

Meta-Analysis in Economics

Practical Tools for Synthesizing Research

Tomas Havranek

Charles University, Prague
Centre for Economic Policy Research, London
Meta-Research Innovation Center, Stanford

Workshop prepared for the University of Osaka

March 16, 2026

Workshop Outline

- 1 Motivation
- 2 Literature Search
- 3 Data Collection
- 4 Conventional Tools
- 5 Publication Bias
- 6 P-hacking
- 7 Heterogeneity

Workshop Outline

- 1 Motivation
- 2 Literature Search
- 3 Data Collection
- 4 Conventional Tools
- 5 Publication Bias
- 6 P-hacking
- 7 Heterogeneity

Meta important for research, policy, practice

Research experience:

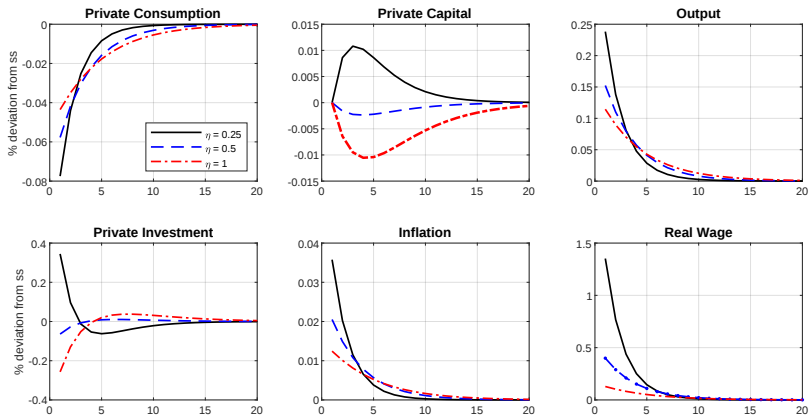
- Affiliate at the Meta-Research Innovation Center, Stanford
- Panel member at ERC Advanced
- Convener of Meta-Analysis in Economics Network
- Meta-Analysis Editor at the Journal of Economic Surveys
- MAIVE method (Nature Communications)
- Over 50 metas published, e.g. REStat, JEEA, JOLE, JIE. . .

Policy experience:

- Former Advisor to the Board, Czech National Bank
- Former member of the Czech National Economic Council
- Affiliate at the Centre for Economic Policy Research, London

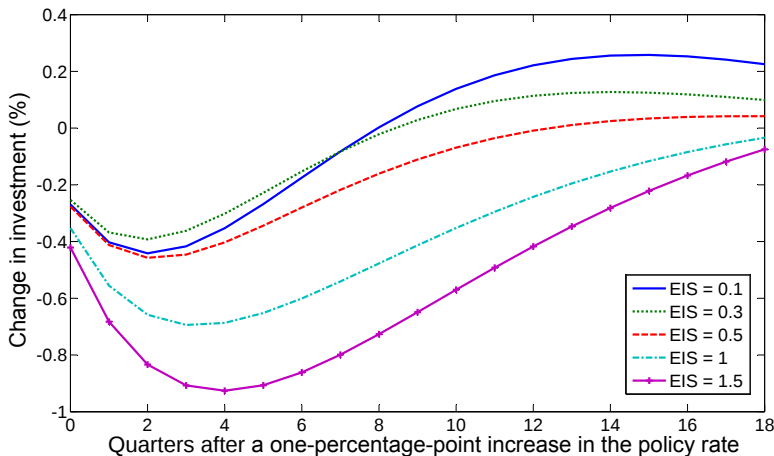
What happens if the government spends more?

Labor supply elasticity is key (but how to calibrate it?):



Details: Elminejad et al. (2023, Review of Economic Dynamics)

What happens if the central bank raises the rate?

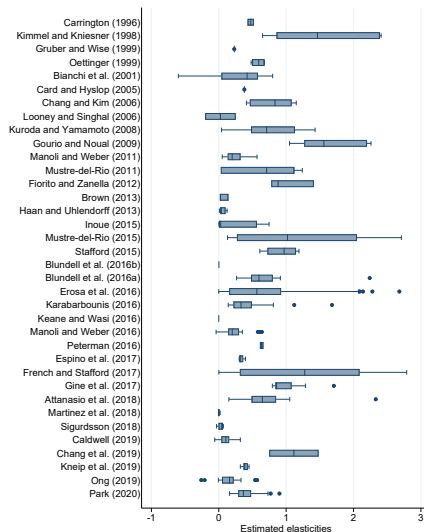


Details: Havranek et al. (2015, Journal of International Econ)

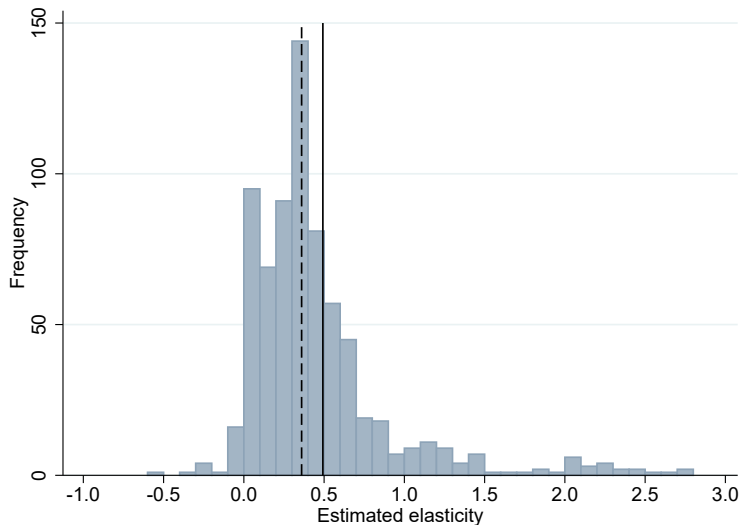
Calibration uncertainty: meta needed!

Parameter		Prior Mean		
Preferences			Calvo wages	θ_w 0.380
Habits	h	0.900	Index. dom. prices	χ_p 0.750
Labour supply coef.	ψ	8.832	Index. imp. prices	χ_M 0.500
Frisch elasticity	ϑ	1.250	Index. exp. prices	χ_x 0.350
Wedges			Index. wages	χ_w 0.920
Euler	κ^{euler}	1.0119	Monetary policy	
Forex	κ^{forex}	1.0034	Taylor rule (int. rates)	γ_R 0.960
Adjustment costs			Taylor rule (output gap)	γ_y 0.220
Investment	κ	20.000	Taylor rule (inflation)	γ_Π 1.150
Capital utilization	γ_2	0.280	Fiscal policy	
Risk premium	Γ_{bw}	0.800	Public consumption	ρ_g 0.750
Elasticities of substitution			Home bias	
Domestic goods	ϵ	5.000	Home bias in consump.	n_c 0.280
Import goods	ϵ_M	9.000	Home bias in invest.	n_i 0.120
Export goods	ϵ_x	9.400	Home bias in export	n_x 0.350
World goods	ϵ_W	0.860	Growth rates	
Consumption goods	ϵ_c	7.600	Invest. spec. tech.	Λ_μ 1.000
Investment goods	ϵ_i	7.600	General tech.	Λ_A 1.009
Labour types	η	7.000	Population	Λ_L 1.000
Price and wage setting			ER appreciation	$\dot{e}_x$ 0.994
Calvo dom. prices	θ_p	0.500	Openness tech.	α_O 1.0035
Calvo exp. prices	θ_x	0.080	Export spec. tech.	α_X 1.0058
Calvo imp. prices	θ_M	0.750		

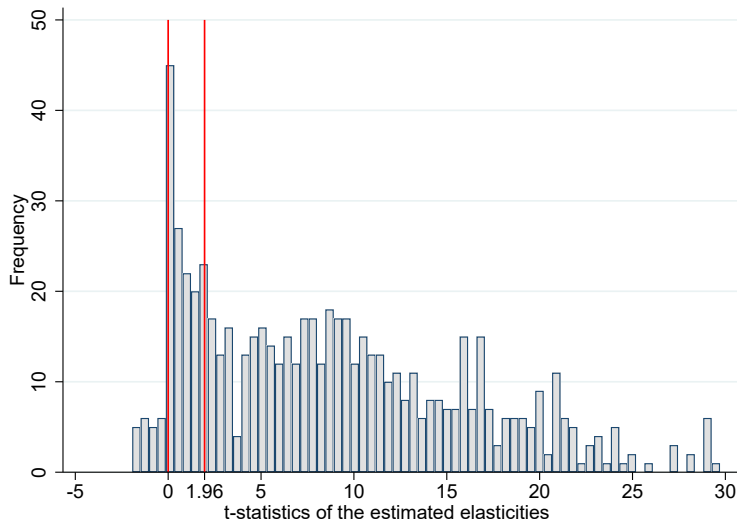
Estimates of the labor supply elasticity vary



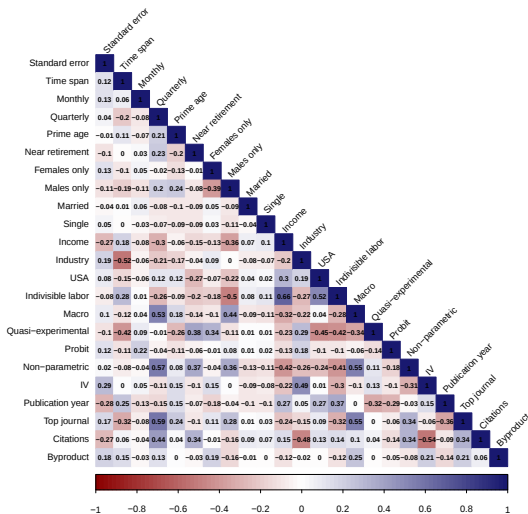
Values around 0.4–0.5 used for calibration (CBO)



Bias: some estimates less likely to be reported



Context: different situations yield different estimates



Implied elasticity: taking bias and context into account

Let's construct a hypothetical study that uses all information reported in the literature but puts more weight on selected aspects (high precision, credible identification, etc.)

	Subjective best practice	Blundell <i>et al.</i> (2016a)
Overall	0.02 (−0.22, 0.25)	−0.03 (−0.25, 0.18)
Near retirement	0.14 (−0.14, 0.42)	0.09 (−0.14, 0.31)
Women	0.12 (−0.08, 0.32)	0.07 (−0.10, 0.24)



ARTICLE | Open Access



Meta-analysis of social science research: A practitioner's guide

Zuzana Irsova, Hristos Doucouliagos, Tomas Havranek, T. D. Stanley

First published: 23 November 2023 | <https://doi.org/10.1111/joes.12595>

SECTIONS



PDF



TOOLS



SHARE

Abstract

This article provides concise, nontechnical, step-by-step guidelines on how to conduct a modern meta-analysis, especially in social sciences. We treat publication bias, p -hacking, and systematic heterogeneity as phenomena meta-analysts must always confront. To this end, we provide concrete methodological recommendations. Meta-analysis methods have advanced substantially over the last four decades. Yet many meta-analyses still rely on

Guiding example: effect of beauty on success



Open Stata and R

Main Findings

- 1 Mean of 1,159 estimates from 67 studies: good looks → earnings ↑ 5%
- 2 Publication bias. After correction: 3%
- 3 Control for ability? Productivity? Beauty premium $\approx$ 1%

Project Website

www.meta-analysis.cz/beauty

Revisions requested at *Nature Human Behaviour*.

Workshop Outline

- 1 Motivation
- 2 Literature Search**
- 3 Data Collection
- 4 Conventional Tools
- 5 Publication Bias
- 6 P-hacking
- 7 Heterogeneity

meta-analysis.cz » resources for research synthesis

INTERTEMPORAL SUBSTITUTION

TRADE ELASTICITY

CAPITAL-LABOR SUBSTITUTION

CLASS SIZE

SKILL SUBSTITUTION

data and codes for meta-analysis

This platform serves as a digital repository for meta-analysis methods and applications co-authored by researchers at the Institute of Economic Studies, Charles University, Prague. Meta-analysis, the quantitative approach to research synthesis, was originally developed in experimental medicine. In economics, meta-analysis proves useful for calibrating structural models, correcting publication bias, and tracing differences in the results of primary studies to their contextual backgrounds. The papers presented here span numerous fields and have been published in journals such as the *Review of Economics and Statistics*, *Journal of the European Economic Association*, and *Journal of International Economics*.

A new meta-analysis approach robust to publication bias, *p*-hacking, and spurious precision:

[Meta-Analysis Instrumental Variable Estimator \(MAIVE\)](#)

Concise, nontechnical, step-by-step guidelines on how to do a meta-analysis:

[The Practitioner's Guide to Modern Meta-Analysis](#)

search

featured articles

[Skill substitution](#)

[Intertemporal substitution](#)

[House prices](#)

[Excess sensitivity](#)

[Habits in consumption](#)

[Border effect](#)

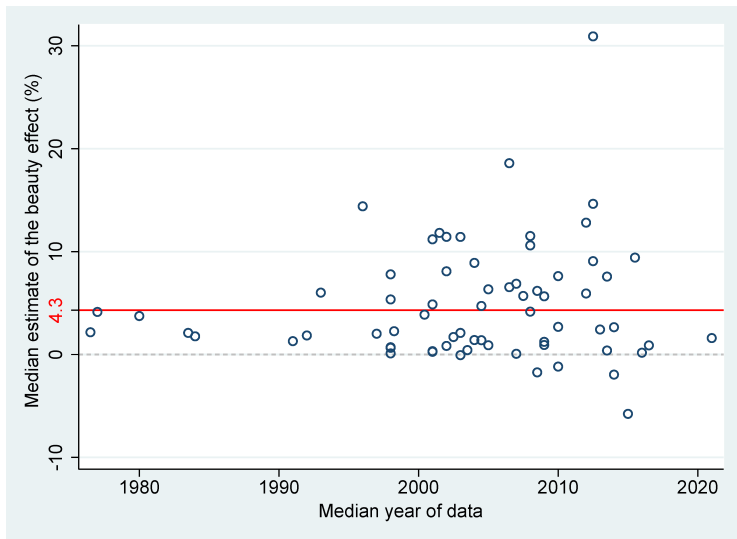
[Trade elasticity](#)

[Capital-labor substitution](#)

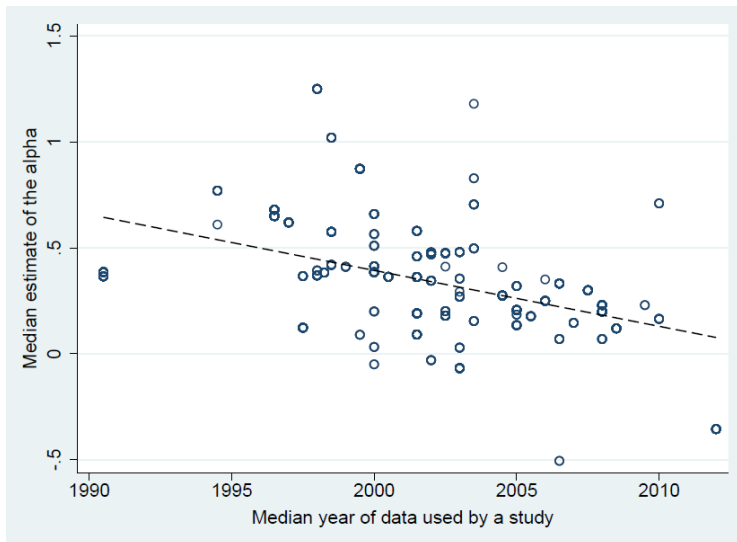
[Labor supply](#)

