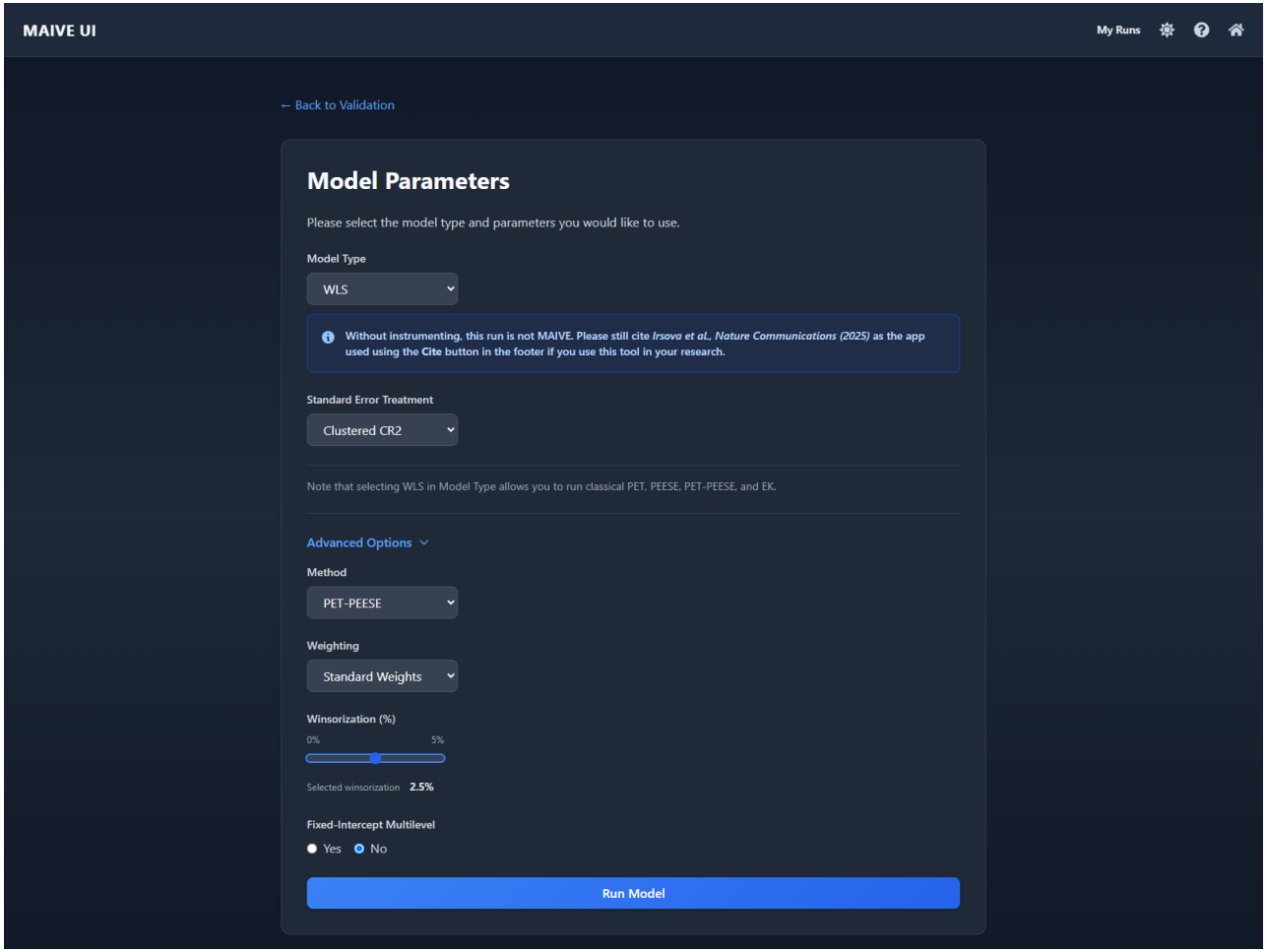
Claude Opus 5, Claude Fable 5.1 and GPT-6 Astra were used in preparing this demo and presentation.

