Lecture 7: Publication Bias

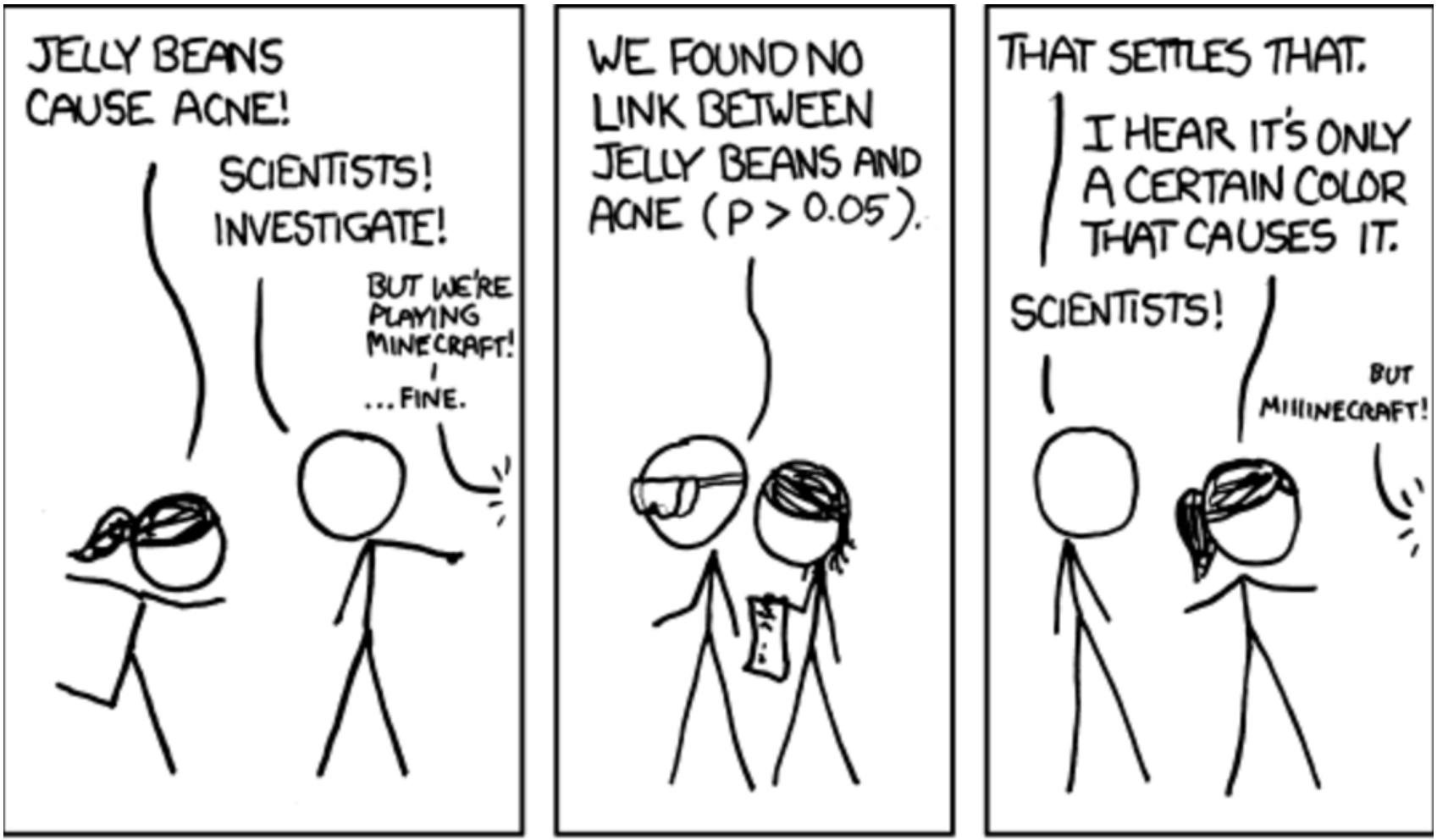
Tomas Havranek

Charles University, CEPR, METRICS