Individual discount rate

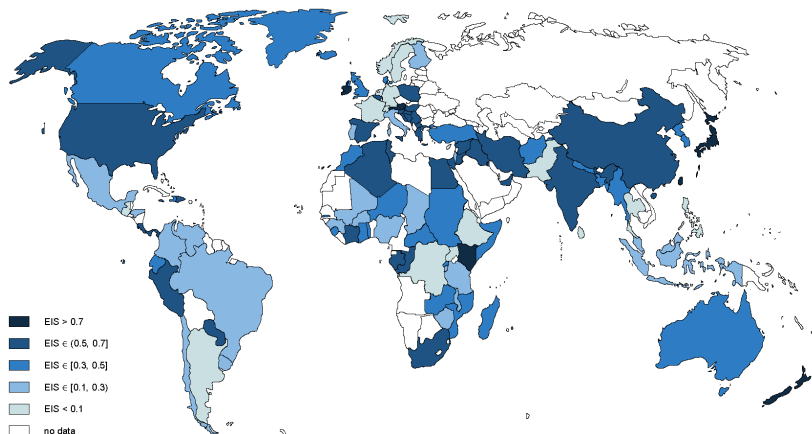
Motivation example: increasing uncertainty



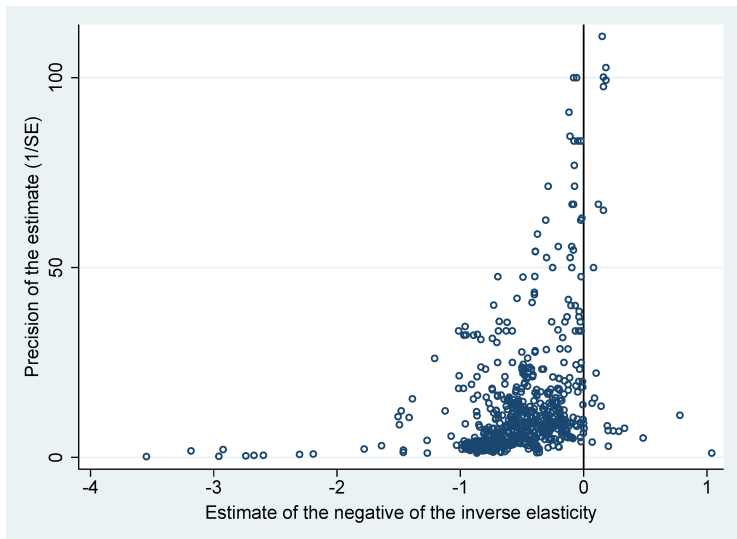
Motivation example: clear trend



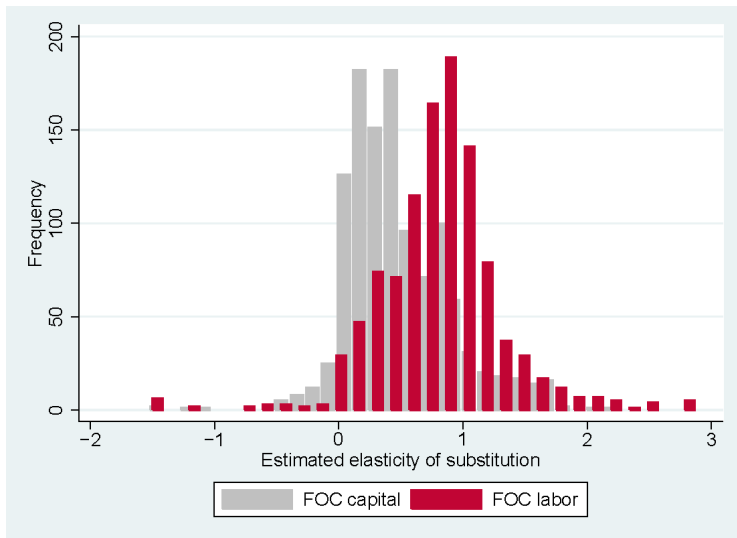
Motivation example: country-level differences



Motivation example: publication bias



Motivation example: method heterogeneity



We use Google Scholar

Advantages:

- Full text, not just the abstract, title, keywords
- Inclusive
- Up-to-date

Disadvantages:

- Algorithm changes in time (replicability issues)
- Too many results
- Sometimes relevance unclear

Search steps (without AI for now)

- 1 Find 5 most relevant studies
- 2 Design a query that shows the 5 studies near the top
- 3 Go through first 500 hits
- 4 Read abstracts
- 5 Download potentially useful studies
- 6 Skim them
- 7 Repeat the procedure for the most recent studies (last 3 years)

Search query 2

"capital" ("labor" OR "labour") "substitution" "elasticity" (CES OR "constant e

Search

Scholar

About 3,730 results (0.07 sec)

Articles

Case law

My library

Any time

Since 2015

Since 2014

Since 2011

Custom range...

Sort by relevance

Sort by date

☐ include patents

☐ include citations

Create alert

Is the US aggregate **production function Cobb-Douglas**? New **estimates** of the **elasticity of substitution** [PDF] from harvard.edu

P Antras - Contributions in Macroeconomics, 2004 - degruyter.com

... 4 **Estimates** under Hicks Neutral Technological Change In this section, I present **estimates of the elasticity of substitution** between **capital** and **labor** based on both classical **regression** analysis and modern time series analysis. ...

Cited by 323 Related articles All 14 versions Cite Save

[PDF] **Capital-labor substitution** and economic efficiency

[PDF] from jstor.org

KJ Arrow, HB Chenery, BS Minhas, RM Solow - The Review of Economics ..., 1961 - JSTOR

... 2) . Geometrically these conditions state that output per unit of **labor** is an increasing function of the input of **capital** per unit of **labor**, convex from ... value of A. This way of generating **production functions** brings to light a connection with the **elasticity of substitution** which does ...

Cited by 2468 Related articles All 14 versions Cite Save

An estimation of US industry-level **capital-labor substitution elasticities**: support for **Cobb-Douglas**

[PDF] from antoniomariacosta.com

EJ Ballistreri, CA McDaniel, EV Wong - The North American Journal of ..., 2003 - Elsevier

... The **elasticity of substitution** between **capital** and **labor** is represented by σ , the coefficient of interest. ... We use the weighted-symmetric test to **determine** the **order** of integration for each series across industries, the **ratio of capital to labor** inputs, and the corresponding relative ...

Cited by 73 Related articles All 17 versions Cite Save

[PDF] **Substitution** among **capital**, **labor**, and natural resource products in American **manufacturing**

[PDF] from jstor.org

DB Humphrey, JR Moroney - The Journal of Political Economy, 1975 - JSTOR

... Allen (1938, p. 504) defined the partial **elasticity of substitution** between inputs i and j by ... LYKKJ The disturbances UL and UK are likely to be correlated, because random deviations from profit maximization should affect both the markets for **labor** and **capital**. ...

Cited by 205 Related articles All 4 versions Cite Save

Primary studies

A	B	C	D	E	F	G
idstudy	study	pubyear	outlet	title	type	
1	Hansen and Singleton (1982)	1982	Econometrica	Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models	article	
2	Hansen and Singleton (1983)	1983	Journal of Political Economy	Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns	article	
3	Mankiw et al. (1985)	1985	Quarterly Journal of Economics	Intertemporal Substitution in Macroeconomics	article	
4	Giovannini (1985)	1985	Journal of Development Economics	Saving and the real interest rate in LDCs	article	
5	Miron (1986)	1986	Journal of Political Economy	Seasonal Fluctuations and the Life Cycle-Permanent Income Model of Consumption	article	
6	Koskela and Viren (1987)	1987	Economics Letters	Inflation and the Euler equation approach to consumption behaviour: Some empirical evidence	article	
7	Ferson and Merrick (1987)	1987	Journal of Financial Economics	Non-stationarity and stage-of-the-business-cycle effects in consumption-based asset pricing relations	article	
8	Hall (1988)	1988	Journal of Political Economy	Intertemporal Substitution in Consumption	article	
9	Whitley (1988)	1988	Journal of Monetary Economics	Some tests of the consumption-based asset pricing model	article	
10	Kugler (1988)	1988	Economics Letters	Intertemporal substitution, taste shocks and cointegration	article	
11	Rossi (1988)	1988	Staff Papers - International Monetary	Government Spending, the Real Interest Rate, and the Behavior of Liquidity-Constrained Consumers in De	article	
12	Zeides (1989)	1989	Journal of Political Economy	Consumption and Liquidity Constraints: An Empirical Investigation	article	
13	Attanasio and Weber (1989)	1989	Economic Journal	Intertemporal Substitution, Risk Aversion and the Euler Equation for Consumption	article	
14	Finn (1989)	1989	Journal of International Money and Fi	An econometric analysis of the intertemporal general-equilibrium approach to exchange rate and current	article	
15	Campbell and Mankiw (1989)	1989	NBER Macroeconomics Annual 1989	Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence	chapter	
16	Koenig (1990)	1990	Quarterly Journal of Economics	Real Money Balances and the Timing of Consumption: An Empirical Investigation	article	
17	Epstein and Zin (1991)	1991	Journal of Political Economy	Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical A	article	
18	Lawrance (1991)	1991	Journal of Political Economy	Poverty and the Rate of Time Preference: Evidence from Panel Data	article	
19	Campbell and Mankiw (1991)	1991	European Economic Review	The response of consumption to income: A cross-country investigation	article	
20	Osano and Inoue (1991)	1991	International Economic Review	Testing between competing models of real business cycles	article	
21	Runkle (1991)	1991	Journal of Monetary Economics	Liquidity constraints and the permanent-income hypothesis	article	
22	Bohn (1991)	1991	Journal of Money, Credit and Banking	On Cash-in-Advance Models of Money Demand and Asset Pricing	article	
23	Gleizer (1991)	1991	Brazilian Review of Econometrics	Saving and real interest rates in Brazil	article	
24	Fauvel and Samson (1991)	1991	Canadian Journal of Economics	Intertemporal substitution and durable goods: an empirical analysis	article	
25	Wirjanto (1991)	1991	Canadian Journal of Economics	Testing the Permanent Income Hypothesis: The Evidence from Canadian Data	article	
26	Kuehlein (1991)	1991	Economics Letters	A test for the presence of precautionary saving	article	
27	Auerbach and Hassett (1991)	1991	National Saving and Economic Perform	Corporate Savings and Shareholder Consumption	chapter	

Snowballing

- No query is perfect
- Idea: cover studies not identified by the query but frequently mentioned in the literature
- For each primary study, download the list of references
- Inspect the 50 studies most frequently cited by the primary studies
- If you can use some of them, add them to your list

Snowballing example

Meta-analysis of social science research: A practitioner's guide

 [View PDF](#)[Full text options](#)[Export](#)[Abstract](#)[Author keywords](#)[Indexed keywords](#)[SciVal Topics](#)[Metrics](#)[Funding details](#)[Funding details](#)

References (114)

[View in search results format](#)

All

[Export](#) [Print](#) [E-mail](#) [Save to PDF](#)[Create bibliography](#)

-  1 Amini, S.M., Parmeter, C.F.

Comparison of model averaging techniques: Assessing growth determinants

(2012) *Journal of Applied Econometrics*, 27 (5), pp. 870-876. Cited 42 times.

doi: 10.1002/jae.2288



-  2 Andrews, I., Kasy, & M.
Identification of and correction for publication bias
(2019) *American Economic Review*, 109 (8), pp. 2766-2794. Cited 180 times.

AI-enhanced literature search

- 1 Use **Deep Research** to explore recent literature and concepts
- 2 Ask **ChatGPT thinking model** to refine your Scholar query
- 3 Download **500 abstracts** from Scholar or Publish or Perish
- 4 Let a **ChatGPT Agent** screen abstracts for relevant estimates
- 5 Ask Agent to download PDFs (or download manually)
- 6 Upload PDFs in **batches of 5**; ask which report usable estimates
- 7 (Optional) Use **ASReview** to double-check or co-train

Using thinking models to design Scholar query

- GPT thinking model helps refine search precision
- Prompt examples:
 - I want to do a meta-analysis of the beauty premium. Suggest a Google Scholar query to find empirical studies on attractiveness and income.
 - Too many irrelevant hits. How can I prioritize regression-based studies only?
- Output improves over iterations: clarify units, methods, outcomes, and keywords

Prompting ChatGPT Agent for screening

- Batch input: 20–50 abstracts at a time from Scholar export

- Example prompt:

*You are an expert in meta-analysis.
For each abstract below and all other
information you have about the paper,
say whether the paper likely reports
quantitative estimates on the effect
of beauty on labor market outcomes that
can be used in a meta-analysis. Justify
briefly.*

- **Output:** Paper 1: Yes (panel data, Table 2)
Paper 2: No (conceptual) Paper 3: Maybe
(ambiguous)

Screening PDFs for quantitative content

- Upload PDFs in small batches (e.g., 5 at a time)

- Example prompt:

For each of these papers, does it contain new empirical estimates of the effect of beauty on wages that I can use in a meta-analysis? Point to specific tables or pages.

- Output format:

- Paper A: Yes, Table 4, OLS and IV estimates on beauty premium
- Paper B: No quantitative analysis
- Paper C: Yes, appendix regression table

Optional enhancements and caveats

- Combine with **ASReview** for machine-assisted filtering
- Use **Semantic Scholar** for better metadata access
- Avoid over-reliance: AI helps, but expert judgment remains key
- Protect against false positives: Always validate flagged studies manually
- End-to-end AI infrastructure for meta-analysis: **otto-SR**
- **AI duels** for meta:
<https://github.com/tjhavranek/research-audit-duel-protocol>
- Forthcoming **MAER-Net guidelines** on AI

Minimum requirements for meta-analysis

For modern meta-analysis techniques to work, you need at least:

- 1 30 estimates
- 2 10 studies
- 3 effect sizes
- 4 standard errors
- 5 sample sizes

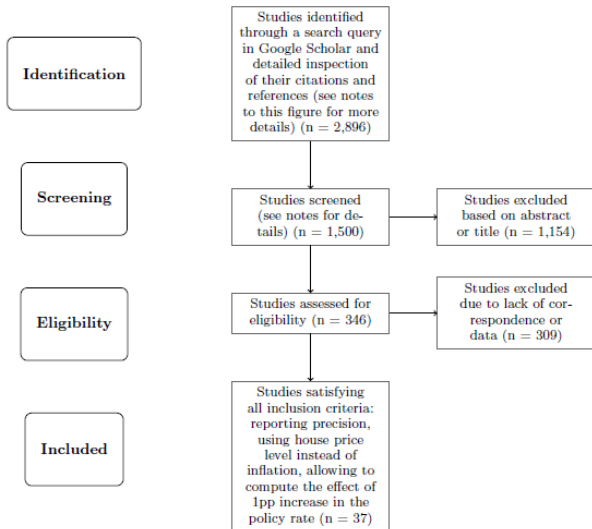
What to do if I have, e.g., just 10 estimates from 4 studies?
Focus on the median or weighted means.

- Some studies are published in poor journals or use weak identification. What to do?
- **Option 1:** exclude these studies ex ante. Not recommended.
- **Option 2:** include these studies with controls for journal quality and methodology. You can always exclude them as a robustness check.
- Note that some fields have standardized quality benchmarks (risk of bias)

Unpublished papers

- Try to collect all studies, published and unpublished
- If unfeasible, it is possible to focus just on published studies (fewer typos, peer-review)
- Including unpublished papers unlikely to alleviate publication bias
- Still too many papers? Recruit a co-author, AI, or use a random subset (last resort)

PRISMA: document your literature search



Workshop Outline

- 1 Motivation
- 2 Literature Search
- 3 Data Collection**
- 4 Conventional Tools
- 5 Publication Bias
- 6 P-hacking
- 7 Heterogeneity

Effects must be quantitatively comparable

What does it mean?