Backup: upload and column mapping



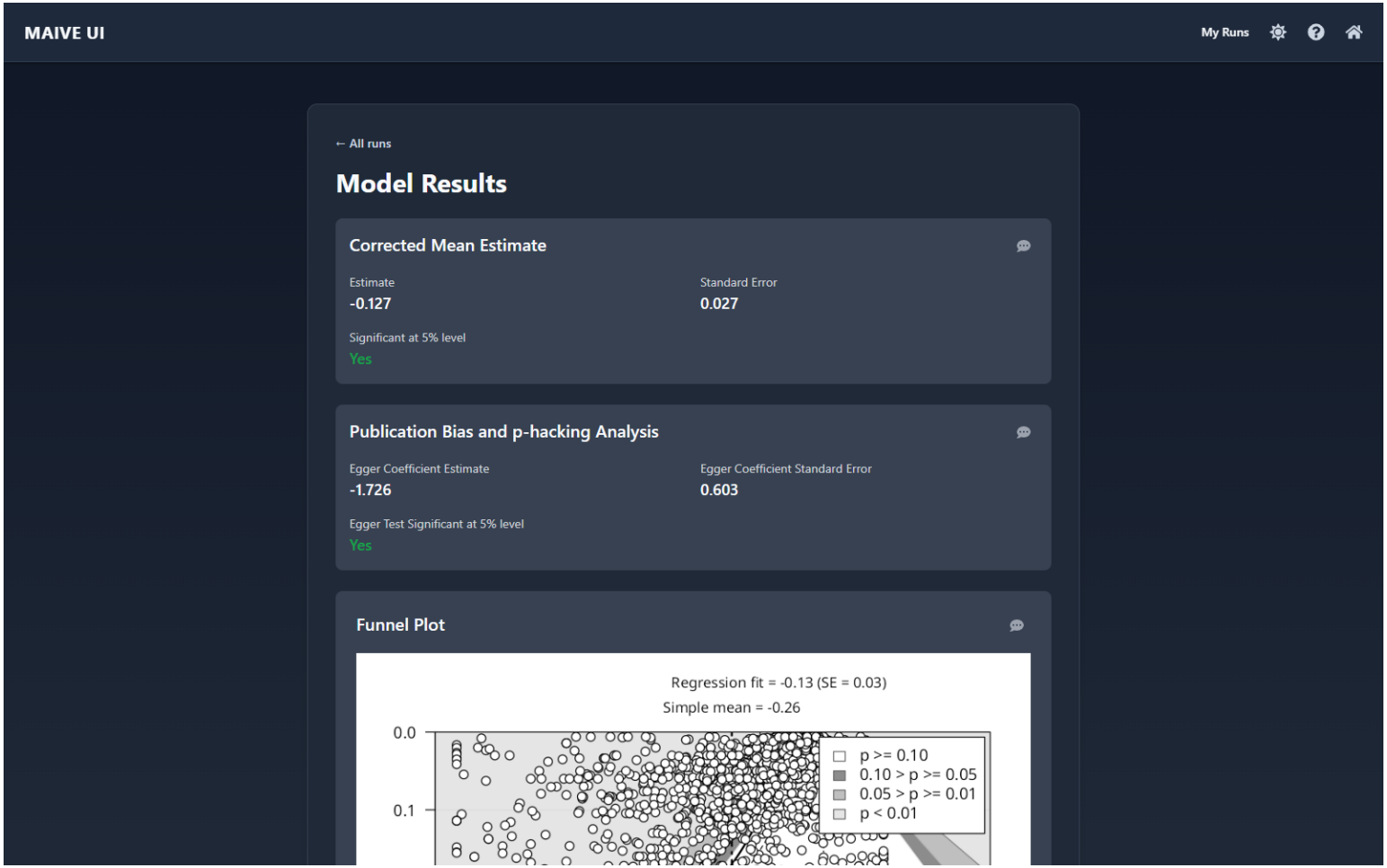
All four columns map automatically.