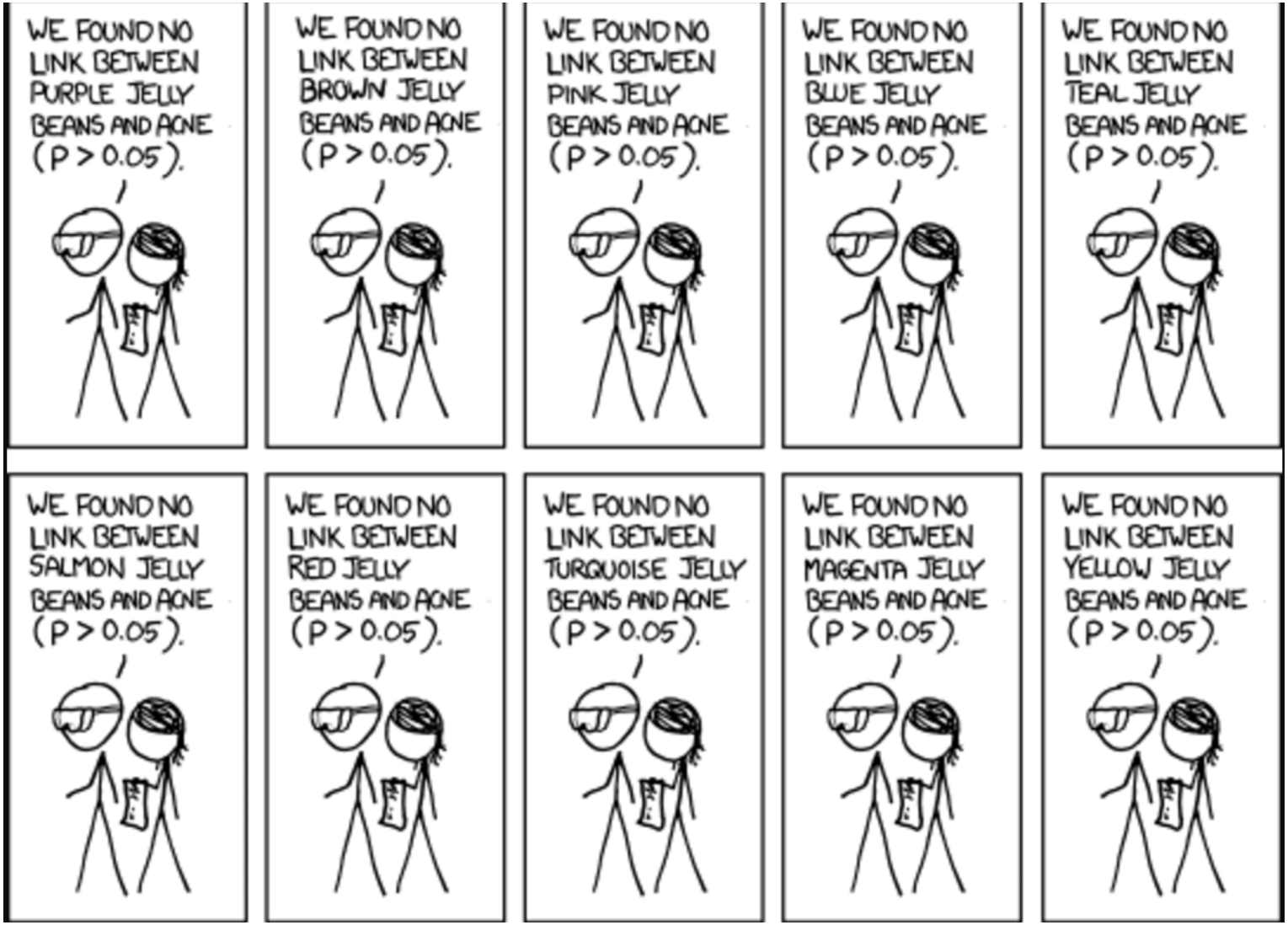
Research Synthesis in Economics and Finance,
University of Canterbury