- When we compare estimates reported in Study 1 and Study 2, it must be clear which one is larger
- It must be clear how much larger it is
- t-statistics are not useful measures of effect size
- Regression coefficients are not comparable in general if studies use different units

$$\text{x-elasticity of } y : \varepsilon = \frac{\frac{\partial y}{y}}{\frac{\partial x}{x}}$$

$$\text{P-elasticity of Q: } \varepsilon = \frac{\frac{\partial Q}{Q}}{\frac{\partial P}{P}} \text{ if continuous}$$

$$\varepsilon = \frac{\frac{Q_2 - Q_1}{Q_1}}{\frac{P_2 - P_1}{P_1}} = \frac{\% \text{ change in quantity Q}}{\% \text{ change in price P}} \text{ if discrete}$$

Elasticities can be recovered from regressing **log(Y) on log(X)**.

Semi-elasticities, dollar values, std.dev. changes

- Semi-elasticity: effect of euro adoption on trade, effects of foreign investment on domestic productivity, effects of borders on trade
- **Dollar values**: statistical value of life, social cost of carbon
- Std.dev. changes: effect of class size on student test scores. What to do here?
- Class size is measured the same way in all studies (number of kids).
- Test scores are measured differently.
- Solution: By **how many standard deviations** do test scores improve if we decrease class size by 10 students?

Standardized mean differences

$$\theta = \frac{\mu_1 - \mu_2}{\sigma},$$

where μ_1 is the mean for one population, μ_2 is the mean for the other population, and σ is a standard deviation based on either or both populations.

- Very popular in psychology, not much used in economics. Why?
- For many questions the treatment and control are not binary classifications (we look at the intensity of treatment)

Partial correlation coefficients

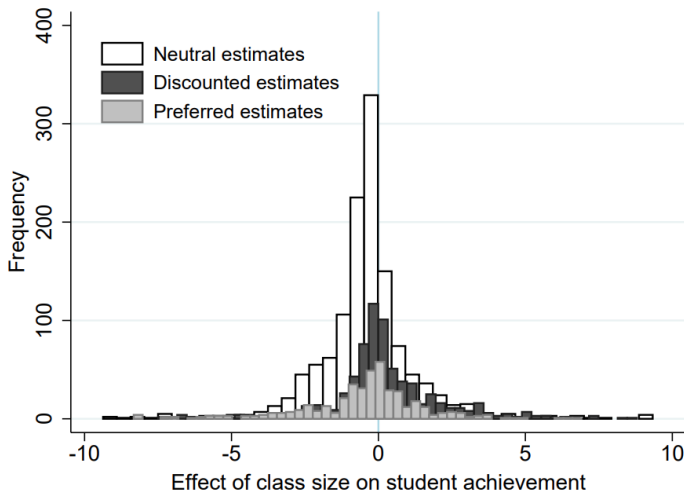
Example: **tuition fees and university enrollment**. Ideally, we would need price elasticities of demand (PED), but many studies are not log on log.

So we transform the estimates to partial correlations using t-statistics and degrees of freedom:

$$\text{PCC(PED)}_{ij} = \frac{T(\text{PED})_{ij}}{\sqrt{T(\text{PED})_{ij}^2 + \text{DF(PED)}_{ij}}}.$$

- Last resort: can be used almost always, but we lose a lot of information
- Difficult to interpret (some guidelines provided by Doucouliagos, 2011)
- Statistical problems for meta-analysis methods (but this is the case also for SMD)
- Include robustness checks: e.g. a subsample of elasticities in the case of tuition fees and university enrollment

Collect all estimates



Standard errors

- Absolutely crucial for meta-analysis
- If not reported directly, can be computed from t-statistics or p-values
- Last resort: approximation based on within-study variation (Matousek et al. 2022)
- Interactions, transformations: delta method

AI-assisted data collection: Overview

- As of March 2026, **human review still essential**
- AI reasonably good at
 - Identifying key study dimensions (estimation method, outcome type, etc.)
 - Extracting basic data with structured prompts and examples
 - Reducing workload to a single reviewer
- Tools: ChatGPT thinking, Agent Mode, NotebookLM
- End-to-end AI infrastructure for meta-analysis: **otto-SR**
- Don't rely on one model (add Claude, Gemini, Grok)!!

Step-by-step data collection workflow

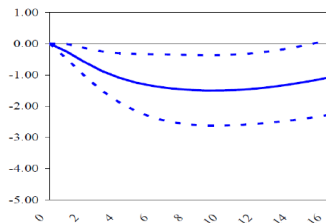
- 1 Use **Deep Research + thinking model** to find key coding dimensions
- 2 Examples: OLS vs IV, cross-section vs panel, outcome types, country, controls
- 3 Upload a few example PDFs + your coded spreadsheet to ChatGPT Agent
- 4 Prompt Agent: "Extract these columns from new PDFs using my examples as a guide."
- 5 Refine extraction prompts based on feedback
- 6 Use **NotebookLM** as an independent extraction check
- 7 Validate discrepancies manually; spot-check the rest

Example prompt for PDF data extraction

Here are 3 PDFs and the corresponding rows from my meta-analysis data sheet. Learn the structure. These data are collected well and I want you to collect data from other papers. I will send 5 new PDFs. Extract estimates of the beauty premium and the corresponding standard error. Collect information on the estimation method used, the definition of the beauty variable, definition of the earnings variable.

- Agent mode allows memory and consistency across batches
- Output can be verified, adjusted, and exported as CSV
- See summary on MAER-Net website.

Graphical estimates, measurement error



Canada		
horizon	response	confidence interval
1	-0.231	(-0.407, -0.016)
2	-0.578	(-1.012, -0.135)
4	-1.076	(-1.853, -0.299)
8	-1.454	(-2.554, -0.359)
12	-1.414	(-2.534, -0.287)
16	-1.084	(-2.259, 0.131)
max	n.a.	n.a.

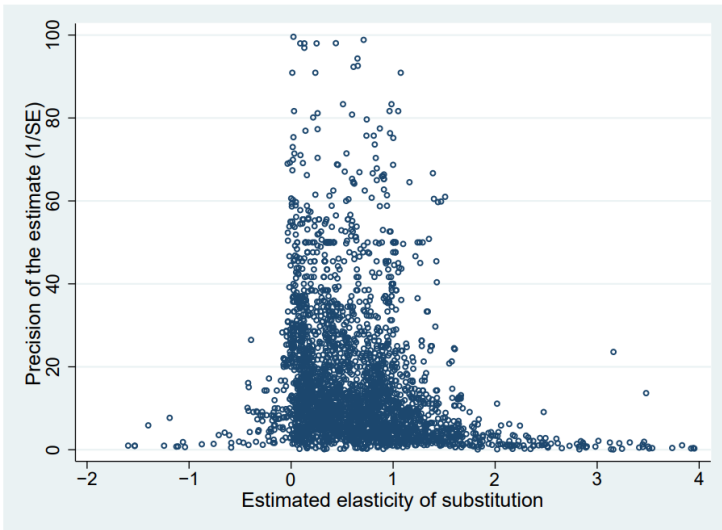
Obtained from: page 50, Figure 2

Frequency: Quarterly

Note: Responses are in %.

Details: Ehrenbergerova et al. (2023, IMF Economic Review)

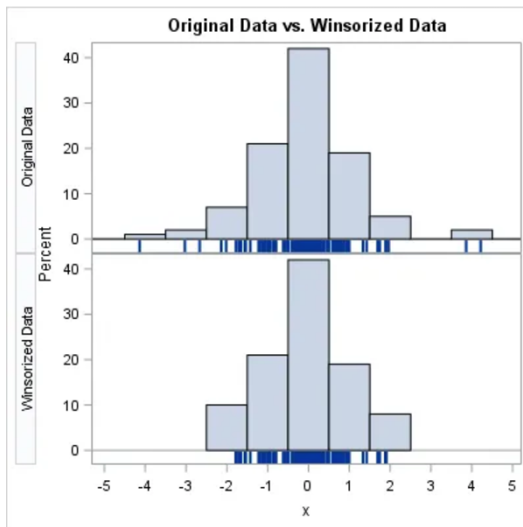
Outliers



What to do with outliers?

- No consensus
- Approach 1: omit very large effects (and standard errors), e.g. more than 3 standard deviations away from the mean
- Approach 2: run robust meta-regression or use tools for outlier detection in regression
- Approach 3: use winsorizing (my personal preference)
- Always report robustness checks

Winsorization



Open Stata

Reflecting context: the effect of beauty on success

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	study_id	study	country	premium	se_premium	precision	tstat	pre_probs	lnnumer	weight	interview	photo_ra	software	self_rate	dummy_1	category	beauty_r	number	salary	study_o	teaching	athletic	electoral	other
2	1	Ahmed e	United St	8.91	2.9601	0.3378	3.01	1806	42.497	0.1667	0	0	1	0	0	1	0	4.4427	1	0	0	0	0	0
3	1	Ahmed e	United St	8.964	5.2723	0.1696	1.7	1806	42.497	0.1667	0	0	1	0	0	1	0	4.4427	1	0	0	0	0	0
4	1	Ahmed e	United St	2.484	3.0233	0.3301	0.82	1806	42.497	0.1667	0	0	1	0	0	1	0	4.4427	1	0	0	0	0	0
5	1	Ahmed e	United St	9.18	4.8316	0.207	1.9	1806	42.497	0.1667	0	0	1	0	0	1	0	4.4427	1	0	0	0	0	0
6	1	Ahmed e	United St	21.6	6.4671	0.1546	3.34	1806	42.497	0.1667	0	0	1	0	0	1	0	4.4427	1	0	0	0	0	0
7	1	Ahmed e	United St	16.012	10.075	0.0993	1.49	1806	42.497	0.1667	0	0	1	0	0	1	0	4.4427	1	0	0	0	0	0
8	2	Ahn & Le	World	0.0375	1.0124	0.9878	0.037	774	27.821	0.0714	0	1	0	0	1	0	1	2.7726	0	0	0	1	0	0
9	2	Ahn & Le	World	0.6186	0.2674	3.7391	2.313	774	27.821	0.0714	0	1	0	0	1	0	1	2.7726	0	0	0	1	0	0
10	2	Ahn & Le	World	0.2825	0.1421	7.0396	1.989	774	27.821	0.0714	0	1	0	0	1	0	1	2.7726	0	0	0	1	0	0
11	2	Ahn & Le	World	0.0057	0.5684	1.7593	0.01	774	27.821	0.0714	0	1	0	0	1	0	1	2.7726	1	0	0	0	0	0
12	2	Ahn & Le	World	0.2691	1.4087	0.7099	0.191	774	27.821	0.0714	0	1	0	0	1	0	1	2.7726	1	0	0	0	0	0
13	2	Ahn & Le	World	0.7858	0.8459	1.1822	0.529	774	27.821	0.0714	0	1	0	0	1	0	1	2.7726	1	0	0	0	0	0
14	2	Ahn & Le	World	0.2622	0.4685	2.1346	0.5537	774	27.821	0.0714	0	1	0	0	1	0	1	2.7726	0	0	0	1	0	0
15	2	Ahn & Le	World	0.4506	0.3763	2.6576	1.1574	774	27.821	0.0714	0	1	0	0	1	0	1	2.7726	0	0	0	1	0	0
16	2	Ahn & Le	World	0.1326	2.3261	0.4299	0.057	774	27.821	0.0714	0	1	0	0	1	0	1	2.7726	1	0	0	0	0	0
17	2	Ahn & Le	World	1.2183	2.1218	0.4713	0.5742	774	27.821	0.0714	0	1	0	0	1	0	1	2.7726	1	0	0	0	0	0
18	2	Ahn & Le	World	0.2301	0.1175	8.5093	1.568	399	18.975	0.0714	0	1	0	0	1	0	1	2.7726	0	0	0	1	0	0
19	2	Ahn & Le	World	0.2023	0.2152	4.6471	0.34	375	19.365	0.0714	0	1	0	0	1	0	1	2.7726	0	0	0	1	0	0
20	2	Ahn & Le	World	-0.561	0.8193	1.2196	-0.684	399	18.975	0.0714	0	1	0	0	1	0	1	2.7726	1	0	0	0	0	0
21	2	Ahn & Le	World	0.3107	0.686	1.4578	0.453	375	19.365	0.0714	0	1	0	0	1	0	1	2.7726	1	0	0	0	0	0
22	3	Anzcova	Czech Rk	0	7.1	0.1408	0	1031	32.109	0.25	1	1	0	1	1	0	0	0.6931	1	0	0	0	0	0
23	3	Anzcova	Czech Rk	9.697	5.1087	0.1957	1.8982	1031	32.109	0.25	1	1	0	1	1	0	0	0.6931	1	0	0	0	0	0
24	3	Anzcova	Czech Rk	17.468	6.9226	0.1445	2.5234	1031	32.109	0.25	1	1	0	1	1	0	0	0.6931	1	0	0	0	0	0
25	3	Anzcova	Czech Rk	5.4461	5.5562	0.18	0.9802	1031	32.109	0.25	1	1	0	1	1	0	1	0.6931	1	0	0	0	0	0
26	4	Arunach	Ecuador	12.412	2.6304	0.3802	4.786	1960	44.272	0.0417	1	0	0	0	0	1	0	0	1	0	0	0	0	0
27	4	Arunach	Ecuador	10.672	2.5897	0.3861	4.1209	1960	44.272	0.0417	1	0	0	0	0	1	0	0	1	0	0	0	0	0
28	4	Arunach	Ecuador	6.4388	3.3209	0.3011	1.9389	1960	44.272	0.0417	1	0	0	0	0	1	0	0	1	0	0	0	0	0
29	4	Arunach	Ecuador	4.7912	3.2635	0.3059	1.4654	1960	44.272	0.0417	1	0	0	0	0	1	0	0	1	0	0	0	0	0
30	4	Arunach	Mexico	28.018	5.1207	0.1953	5.4715	923	30.381	0.0417	1	0	0	0	0	1	0	0	1	0	0	0	0	0
31	4	Arunach	Mexico	16.416	5.8208	0.1718	2.8202	923	30.381	0.0417	1	0	0	0	0	1	0	0	1	0	0	0	0	0
32	4	Arunach	Mexico	16.416	5.8208	0.1718	2.8202	923	30.381	0.0417	1	0	0	0	0	1	0	0	1	0	0	0	0	0
33	4	Arunach	Mexico	11.016	6.6609	0.1501	1.6538	923	30.381	0.0417	1	0	0	0	0	1	0	0	1	0	0	0	0	0
34	4	Arunach	Ecuador	10.885	3.0023	0.3331	3.6254	1960	44.272	0.0417	1	0	0	0	1	0	0	0	1	0	0	0	0	0

Measurement

Measurement of beauty

- Interviewer-rated beauty
- Photo-rated beauty
- Software-rated beauty
- Self-rated beauty
- Dummy beauty
- Categorical beauty
- Beauty premium
- Beauty penalty

Measurement of success

- Earnings
- Study outcomes
- Teaching & research outcomes
- Athletic success
- Electoral success
- Other outcomes

Data, estimation, publication

Data characteristics

- Male subjects
- Female subjects
- Mixed gender
- High-skilled workers
- Prostitutes
- Other dressy occupations
- Non-dressy occupations
- Western culture
- Other cultures
- Panel data
- Cross-section