Backup: WLS, PET-PEESE, winsorize 2.5%



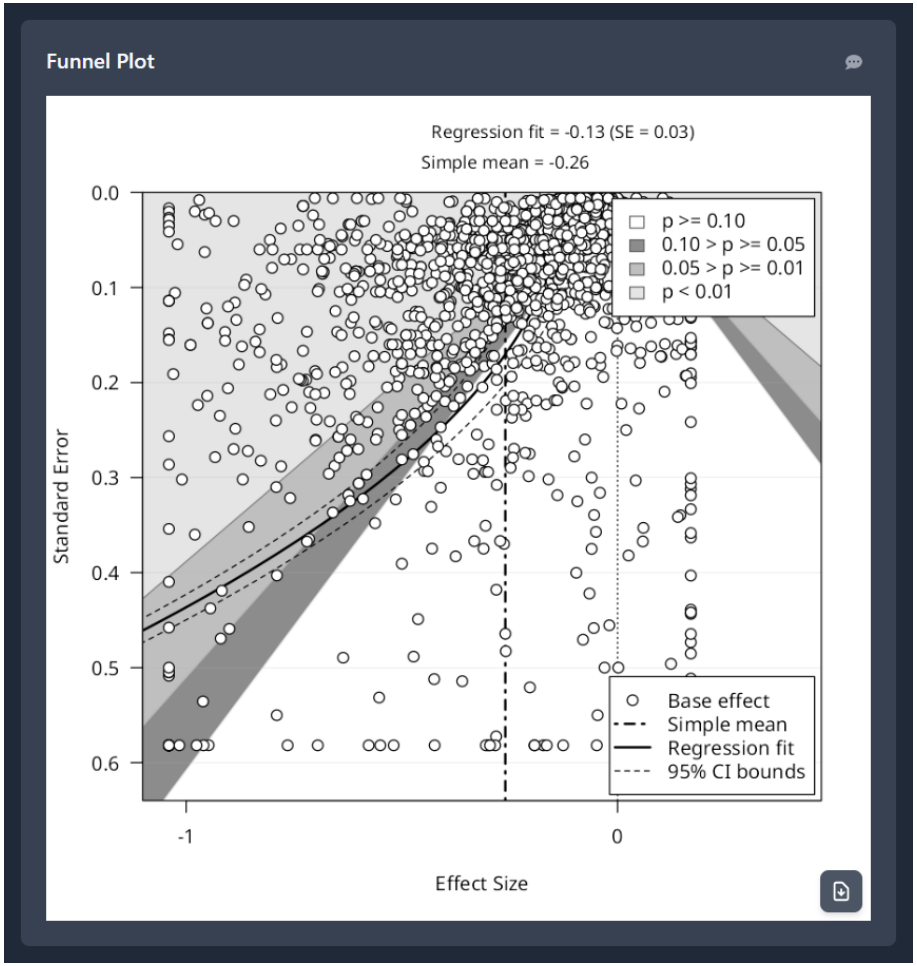
Settings for the first run.

Backup: PET-PEESE



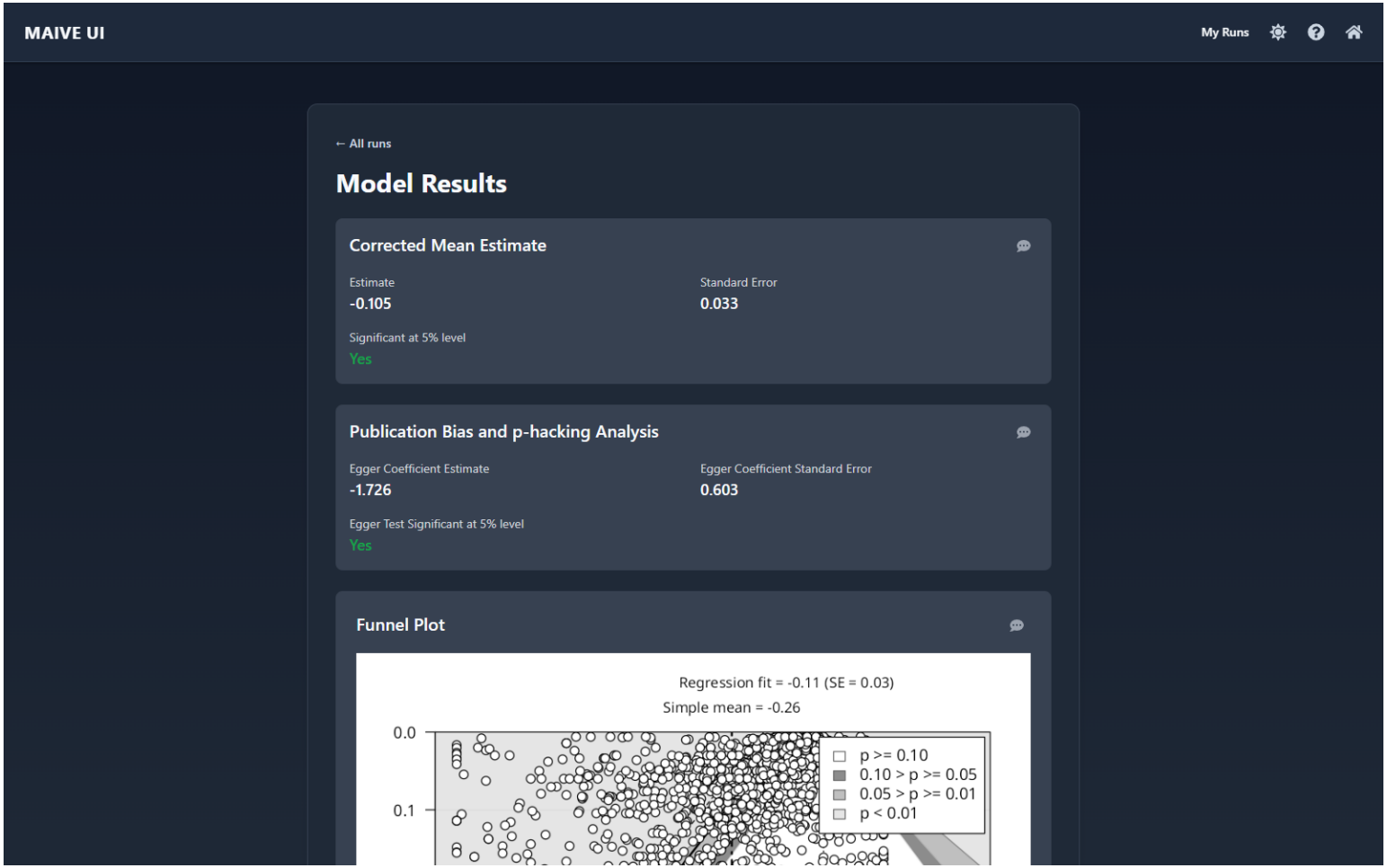
PET-PEESE picks PEESE: −0.127.