March 10, 2025

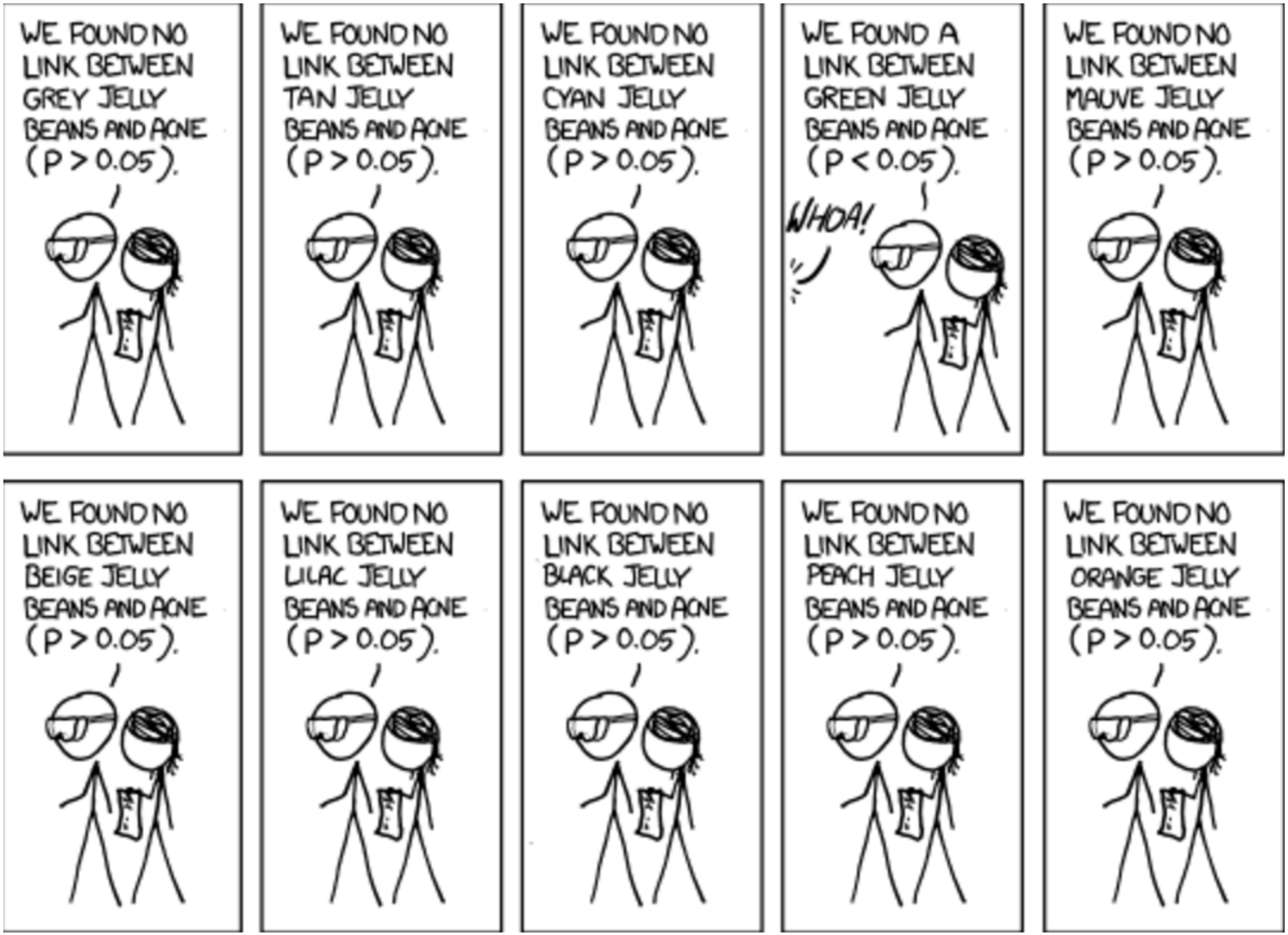
Intuition


JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
...FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIIINECRAFT!