Estimation technique

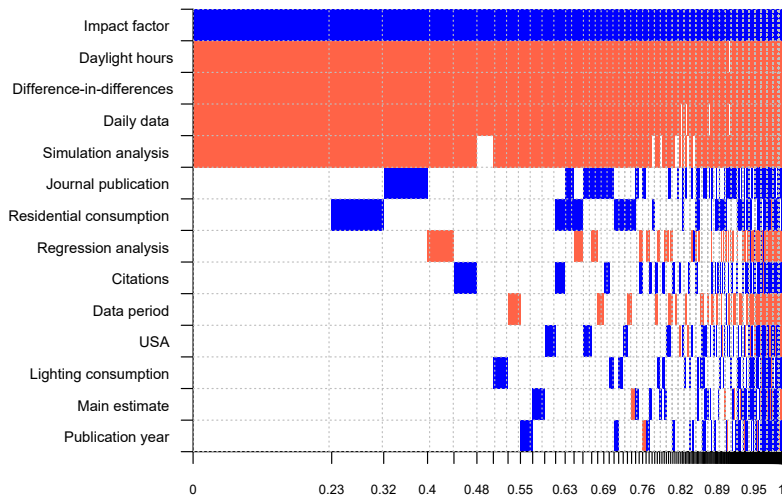
- Ordinary least squares
- Instrumental variables
- Difference-in-differences
- Other method
- Cognitive skill control
- No cognitive skill control

Publication characteristics

- Published study
- Unpublished study
- High-quality peer review

- Meta-analysis can be used to test hypotheses that primary studies cannot investigate
- Example: the effect of daylight saving time on energy consumption
- Individual studies have data for individual countries (or provinces)
- In meta-analysis, we have data for dozens of countries
- We can investigate the sources of cross-country differences

DST savings and latitude



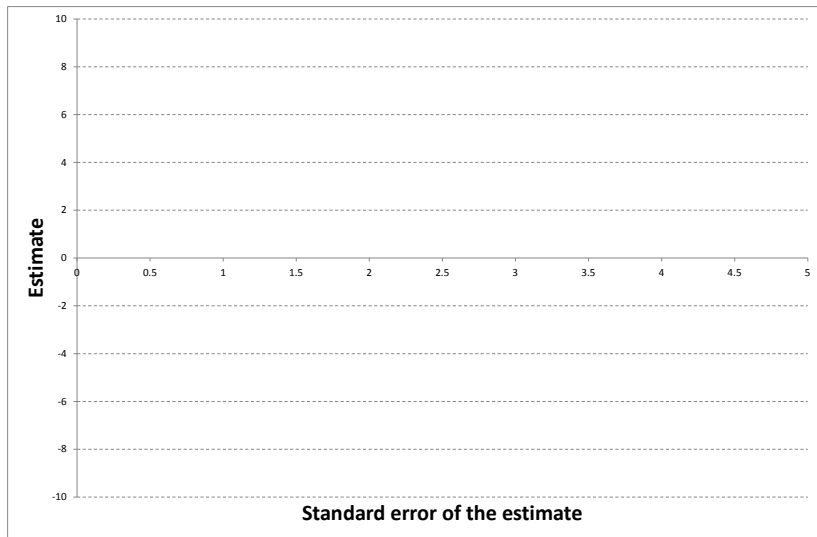
Workshop Outline

- 1 Motivation
- 2 Literature Search
- 3 Data Collection
- 4 Conventional Tools**
- 5 Publication Bias
- 6 P-hacking
- 7 Heterogeneity

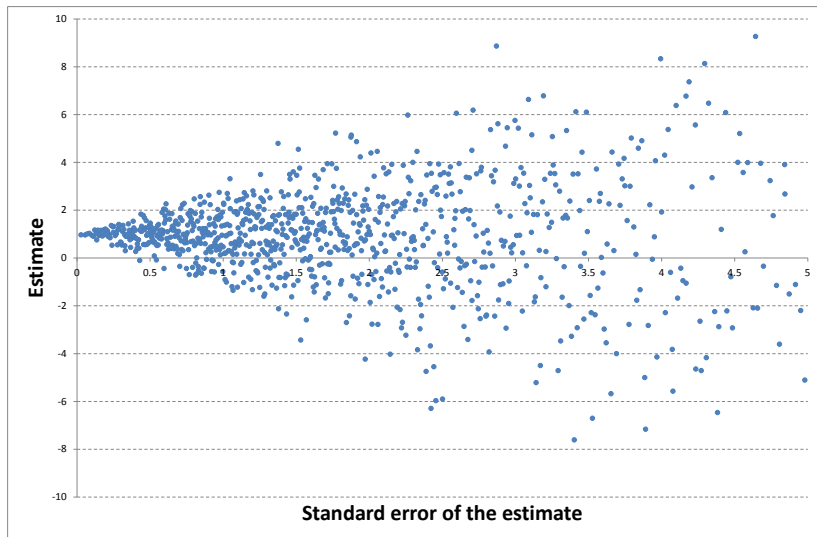
Summary statistics for meta-analysis

- Simple mean of reported estimates: natural starting point
- But inefficient and affected by outliers, publication bias, p-hacking
- Median: surprisingly robust to most problems in meta-analysis
- If you have only a few studies or estimates, focus on the (unweighted) median!
- Classical meta-analysis approach: use inverse variance as the weight

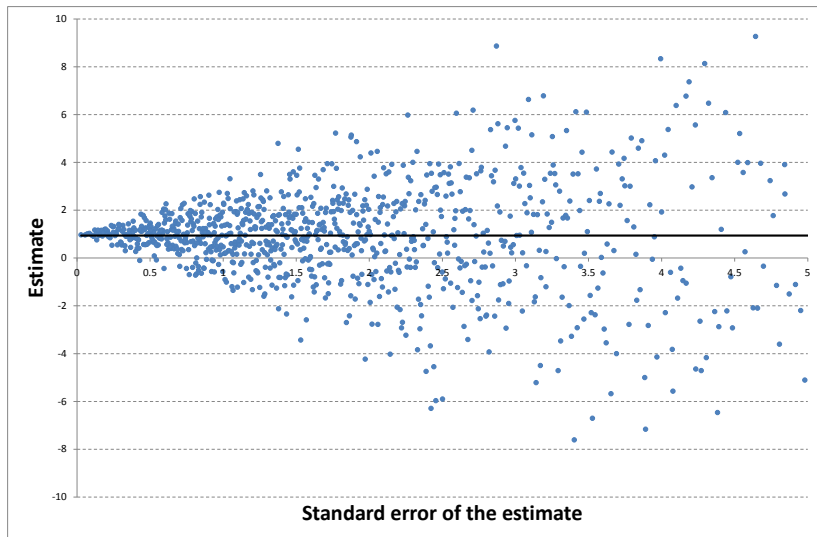
More precision -> more weight



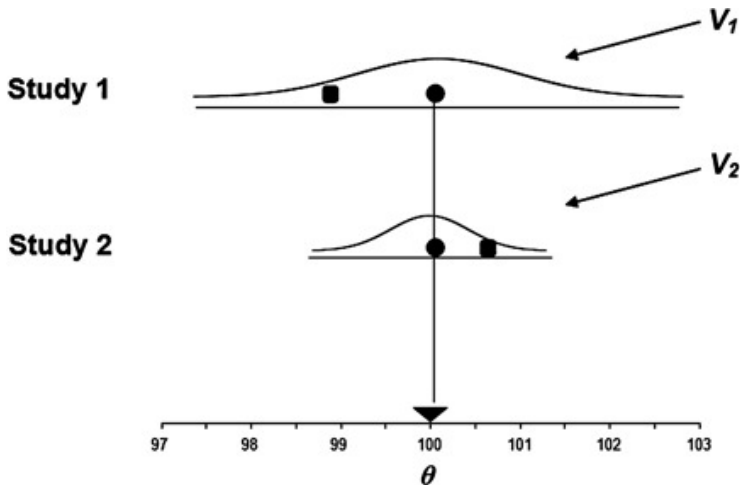
More precision -> more weight



More precision -> more weight



“Fixed effect” estimator in meta-analysis



“Fixed effect” estimator in meta-analysis

Mean weighted by inverse variance of individual estimates.
Very sensitive to outliers in precision (including typos).

$$Y_i = \theta + \varepsilon_i, \quad (1)$$

$$V_i = \frac{\sigma^2}{n}, \quad (2)$$

$$W_i = \frac{1}{V_i}, \quad (3)$$

$$M = \frac{\sum_{i=1}^k W_i Y_i}{\sum_{i=1}^k W_i}, \quad (4)$$

$$V_M = \frac{1}{\sum_{i=1}^k W_i}. \quad (5)$$

Open Stata

FE vs. UWLS: point estimate

- Unrestricted weighted least squares yield the same point estimate:

$$\hat{\theta} = \frac{\sum w_i \theta_i}{\sum w_i}, \quad w_i = \frac{1}{SE_i^2}$$

- Difference: Variance estimation.

Method	Point Estimate
Fixed-Effect (FE)	$\hat{\theta} = \frac{\sum w_i \theta_i}{\sum w_i}$
UWLS	$\hat{\theta} = \frac{\sum w_i \theta_i}{\sum w_i}$

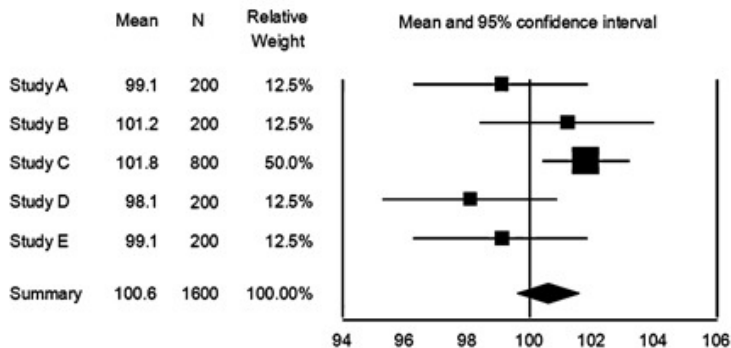
Same estimate, different variance!

FE vs. UWLS: variance

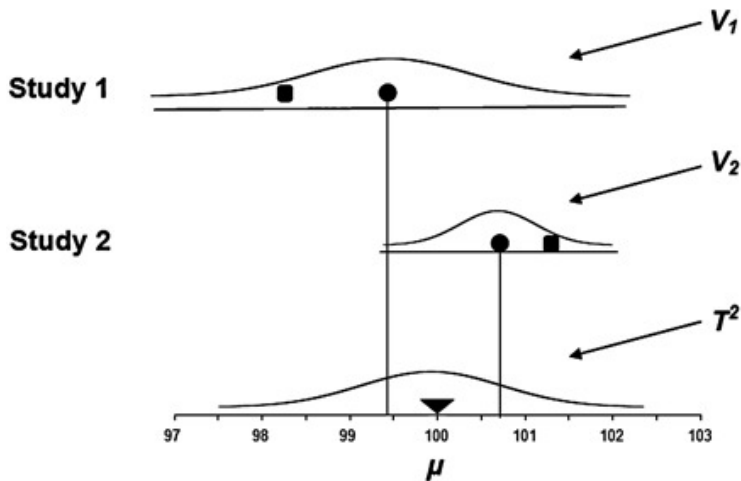
Feature	Fixed-Effect (FE)	UWLS
Assumption	Identical θ	Allows heterogeneity
Variance formula	$\frac{1}{\sum w_i}$	$\frac{\sum w_i(\theta_i - \hat{\theta})^2}{\sum w_i}$
Variance components	Within-study only	Within + between-study
Robustness	Sensitive to mis-specification	More robust

FE typically underestimates variance. Use UWLS!

Aptitude score at one college



Random effects



Random effects

The weight is now diluted by a heterogeneity term. More robust to outliers in precision and heterogeneity, less robust to publication bias.

$$Y_i = \mu + \xi_i + \varepsilon_i, \quad (1)$$

$$V_i = \frac{\sigma^2}{n}, \quad (2)$$

$$W_i = \frac{1}{V_i + \tau^2}, \quad (3)$$

$$M = \frac{\sum_{i=1}^k W_i Y_i}{\sum_{i=1}^k W_i}, \quad (4)$$

$$V_M = \frac{1}{\sum_{i=1}^k W_i}. \quad (5)$$

Estimating τ^2 in random-effects model

Between-study variance τ^2 accounts for heterogeneity across studies. Common estimators:

- **DerSimonian-Laird (DL):**

$$\tau_{DL}^2 = \frac{Q - (k - 1)}{C}, \quad C = \sum w_i - \frac{\sum w_i^2}{\sum w_i}$$

- Simple, but underestimates τ^2 when k is small.

- **Restricted Maximum Likelihood (REML):**

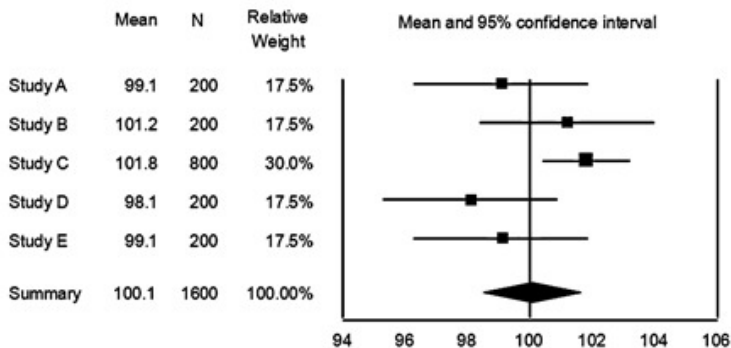
$$\max_{\tau^2} \sum \log(w_i^* + \tau^2) + \sum \frac{(y_i - \hat{\theta})^2}{w_i^* + \tau^2}$$

- More accurate, default in software.

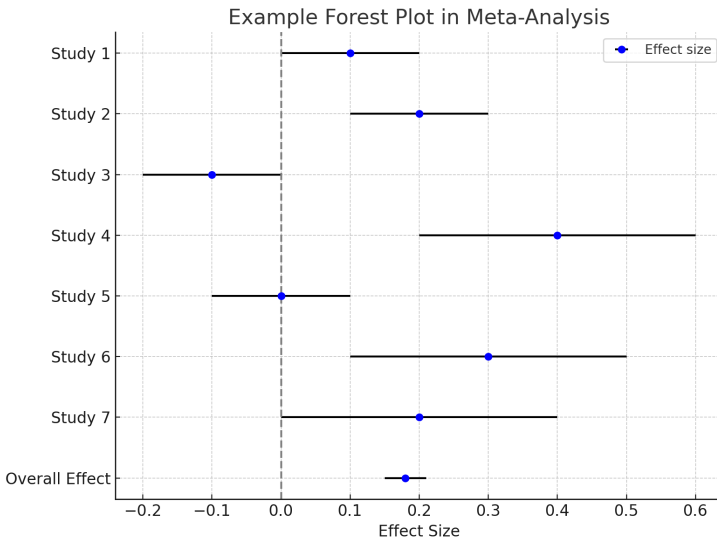
Rule of thumb: Use REML unless simplicity is needed!