Backup: funnel plot in the app



Same estimates as on the demo slide; the app puts SE on the vertical axis, 0 at the top.

Backup: EK

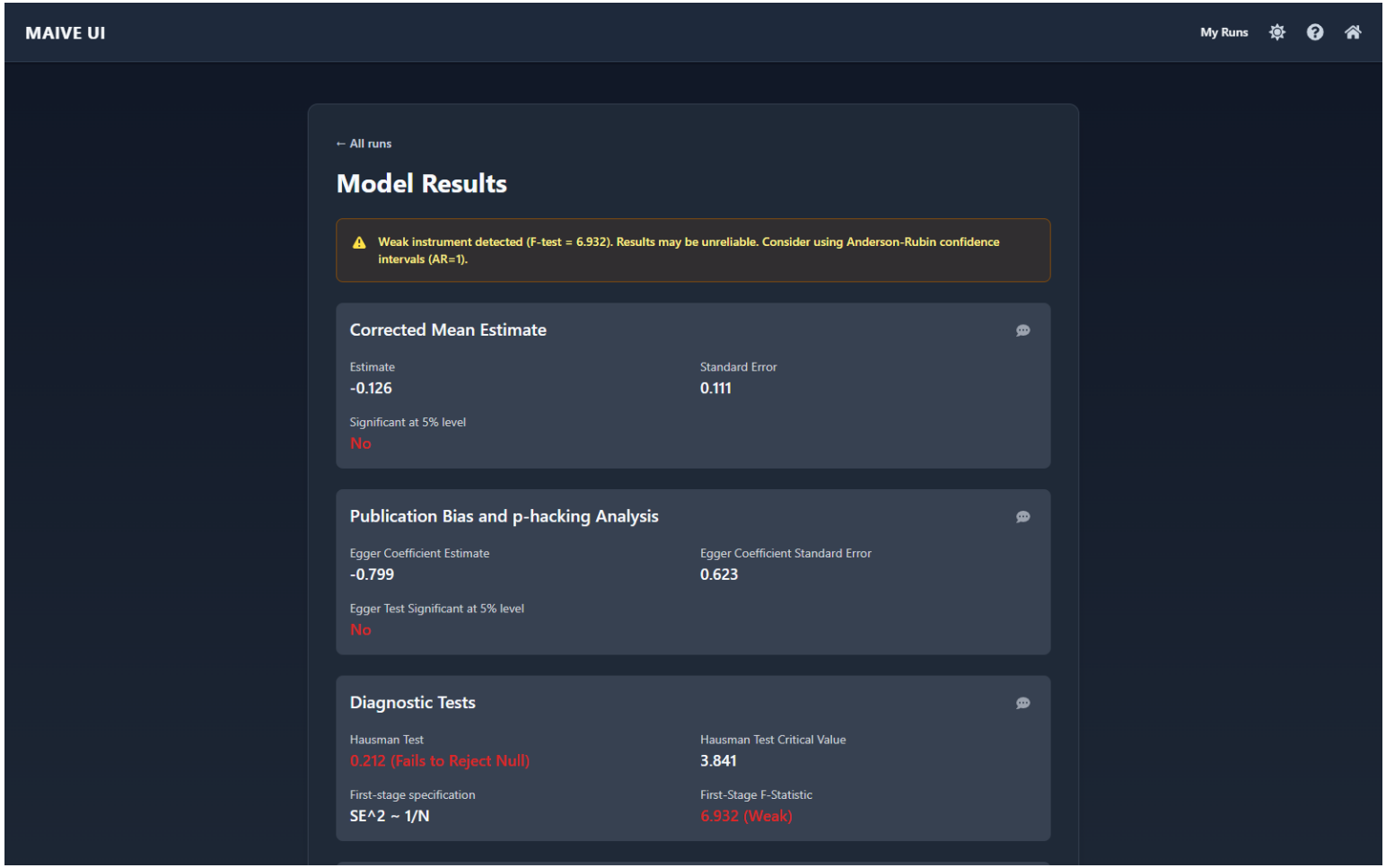


EK finds no kink, so it equals PET: -0.105.

Backup: what the API returned

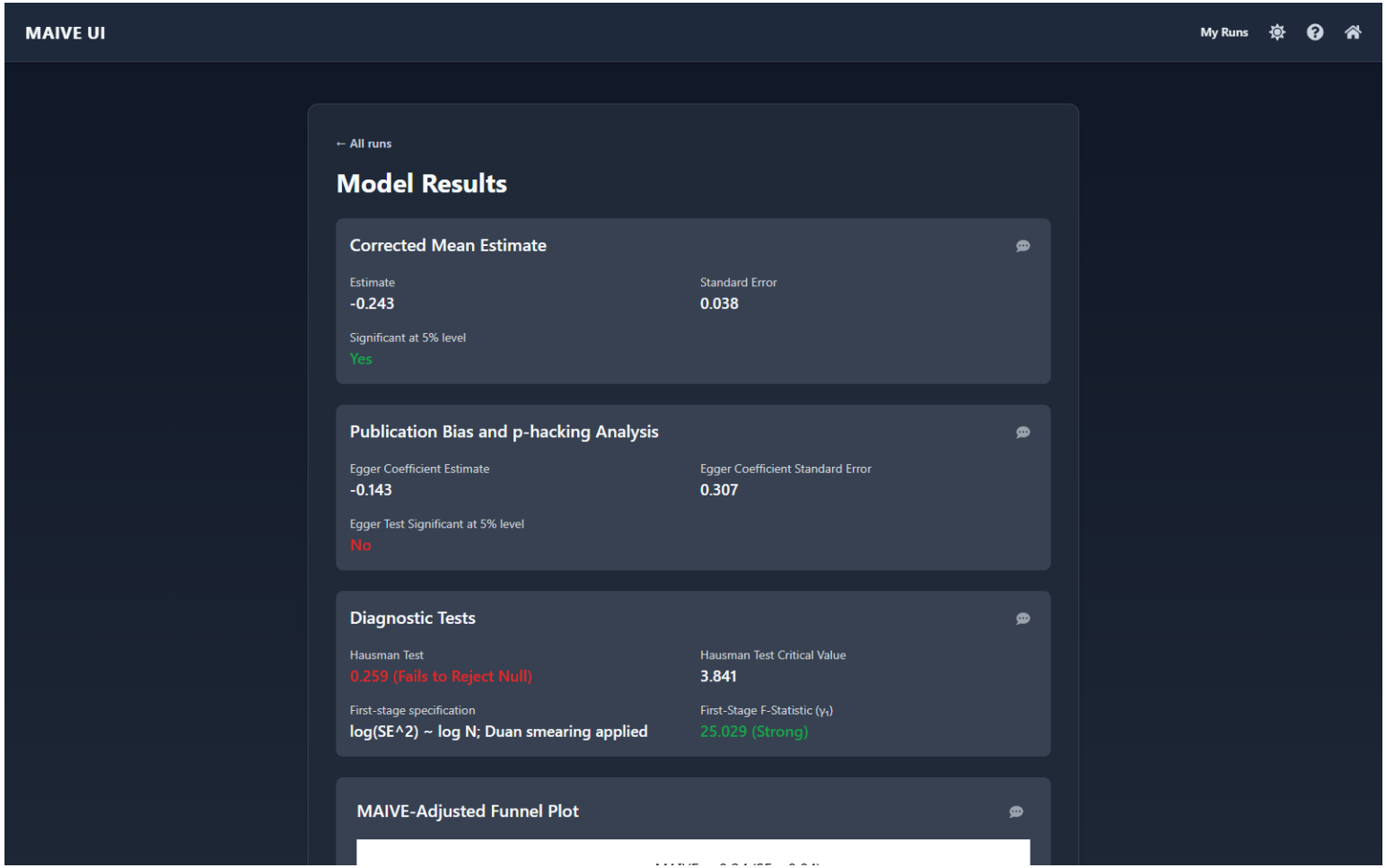
```
POST https://api.maive.eu/v1/run-model
1,647 rows, winsorize 2.5

recipe PET-PEESE    HTTP 200
  effectEstimate    -0.127104860252972
  standardError      0.026919174216364 4      (PEESE selected)

recipe EK           HTTP 200
  effectEstimate    -0.10525106160908
  standardError      0.032703393098768 9

MAIVE, log first stage    HTTP 200
  effectEstimate    -0.24255281691904
  standardError      0.037775202532599
  firstStageFStatistic    25.03

WAIVE, same settings    HTTP 200
  effectEstimate    -0.190975882295736
```