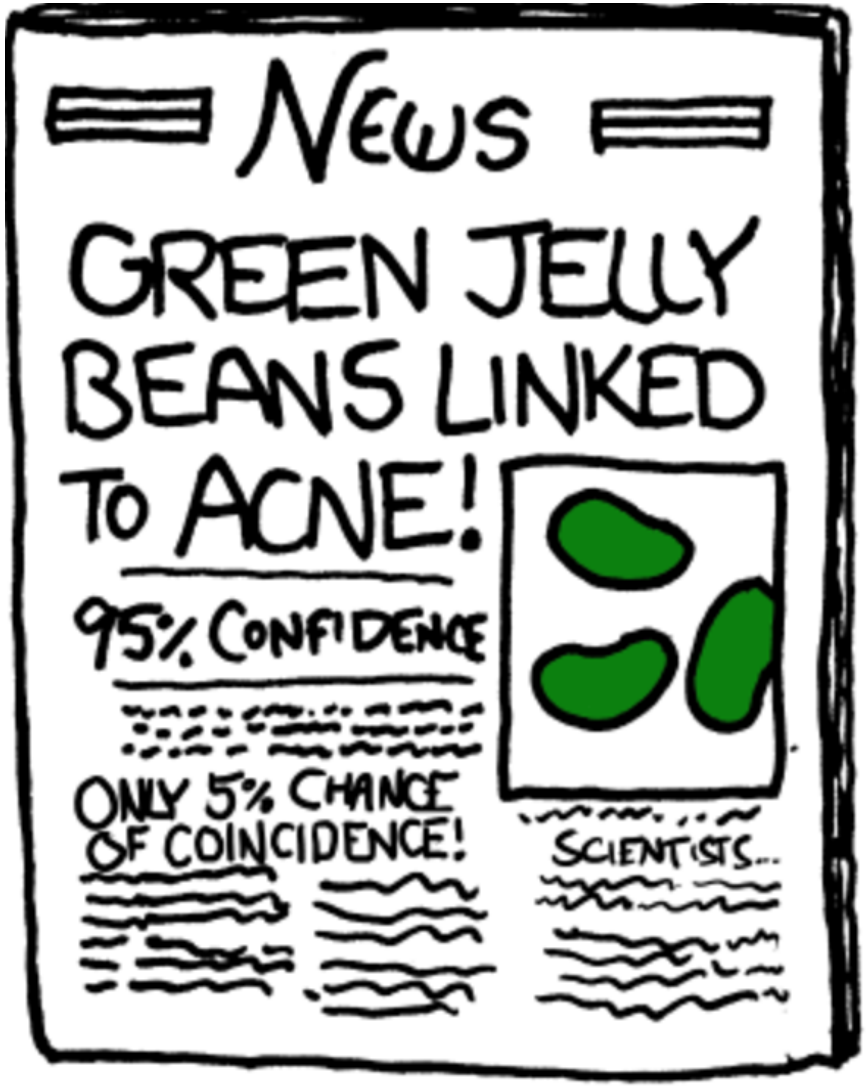
Intuition


WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).

Intuition


WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).

Intuition


NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

Paxil scandal

- **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.

Paxil scandal

- **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 correlation (?)