Random effects

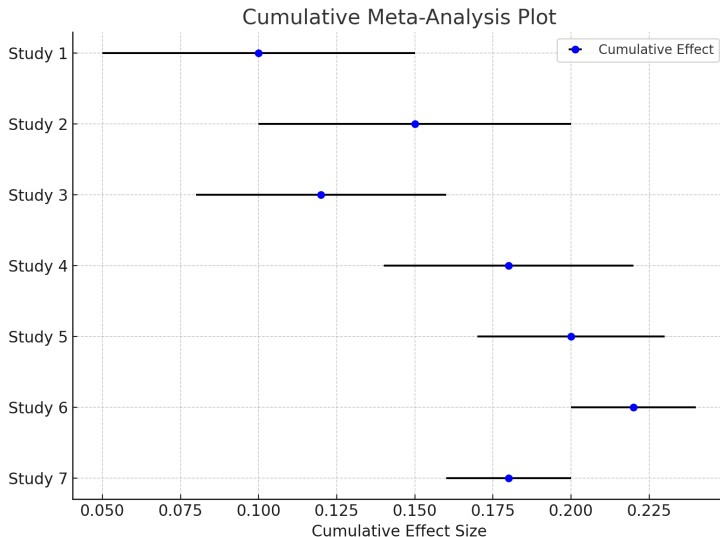
Aptitude score at all colleges



Forest plot



Cumulative meta-analysis plot



Workshop Outline

- 1 Motivation
- 2 Literature Search
- 3 Data Collection
- 4 Conventional Tools
- 5 Publication Bias**
- 6 P-hacking
- 7 Heterogeneity

- **Background:**

- Paxil (paroxetine): Antidepressant approved for adolescents, developed by GlaxoSmithKline (GSK).
- Study 329 (1998-2001): Clinical trial funded by GSK to test Paxil's safety and efficacy in adolescents.

- **What Happened?**

- Published study claimed Paxil was *"well-tolerated and effective."*
- Internal documents later revealed:
 - Negative results (e.g., increased suicidal ideation) were downplayed.
 - Positive outcomes were selectively reported.

- **Discovery of Bias:**
 - Legal proceedings in 2012 forced disclosure of internal data.
 - Independent reanalysis of Study 329 showed:
 - No significant benefit over placebo.
 - Serious safety concerns (e.g., suicidal ideation).
- **Key Lessons:**
 - Publication bias can harm patients and public trust.
 - Highlights the need for:
 - Transparency in research data.
 - Independent replication and reanalysis.

No bias, no E-SE correlation (?)

Researchers estimate

$$Earnings_j = \gamma Beauty_j + \dots + w_j.$$

They assume that $\hat{\gamma}/SE_{\hat{\gamma}}$ has a t -distribution.

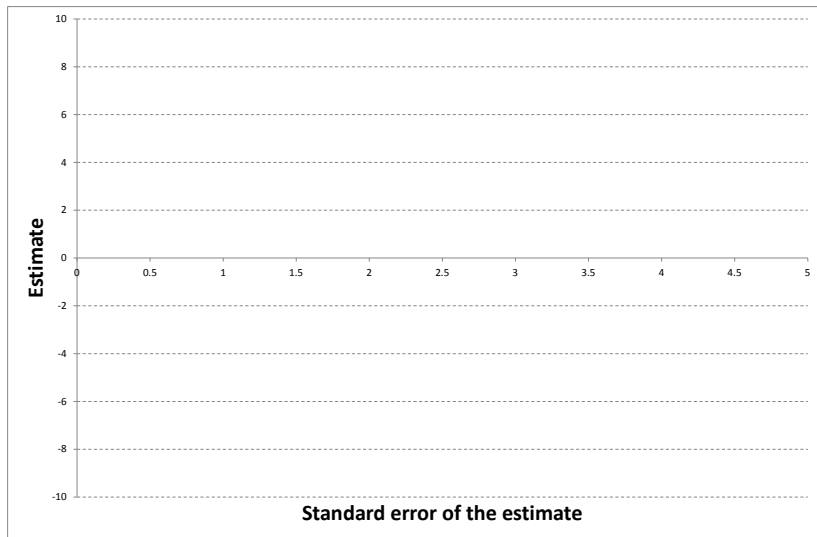
→ estimates should not be correlated with standard errors
(Card & Krueger, 1995)

Types of publication bias

Type I: Selection for the “correct sign.”

Type II: Selection for statistical significance.

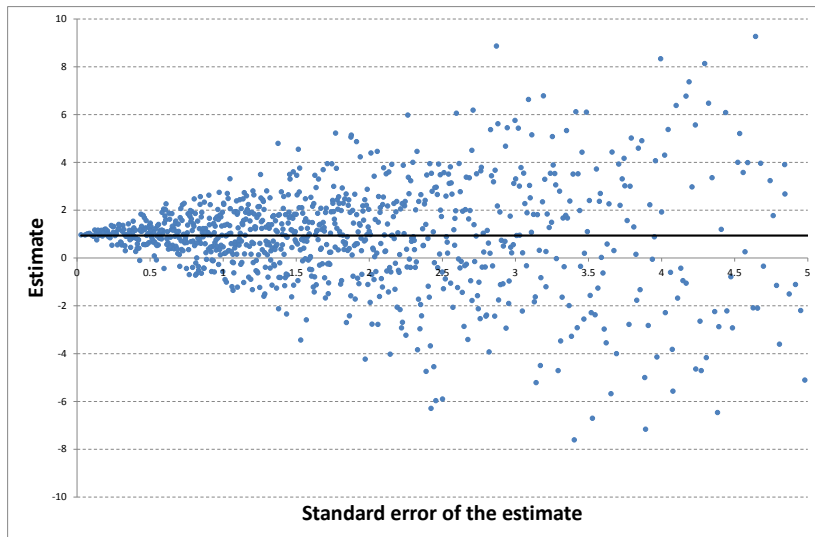
Simulation



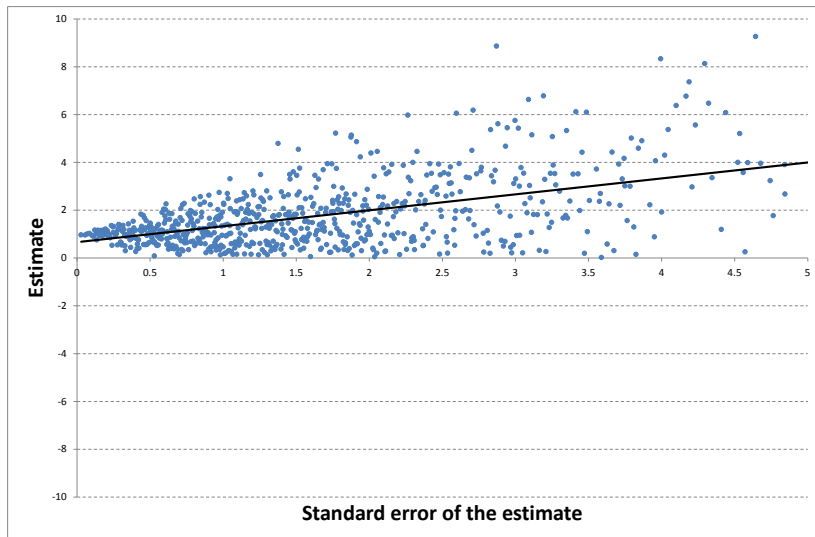
Simulation

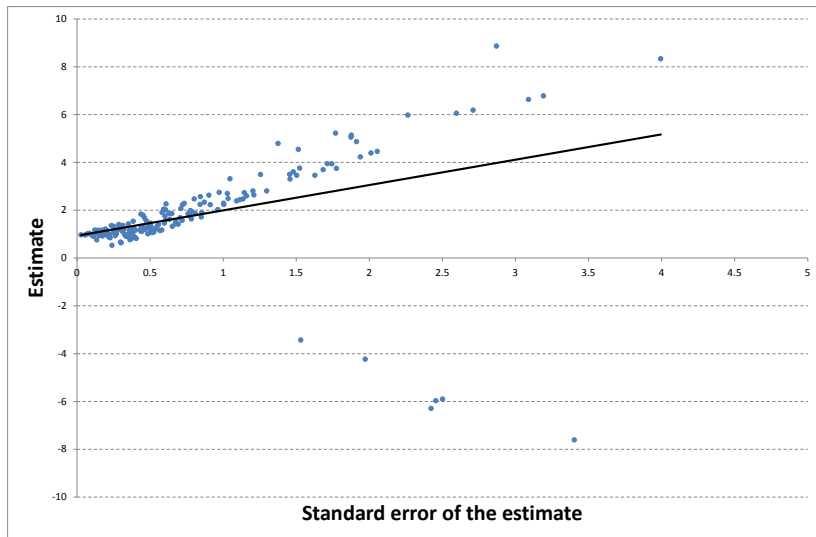


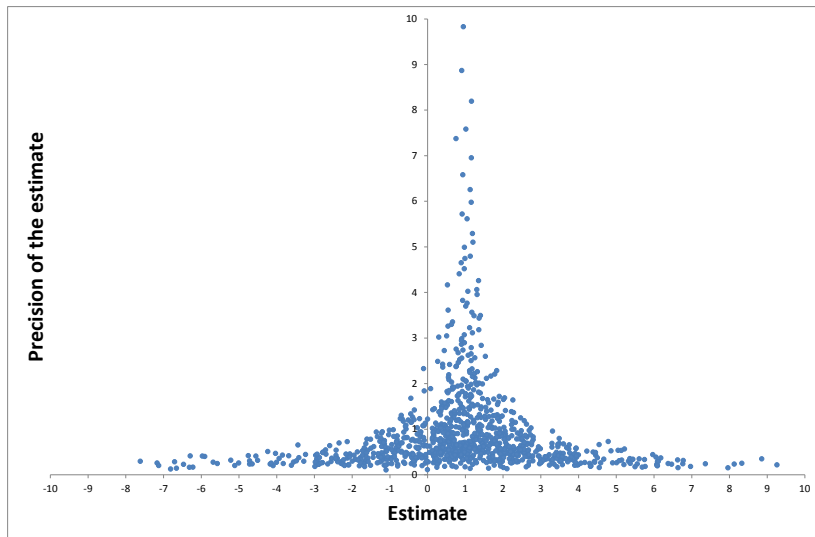
Simulation



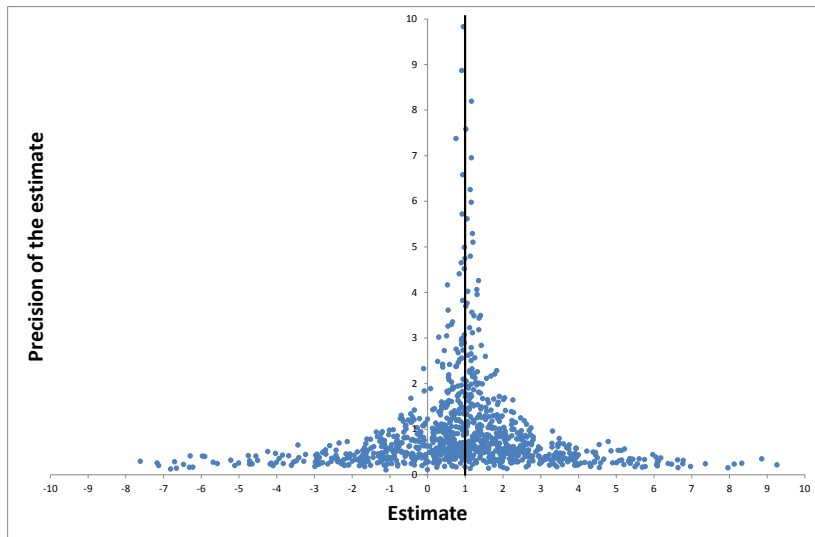
Simulation



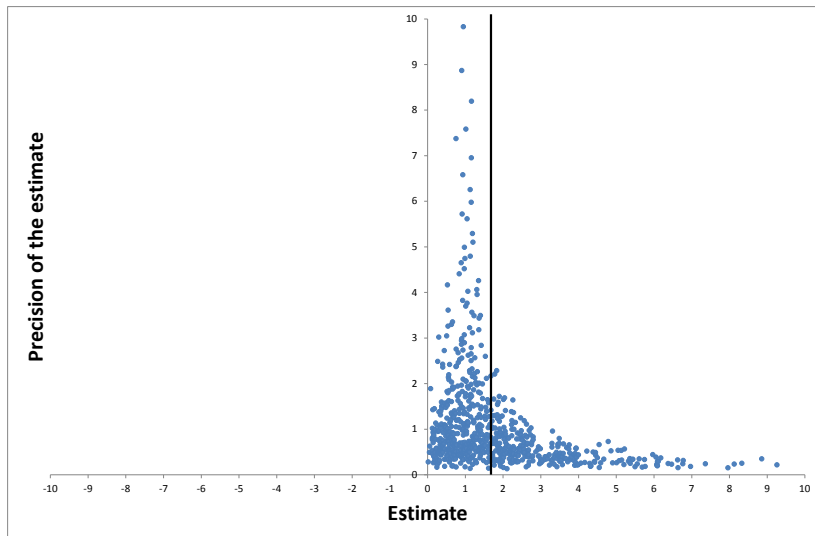




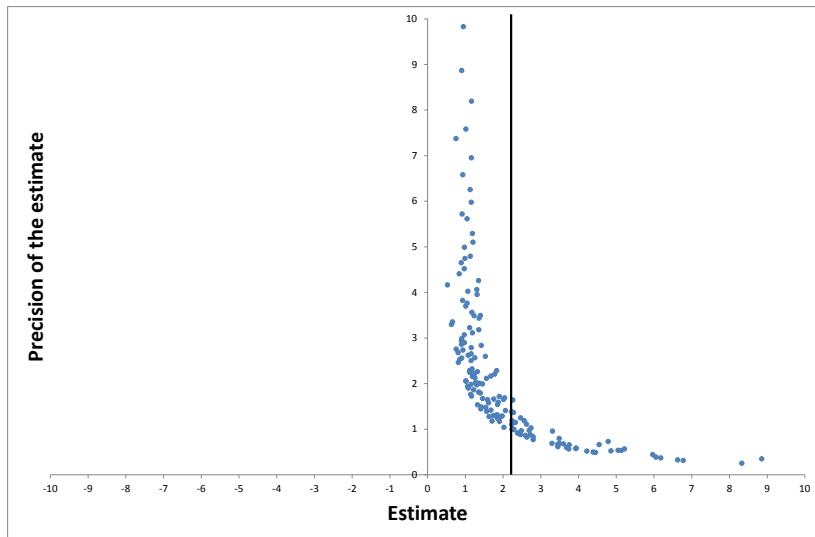
Simulation



Simulation



Simulation



Funnel asymmetry test

In the absence of publication bias the funnel should be symmetrical:

$$\text{estimate}_i = \underbrace{\beta}_{\text{true effect}} + \underbrace{\beta_0 SE_i}_{\text{publication bias}} + \mu_i.$$

The equation is heteroscedastic. Weighted least squares yield

$$t_i = \beta_0 + \beta \left(\frac{1}{SE_i} \right) + \vartheta_i.$$

This is call the funnel asymmetry test—precision effect test (FAT-PET).

Open Stata

Default nonlinear correction model (PEESE)

Publication bias unlikely to be linear in SE. Precision effect test with standard error:

$$\text{estimate}_i = \underbrace{\beta}_{\text{true effect}} + \underbrace{\beta_0 SE_i^2}_{\text{publication bias}} + \mu_i.$$

Weighted least squares yield

$$t_i = \beta_0 SE_i + \beta \left(\frac{1}{SE_i} \right) + \vartheta_i.$$

Works well under most circumstances.

WAAP (Ioannidis et al., 2017)

Weighted **A**verage of **A**dequately **P**owered (WAAP) estimates.

- Focus on studies with adequate statistical power to reduce the impact of selective reporting.
- Combines results only from studies with at least 80% power.

Issues:

- Simple and robust.
- Need to estimate the true effect first and then compute power.

Stem-based correction technique (Furukawa, 2019)

Stem-based method:

- A non-parametric estimator that exploits the variance-bias trade-off.
- Focuses on the most precise estimates, referred to as the “stem” of the funnel plot.

Key features:

- Robust to various publication selection processes.
- Does not rely on specific distributional assumptions.

Application:

- Filters studies with the highest precision (lowest standard errors) to estimate the true effect.
- Reduces influence from noisier, potentially biased studies in the broader funnel plot.