Backup: MAIVE, levels first stage



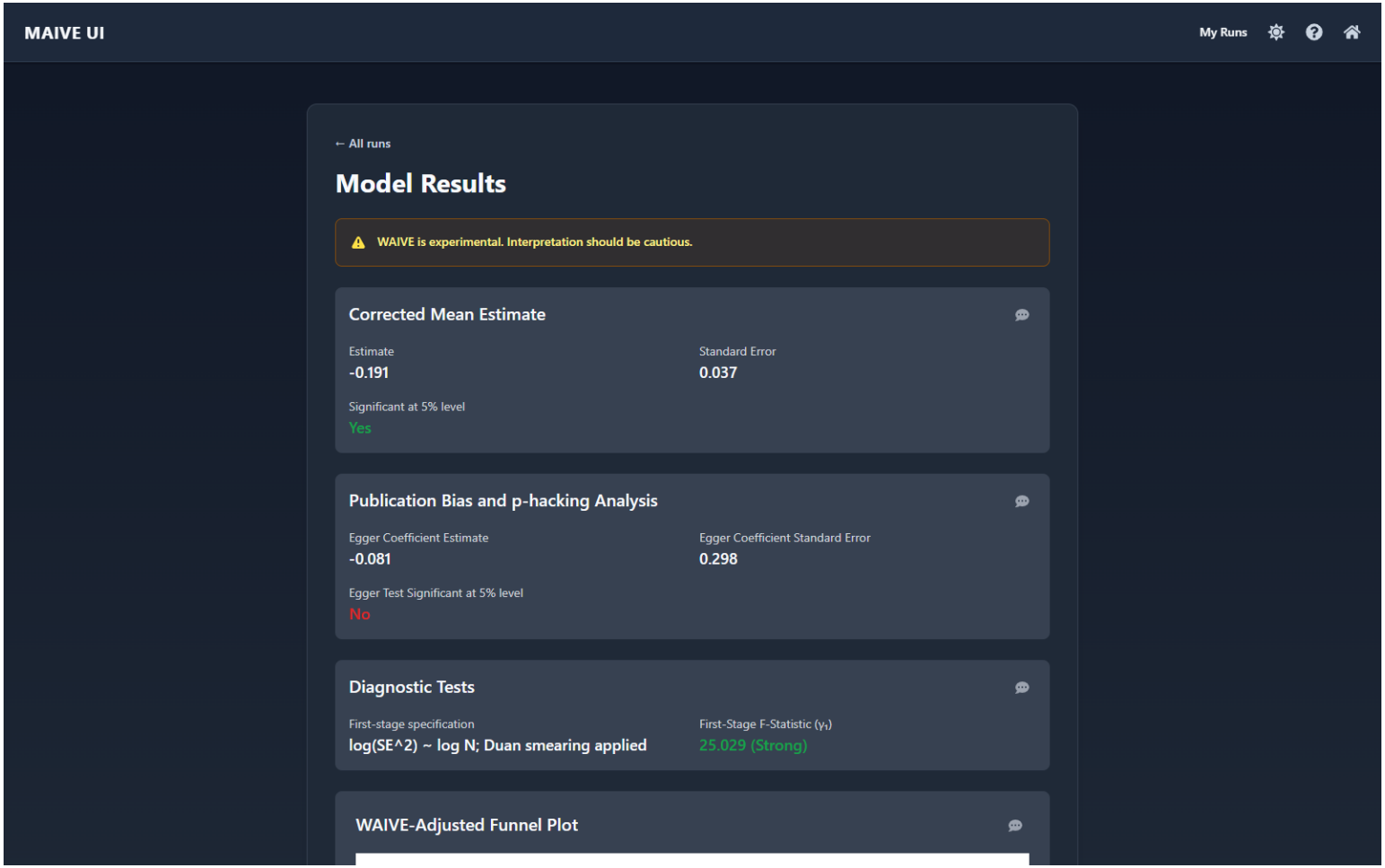
Levels first stage: MAIVE -0.126, $F = 6.9$, and the app warns.