Researchers estimate

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

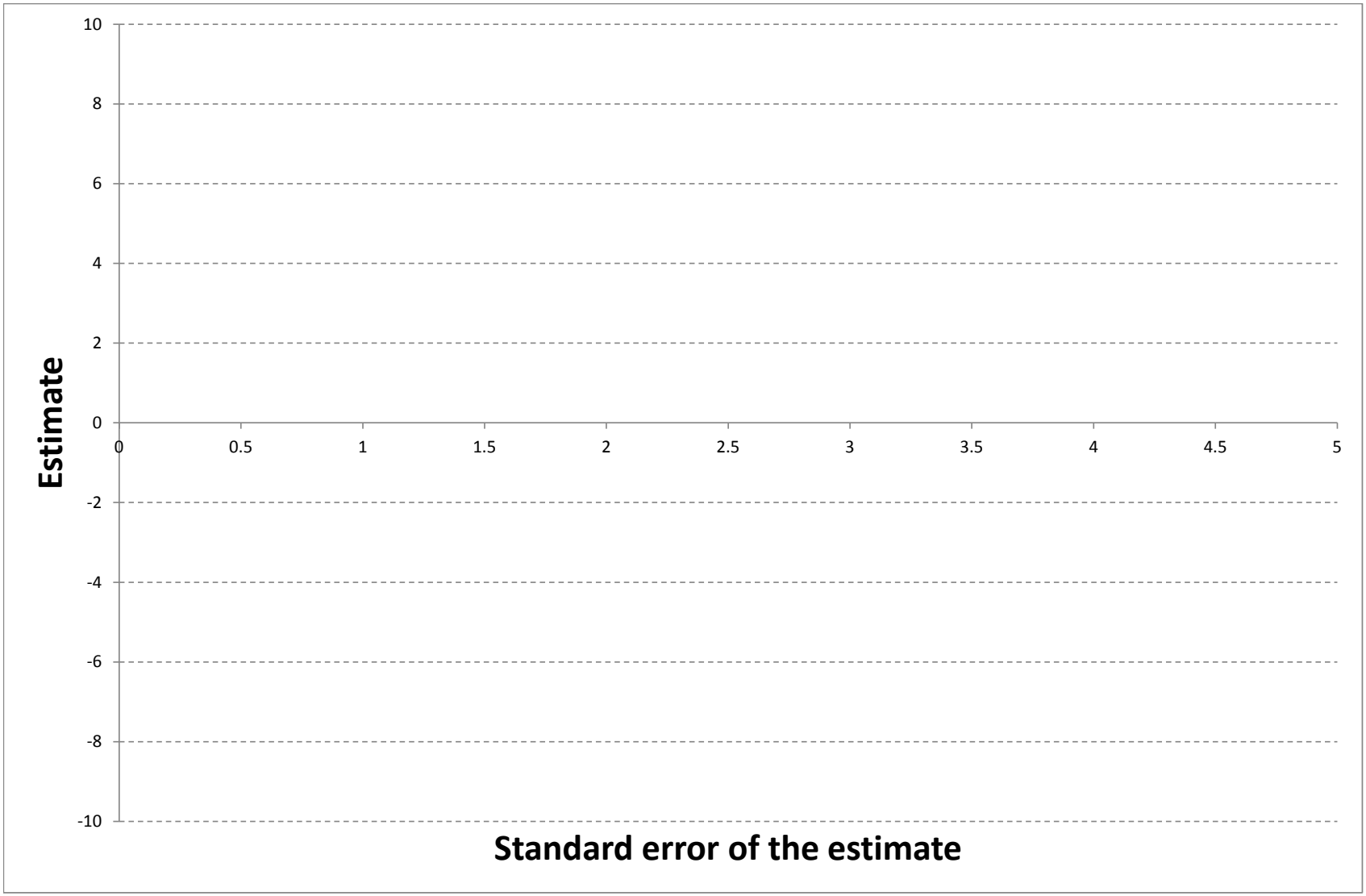
They assume that $\widehat{\gamma}/SE_{\widehat{\gamma}}$ has a t-distribution.