Endogenous kink model (Bom & Rachinger, 2019)

Overview:

- Fits a piecewise linear model with a kink determined by the data.
- Identifies where publication selection is less likely to occur.

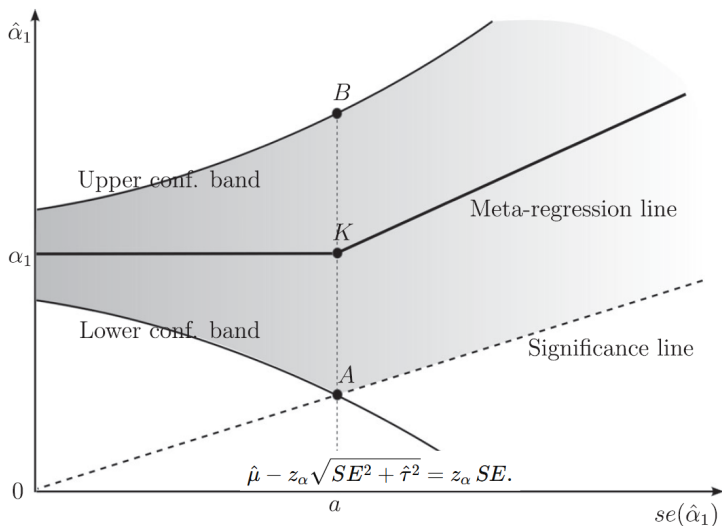
Key features:

- Endogenously estimates the cutoff based on standard errors.
- Often reduces bias and improves efficiency compared to traditional methods.

Application:

- Highlights genuine effects by focusing on less selective segments of the data.
- Effective in meta-analyses with substantial heterogeneity in study precision.

Endogenous kink



Selection model (Andrews & Kasy, 2019)

Overview:

- Models the selection process explicitly to address publication bias.
- Assumes researchers are more likely to publish statistically significant results.

Key features:

- Uses the likelihood of publication as a function of statistical significance.
- Adjusts estimates to reflect the missing studies.

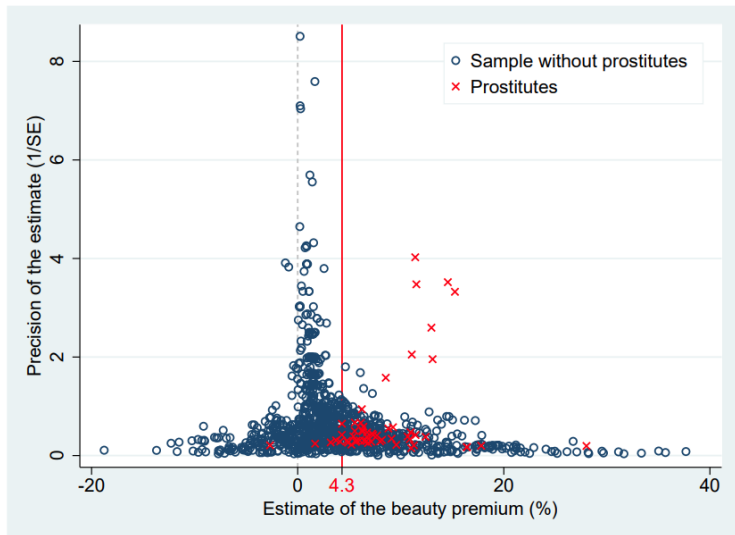
Comparison to funnel-based models:

- Selection models are more rigorous but require stronger assumptions about the publication process.
- Funnel methods are simpler and more robust to model misspecification.

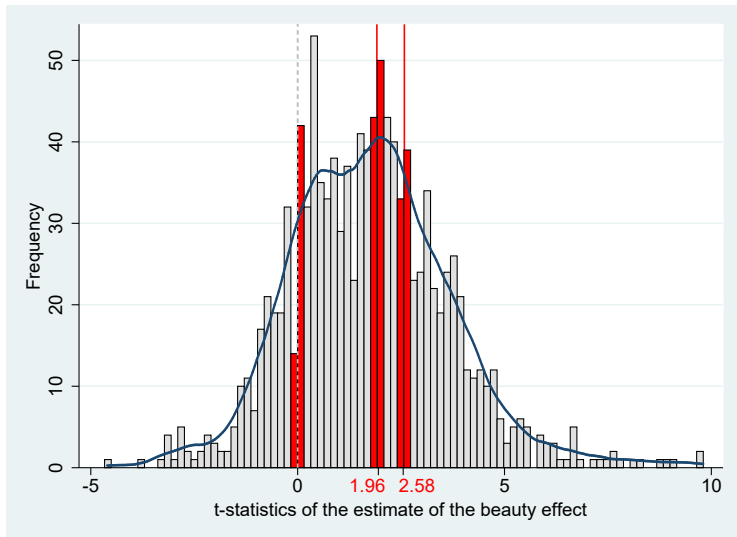
Beauty effect: corrected mean probably below 3

Panel A	OLS	FE	BE	MAIVE	Weighted
Publication bias (<i>standard error</i>)	0.377*** (0.118) [0.116, 0.634]	0.208* (0.119)	0.732*** (0.164)	0.758 (0.489) {-0.064, 2.160}	0.656** (0.272) [-0.009, 1.268]
Effect beyond bias (<i>constant</i>)	2.865*** (0.464) [1.916, 3.781]	3.497*** (0.446)	2.243** (0.871)	1.436 (1.866) {-0.378, 3.251}	3.424*** (1.156) [0.776, 6.339]
First-stage robust F-stat				15.8	
Observations	1,159	1,159	1,159	1,159	1,159
Panel B	Precision-weighted	WAAP	Stem	Kink	Selection
Publication bias	1.720*** (0.084) [1.262, 2.166]			1.720*** (0.084) [1.262, 2.166]	P = 0.142 (0.038)
Effect beyond bias	0.343 (0.118) [-0.039, 1.635]	0.323** (0.147)	0.055 (1.276)	0.343 (0.118) [-0.039, 1.635]	0.493 (0.410)
Observations	1,159	1,159	1,159	1,159	1,159

Sex workers stand apart



Bias caused by selection for a positive sign



Workshop Outline

- 1 Motivation
- 2 Literature Search
- 3 Data Collection
- 4 Conventional Tools
- 5 Publication Bias
- 6 P-hacking**
- 7 Heterogeneity

Web app for p-hacking correction: EasyMeta.org

Seamless Meta-Analysis with MAIVE

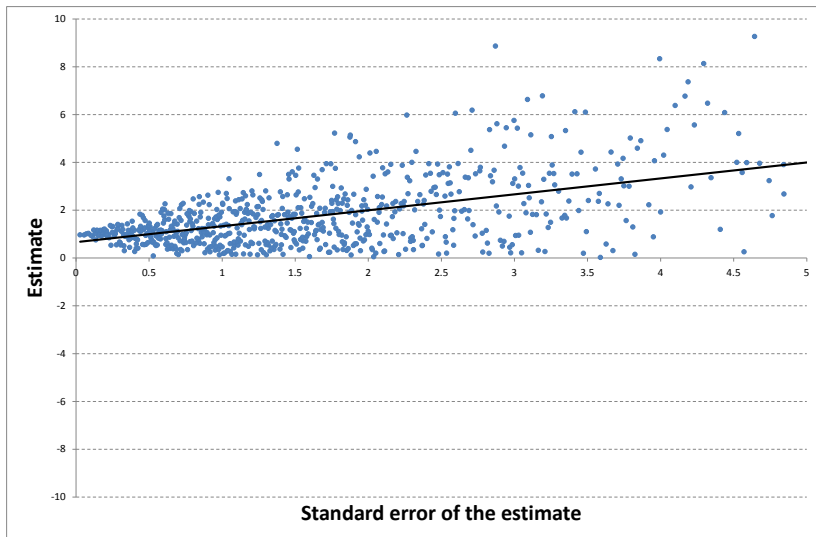
Adjust your data for publication bias, p-hacking, and spurious precision.
Powered by the MAIVE estimator (*Nature Communications*).

Upload Your Data

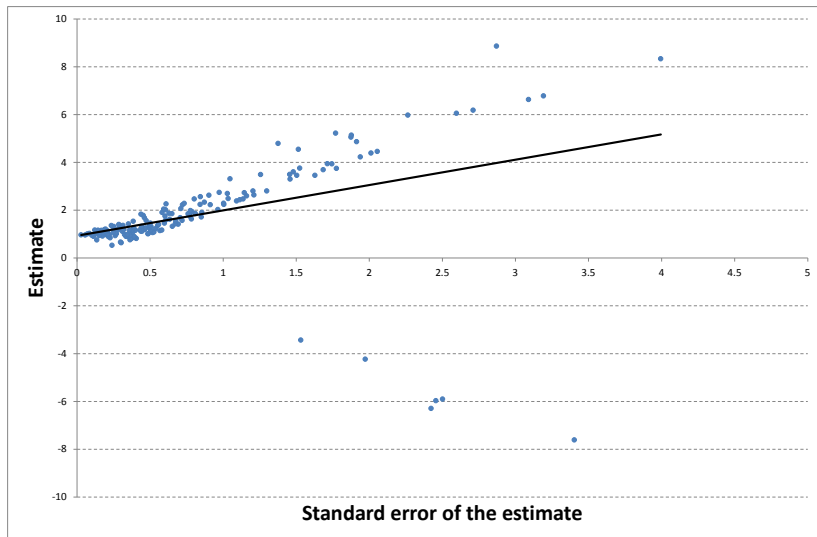
 What is MAIVE?

 Run a demo

Recall publication bias



Recall publication bias

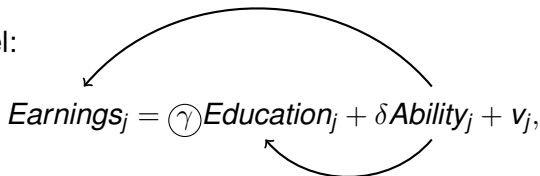


Key difference

- **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)

Example: education premium

True model:

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


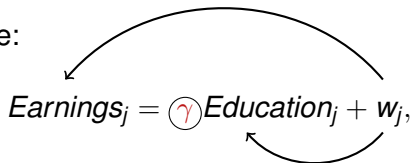
Ability **not observed**.

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.

Example: education premium

Proxy:


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

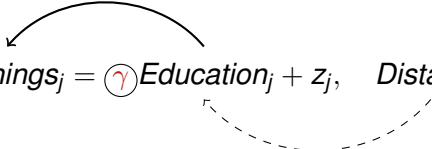

Ability **not observed**.

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.

Example: education premium

Instrument:

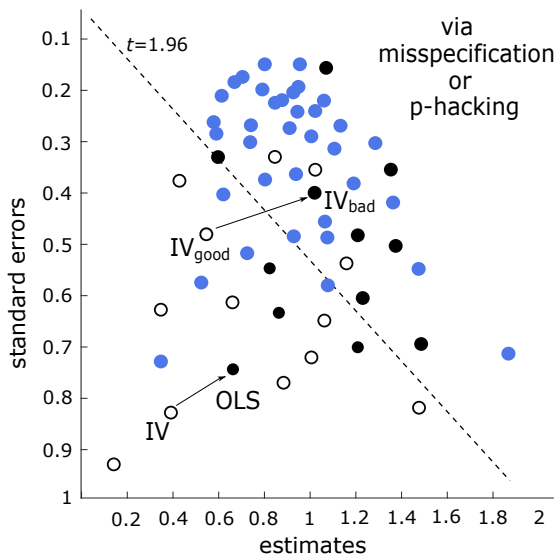
$$\text{Earnings}_j = \gamma \text{Education}_j + z_j, \quad \text{Distance}_j$$


Ability **not observed**.

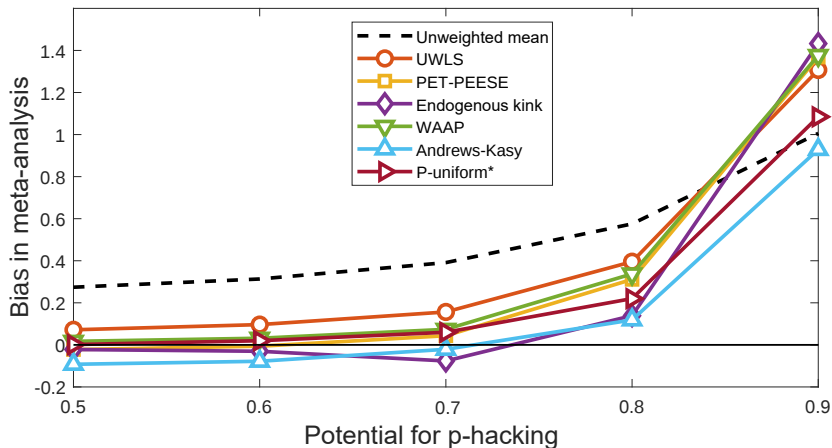
Primary studies:

- 1 ignore ability $\rightarrow \hat{\gamma}$ too large, $\text{SE}(\hat{\gamma})$ too small.
- 2 include a proxy $\rightarrow \hat{\gamma}$ smaller, $\text{SE}(\hat{\gamma})$ larger.
- 3 quasi-experiment $\rightarrow \hat{\gamma}$ even smaller, $\text{SE}(\hat{\gamma})$ 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 (2.19)	5.37 (1.26)	5.36 (1.21)	5.37 (1.19)
White/Asian			8.35 (1.35)	8.44 (1.36)
Girl			4.48 (0.63)	4.39 (0.63)
Free lunch			-13.15 (0.77)	-13.07 (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

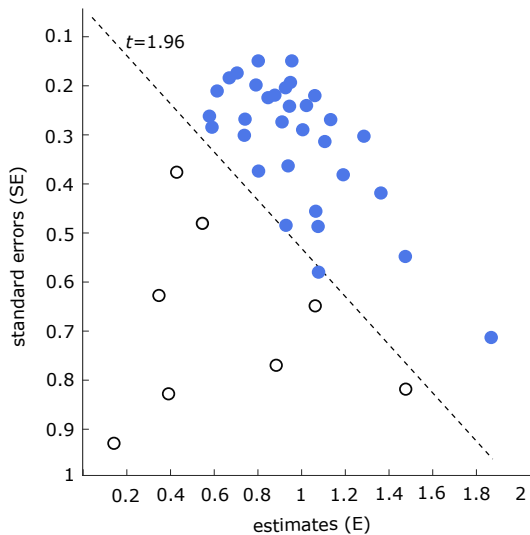
	Krueger (1999)	Replications using different computations of SE				
	(1)	(2) Bootstrap of class clusters	(3) Class clusters	(4) School clusters	(5) Huber-White SE	(6) Plain vanilla SE
Small class	4.82 (2.19)	4.71 (2.00)	4.71 (1.88)	4.71 (1.38)	4.71 (0.79)	4.71 (0.76)
Sample	5,861	5,743	5,743	5,743	5,743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

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

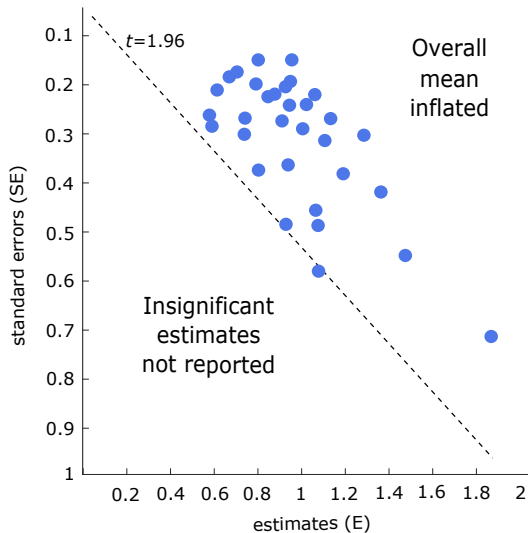
Meta-analysis of the class size effect

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

Funnel plot

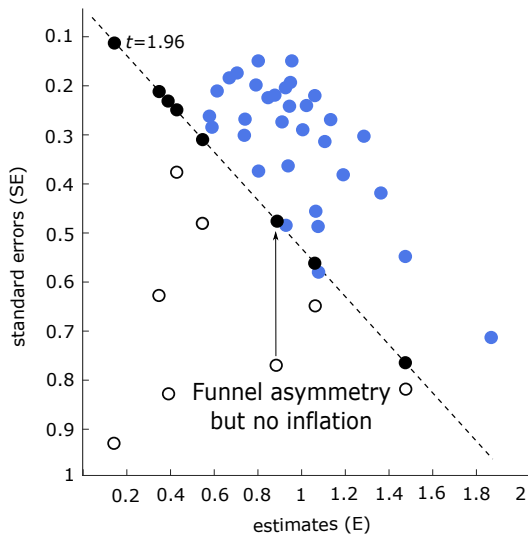


Publication bias





Spurious precision



Key meta assumption broken

PEESE:

$$\hat{E}_i = \textcircled{E_0} + \beta SE(\hat{E}_i)^2 + u_i,$$

Methods or p-hacking

Methods or p-hacking

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

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

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 → MAIVE.