Backup: MAIVE, log first stage



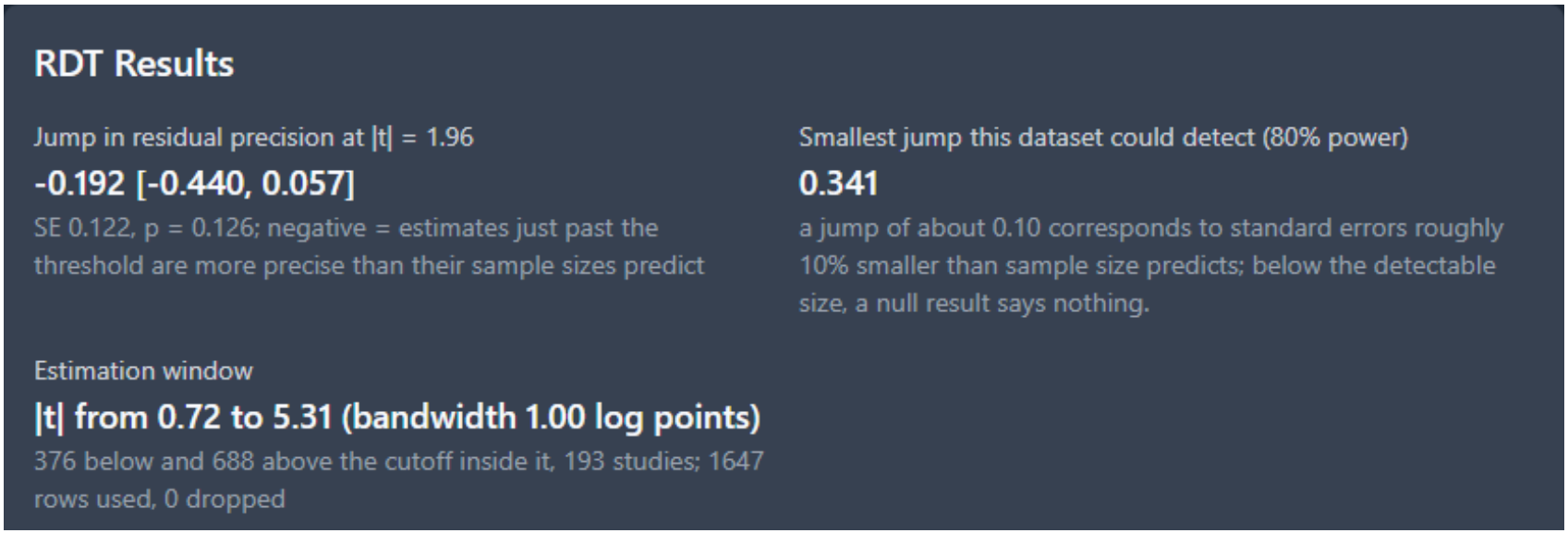
Log first stage: $F = 25$, MAIVE -0.24.

Backup: WAIVE



WAIVE -0.19; experimental, weights inverted (MAIVE issue #30).

Backup: RDT



Jump -0.19 ($p = 0.13$): no clear jump at $|t| = 1.96$.