$\rightarrow$ estimates should not be correlated with standard errors (Card & Krueger, 1995)

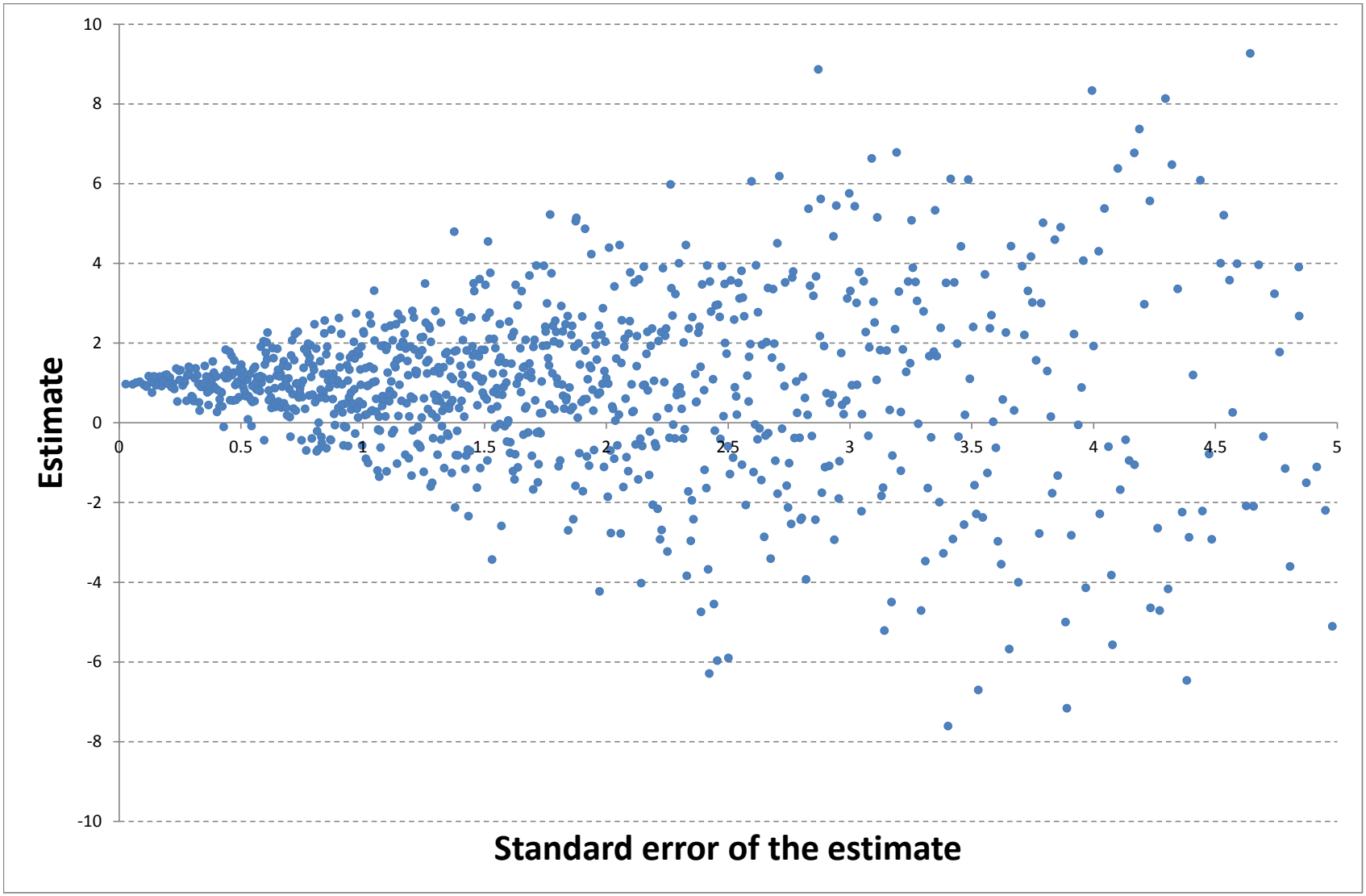
Types of publication bias

Type I: Selection for the "correct sign."
Type II: Selection for statistical significance.