Meta-analysis instrumental variable estimator

MAIVE intuition:

Definition of SE

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

- 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:

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

hacking or misspecifications



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

Meta-analysis instrumental variable estimator

MAIVE adjustment:

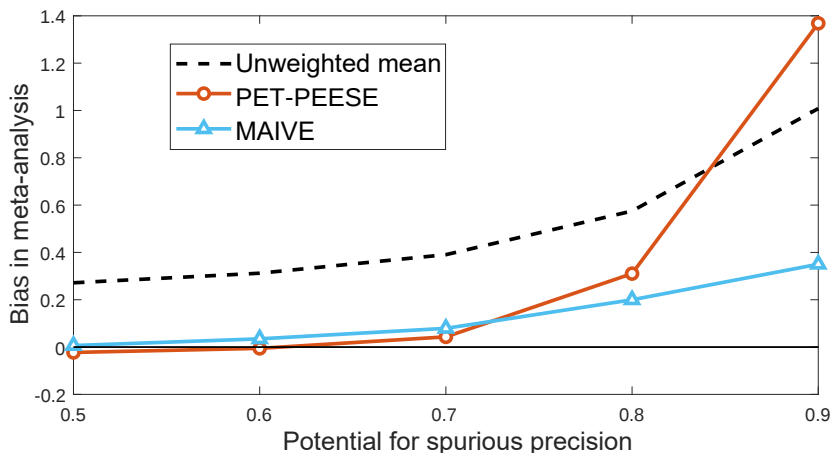
$$SE(\hat{E})_{adj,i}^2 = \hat{\alpha}_0 + \hat{\alpha}_1 (1/N_i) + \cancel{\pi_i}.$$

hacking or misspecifications

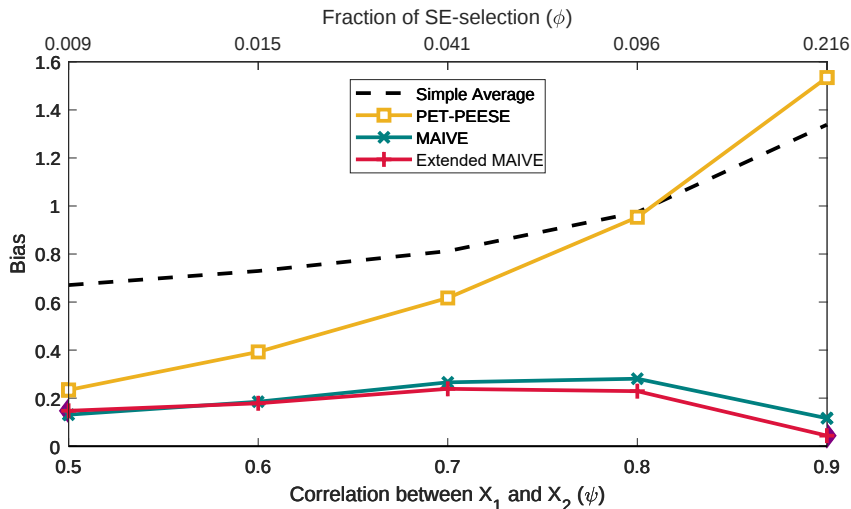


- MAIVE + PEESE = classical IV.
- Can add **controls in the 2nd stage**.
- Plug MAIVE-adjusted SEs into other estimators.

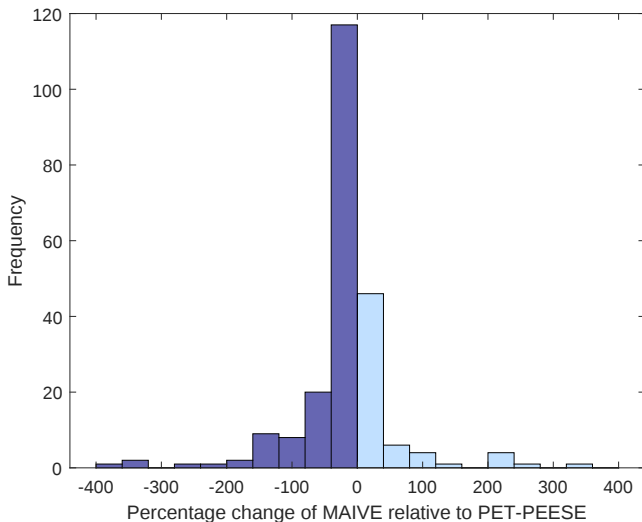
MAIVE alleviates the bias



Extended MAIVE (WAIVE: Weighted Adjustment)



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

Published in *Nature Communications*.

Open Stata

(MAIVE web app at easymeta.org)

RTMA: Maya Mathur's p-hacking correction

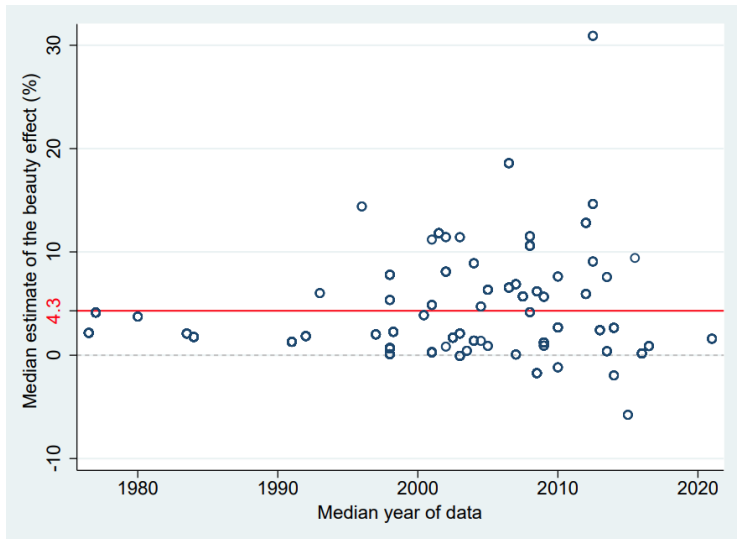
- **Right-truncated meta-analysis** (Mathur 2024, RSM)
- Assumption 1: insignificant estimates are NOT p-hacked
- Assumption 2: estimates are normally distributed
- Assumption 3: there are enough insignificant estimates published that we can recover the underlying distribution
- Bayesian techniques used for computation feasibility (ML would typically need huge samples)
- Digression: Robust Bayesian Meta-Analysis (**RoBMA**)

Open R

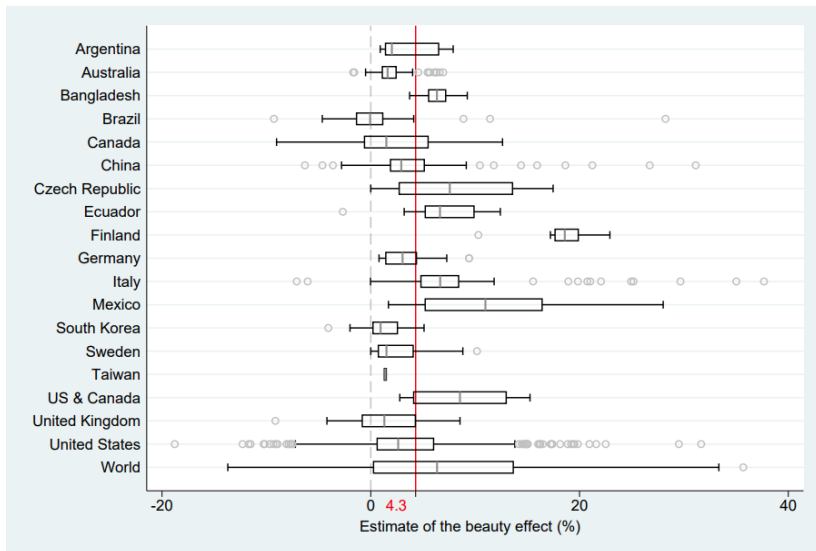
Workshop Outline

- 1 Motivation
- 2 Literature Search
- 3 Data Collection
- 4 Conventional Tools
- 5 Publication Bias
- 6 P-hacking
- 7 Heterogeneity**

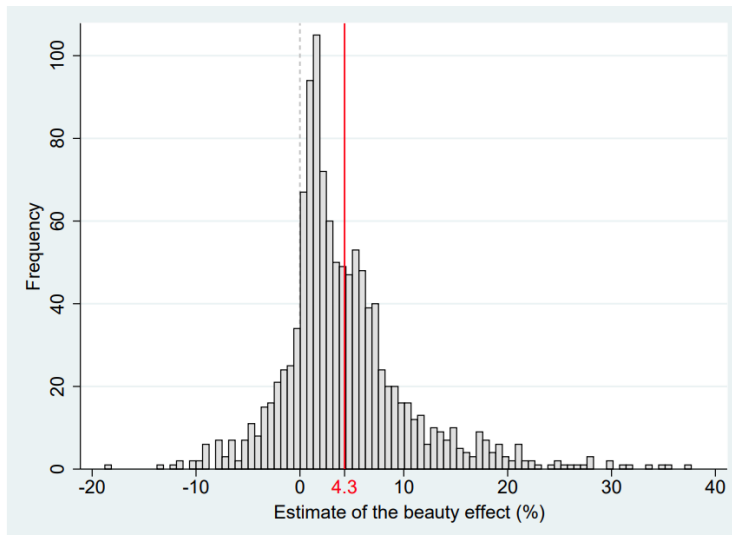
Guiding example: beauty and success



Heterogeneity both within and across



Wide dispersion



Measurement of beauty

			Unweighted		
	Est.	Stud.	Mean	95% conf. int.	
<i>Measurement of beauty</i>					
Interviewer-rated beauty	526	24	4.33	3.80	4.86
Photo-rated beauty	481	37	4.28	3.71	4.85
Software-rated beauty	104	8	5.49	4.15	6.83
Self-rated beauty	56	6	2.04	0.79	3.29
Dummy beauty	460	23	3.63	3.05	4.22
Categorical beauty	699	52	4.70	4.24	5.16
Beauty premium	954	67	4.48	4.07	4.89
Beauty penalty	205	16	3.33	2.56	4.10

Measurement of success

			Unweighted		
	Est.	Stud.	Mean	95% conf. int.	
<i>Measurement of success</i>					
Earnings	688	43	4.33	3.92	4.75
Study outcomes	189	9	3.19	1.98	4.41
Teaching & research outcomes	139	8	4.95	4.05	5.85
Athletic success	33	2	6.28	3.03	9.54
Electoral success	37	4	7.00	4.26	9.75
Other outcomes	73	6	2.98	2.24	3.73

Data characteristics

			Unweighted		
	Est.	Stud.	Mean	95% conf. int.	
<i>Data characteristics</i>					
Male subjects	375	44	3.66	3.10	4.22
Female subjects	447	48	4.10	3.45	4.75
Mixed gender	337	36	5.21	4.57	5.85
High-skilled workers	335	27	4.30	3.78	4.81
Prostitutes	55	4	8.55	7.27	9.82
Other dressy occupations	138	15	4.92	3.89	5.96
Non-dressy occupations	966	51	3.94	3.55	4.34
Western culture	874	47	3.85	3.46	4.23
Other cultures	285	21	5.60	4.72	6.48
Panel data	998	55	3.86	3.47	4.25
Cross-section	161	13	6.87	5.90	7.83

Estimation technique

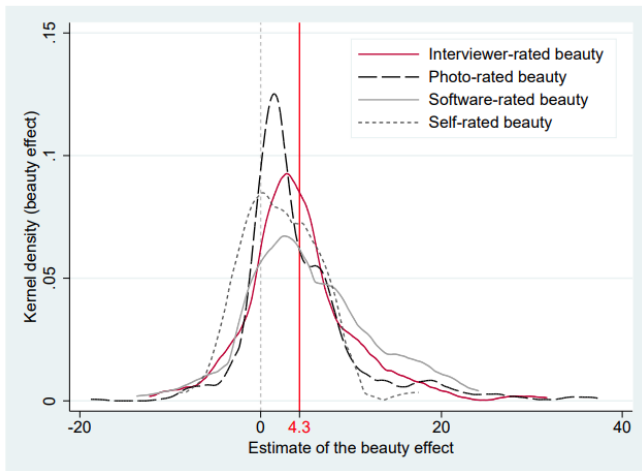
<i>Estimation technique</i>	Est.	Stud.	Unweighted		
			Mean	95% conf. int.	
Ordinary least squares	903	60	4.17	3.77	4.56
Instrumental variables	83	10	4.86	3.07	6.64
Difference-in-differences	12	2	1.24	0.63	1.84
Other method	161	13	4.84	3.78	5.90
Cognitive skill control	445	26	2.77	2.17	3.36
No cognitive skill control	714	53	5.22	4.78	5.66

Publication characteristics

			Unweighted		
	Est.	Stud.	Mean	95% conf. int.	
<i>Publication characteristics</i>					
Published study	1,027	58	4.19	3.80	4.57
Unpublished study	132	9	4.99	3.96	6.03
High-quality peer review	151	7	5.40	4.56	6.25