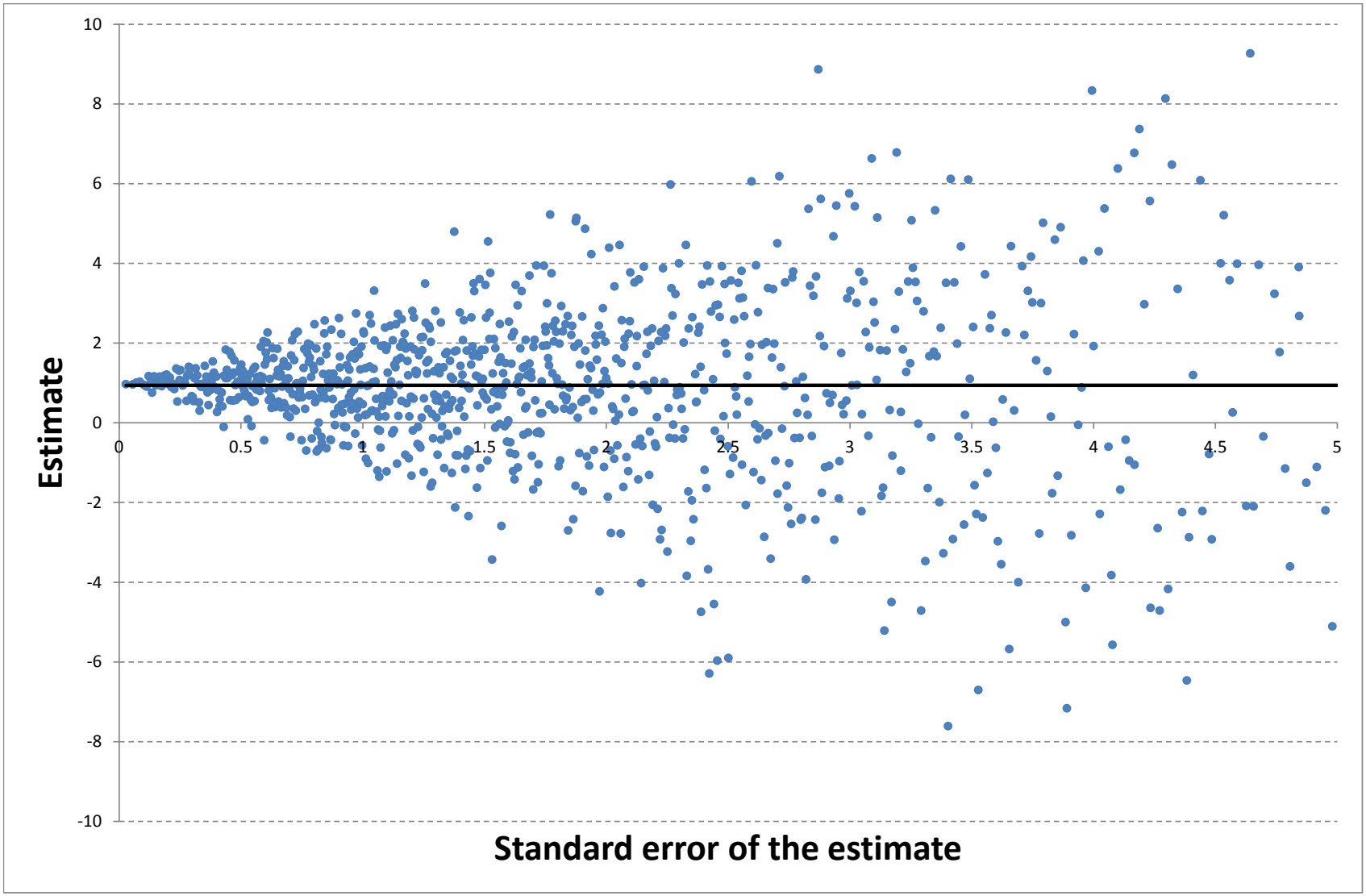
Simulation


10
8
6
4
2
0
-2
-4
-6
-8
-10
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Estimate
Standard error of the estimate