Beauty measurement doesn't matter

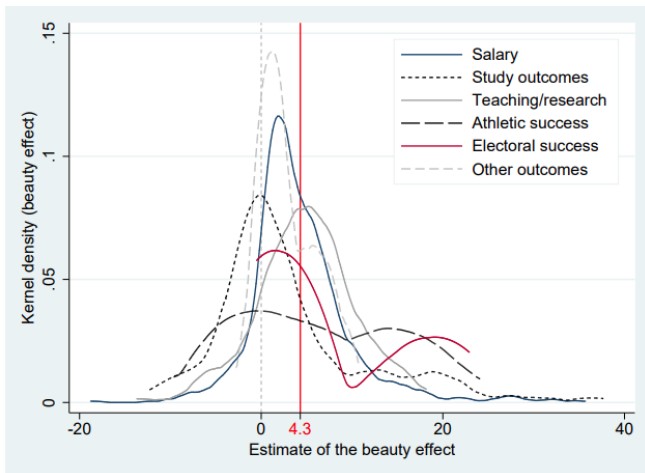
(a) Beauty measurement



Open Stata

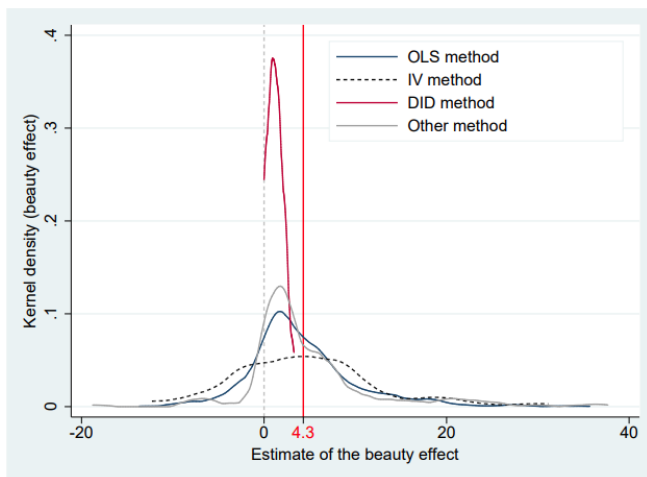
Success measurement doesn't matter

(b) Success measurement



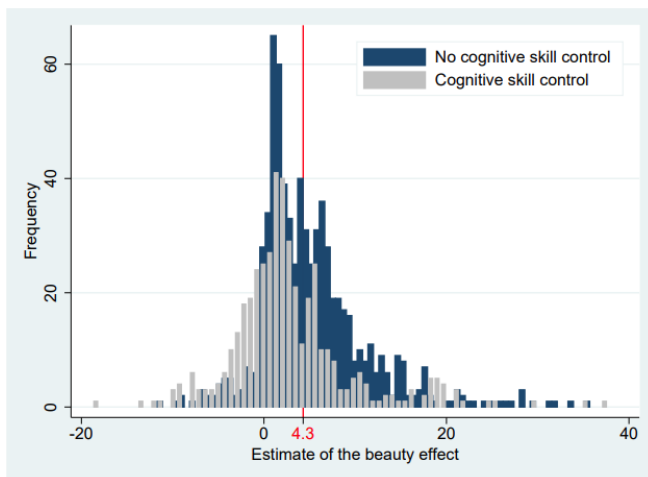
Method doesn't matter

(c) Method choice



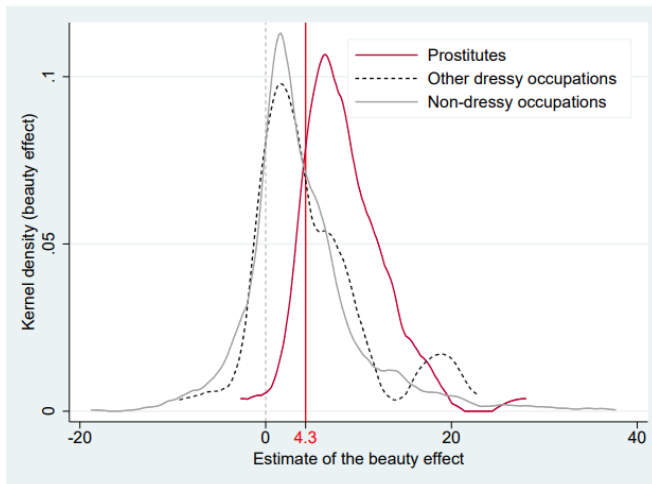
Ability control matters

(g) Ability control

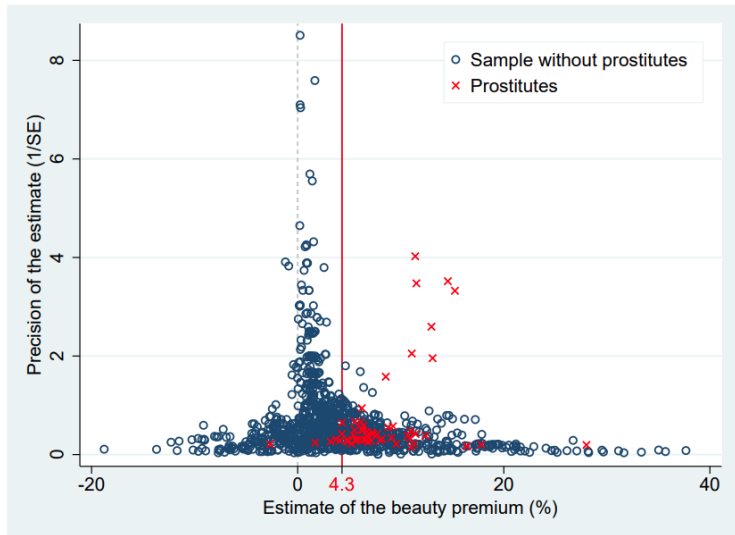


Occupation matters

(d) Occupation



Sex workers stand apart



Option 1: subsamples

Part 1. Subsample of prostitutes					
Panel A	OLS	FE	BE	MAIVE	Weighted
Publication bias (<i>standard error</i>)	0.148 (0.776) [-2.956, 2.671]	1.467 (0.782)	-0.756 (1.032)	-1.567 (0.967) {-7.405, -0.514}	-0.468** (0.210) [-3.545, 0.383]
Effect beyond bias (<i>constant</i>)	8.136*** (2.767) [1.525, 15.17]	4.488 (2.164)	11.260* (2.686)	12.880*** (1.392) {1.110, 17.173 }	11.760*** (0.414) [5.245, 13.97]
First-stage robust F-stat				15.8	
Observations	55	55	55	55	55
Panel B	Precision-weighted	WAAP	Stem	Kink	Selection
Publication bias	-2.154*** (0.745) [-4.057, -0.568]			-2.126 (1.878) [-4.326, -0.755]	P = 1.215 (0.091)
Effect beyond bias	13.290*** (0.298) [-22.03, 16.87]	12.168*** (0.372)	11.205*** (0.905)	12.214*** (0.349) [-22.49, 17.96]	8.490*** (1.691)
Observations	55	55	55	55	55

Option 1: subsamples

Part 2. Sample without prostitutes					
Panel A	OLS	FE	BE	MAIVE	Weighted
Publication bias (<i>standard error</i>)	0.391*** (0.117) [0.120, 0.642]	0.204* (0.119)	0.800*** (0.161)	0.875* (0.450) {0.118, 2.078}	0.718*** (0.271) [0.061, 1.331]
Effect beyond bias (<i>constant</i>)	2.581*** (0.447) [1.660, 3.501]	3.289*** (0.453)	1.603* (0.875)	0.741 (1.728) {0.048, 3.696}	2.617** (1.061) [0.333, 5.113]
First-stage robust F-stat				16.2	
Observations	1,104	1,104	1,104	1,104	1,104
Panel B	Precision-weighted	WAAP	Stem	Kink	Selection
Publication bias	1.610*** (0.202) [1.184, 2.051]			1.610*** (0.202) [1.184, 2.051]	P = 0.307 (0.039)
Effect beyond bias	0.152 (0.118) [-0.050, 0.963]	0.229*** (0.07)	0.008 (0.798)	0.152 (0.118) [-0.050, 0.963]	0.669*** (0.231)
Observations	1,104	1,104	1,104	1,104	1,104

Option 1: subsamples

Part 1. Subsample of penalties

Panel A	OLS	FE	BE	MAIVE	Weighted
Publication bias (<i>standard error</i>)	0.165** (0.0678) [0.021, 0.406]	-0.354 (0.235)	0.702*** (0.168)	0.578 (1.285) {-1.941, 3.096}	0.354*** (0.131) [0.028, 1.395]
Effect beyond bias (<i>constant</i>)	2.647*** (0.733) [1.225, 4.378]	4.801*** (0.977)	0.447 (0.877)	0.933 (5.565) {-3.133, 4.998}	2.655** (1.153) [0.038, 5.782]
First-stage F-stat				0.8	
Observations	205	205	205	205	205
Panel B	Precision-weighted	WAAP	Stem	Kink	Selection
Publication bias	1.097*** (0.164) [0.709, 1.489]			1.097*** (0.12) [0.709, 1.489]	P = 0.369 (0.019)
Effect beyond bias	0.139*** (0.0531) [-0.405, 1.425]	0.244*** (0.021)	0.245 (0.323)	0.139 (0.097) [-0.405, 1.425]	0.236*** (0.067)
Observations	205	205	205	205	205

Option 1: subsamples

Part 2. Sample without penalties

Panel A	OLS	FE	BE	MAIVE	Weighted
Publication bias (<i>standard error</i>)	0.437*** (0.139) [0.125, 0.745]	0.282* (0.160)	0.745*** (0.172)	0.772* (0.440) {0.032, 1.949}	0.666** (0.281) [-0.044, 1.295]
Effect beyond bias (<i>constant</i>)	2.881*** (0.555) [1.745, 4.058]	3.448*** (0.585)	2.228** (0.914)	1.652 (1.580) {0.068, 4.171}	3.470*** (1.202) [0.845, 6.305]
First-stage robust F-stat				17.3	
Observations	954	954	954	954	954
Panel B	Precision-weighted	WAAP	Stem	Kink	Selection
Publication bias	1.881*** (0.246) [1.351, 2.378]			1.881*** (0.246) [1.351, 2.378]	P = 0.300 (0.037)
Effect beyond bias	0.346 (0.294) [-0.051, 2.285]	0.38* (0.229)	0.013 (1.323)	0.346 (0.294) [-0.051, 2.285]	0.200 (0.828)
Observations	954	954	954	954	954

Option 2: Bayesian model averaging

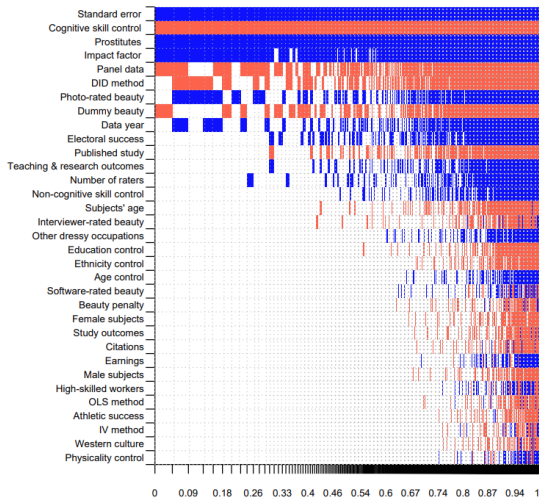
- We want to regress estimates of the beauty effect on variables reflecting heterogeneity
- But too many variables; model uncertainty
- Solution: run regressions with different combinations of the variables
- Give more weight to those that fit the data well and are parsimonious

Option 2: Bayesian model averaging

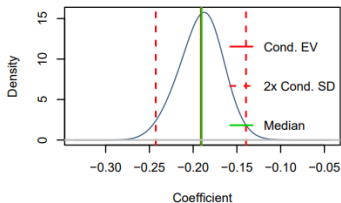
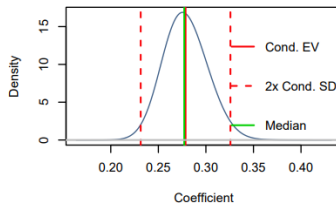
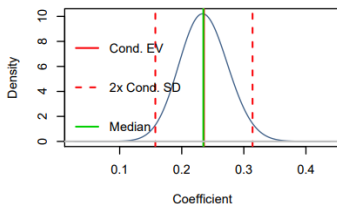
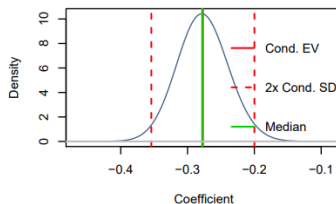
- Model averaging can be frequentist but computationally difficult
- Intuition: run different combinations and weight by adjusted R^2 or information criteria
- In the Bayesian setting we can use the MCMC algorithm
- **G-prior:** all coefficients are zero. Weight? UIP = the prior has the same weight as one observation
- **Model prior:** uniform = all models have the same weight
- Dilution prior: models with collinearity get less weight

Open R

Model inclusion in BMA



Posterior densities

Marginal Density: Industry data (PIP 100 %)**Marginal Density: FOC_L_w (PIP 100 %)****Marginal Density: Linear approx. (PIP 100 %)****Marginal Density: Normalized (PIP 100 %)**

Regression results

Table 5: Why reported beauty premiums vary

Response variable: Beauty premium	Bayesian model averaging (baseline model)			Stepwise regression (frequentist check)		
	P. mean	P. SD	PIP	Mean	SE	p-value
Constant	2.759	NA	1.000	3.582	0.562	0.000
Standard error	0.435	0.042	1.000	0.426	0.112	0.000
<i>Measurement of beauty</i>						
Interviewer-rated beauty	-0.033	0.204	0.036			
Photo-rated beauty	0.591	0.745	0.424			
Software-rated beauty	0.009	0.149	0.014			
Dummy beauty	-0.485	0.664	0.387			
Beauty penalty	-0.007	0.089	0.012			
Number of raters	0.051	0.140	0.134			
<i>Measurement of success</i>						
Earnings	0.006	0.098	0.010			
Study outcomes	-0.006	0.114	0.011			
Teaching & research	0.238	0.645	0.141			
Athletic success	-0.004	0.104	0.007			
Electoral success	0.670	1.361	0.224			
<i>Data characteristics</i>						
Male subjects	-0.005	0.064	0.010			
Female subjects	-0.006	0.069	0.012			
Subjects' age	-0.075	0.332	0.061			
High-skilled workers	0.002	0.064	0.009			
Prostitutes	4.793	1.058	0.999	4.386	1.199	0.001
Other dressy occupations	0.035	0.227	0.032			
Western culture	-0.001	0.039	0.006			
Panel data	-1.131	1.016	0.606			
Data year	0.337	0.504	0.346			

Best practice (implied estimate)

- From BMA results we know how estimates depend on study design
- Some study designs are better: identification, data, generally quality
- Select best practice based on the literature (quasi-experimental studies, etc.)
- **Compute fitted values** and confidence intervals (lincom command in Stata)

Open Stata

Best practice (class size effect)

	Effect	95% cred. int.	
STAR experiment	-1.50	-3.04	0.04
Regression discontinuity	-0.18	-1.36	1.00
Instrumental variable	-0.42	-1.59	0.75
Fixed effects	0.01	-1.17	1.19
OLS	0.01	-1.19	1.21
Kindergarten	0.18	-1.10	1.46
Primary school	-0.46	-1.65	0.72
Secondary school	0.18	-0.99	1.35
Disadvantaged students	-0.19	-1.42	1.03
USA	-0.21	-1.30	0.88
Scandinavian countries	-0.21	-1.47	1.05
Other countries	-0.21	-1.43	1.02
Math	-0.21	-1.38	0.97
Reading	-0.21	-1.39	0.96
Writing	-0.20	-1.50	1.11
Languages	-0.21	-1.40	0.99
Other subject	-0.21	-1.43	1.01
Class size = 15	-0.99	-2.18	0.21
Class size = 20	-0.45	-1.62	0.72
Class size = 25	-0.18	-1.35	1.00
Class size = 30	0.01	-1.18	1.19
Class size = 35	0.15	-1.05	1.34
Overall corrected mean	-0.21	-1.38	0.97

Summary: How to do a modern meta-analysis

- 1 Search for primary studies using Google Scholar, snowballing, AI agent
- 2 Use comparable estimates (elasticities, mean differences, correlations)
- 3 Code standard errors, sample sizes, main differences in data and methods
- 4 Winsorize estimates and standard errors at the 1% level
- 5 Correct for publication bias and p-hacking (MAIVE; if weak instrument $\rightarrow$ RTMA)
- 6 Examine heterogeneity using BMA with the dilution prior
- 7 Report implied effects for different contexts

Thank you for inviting me

Prepared for the University of Osaka. Published at meta-analysis.cz/teaching under CC BY 4.0.