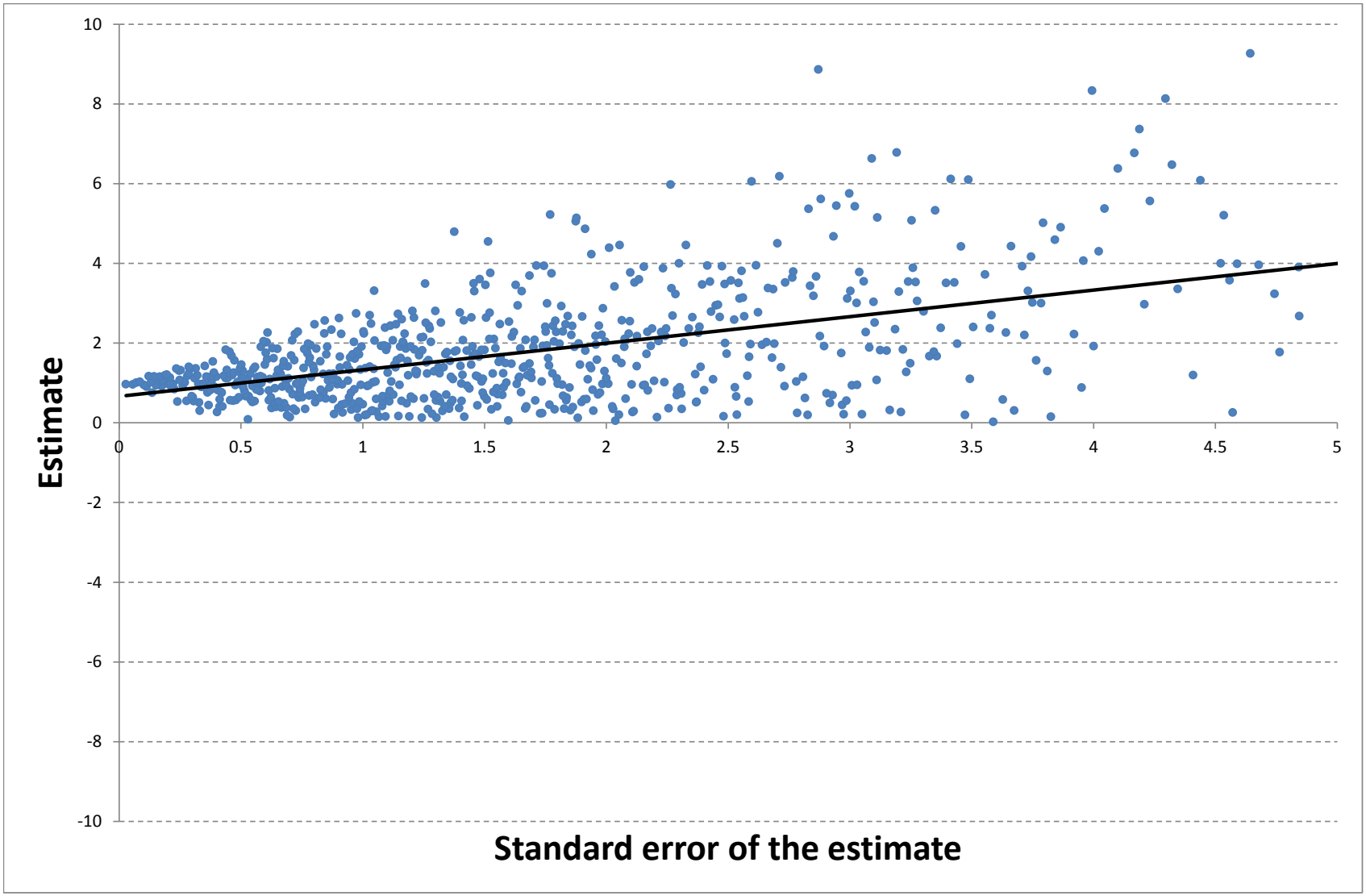
Simulation


Estimate
Standard error of the estimate
10
8
6
4
2
0
-2
-4
-6
-8
-10
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5

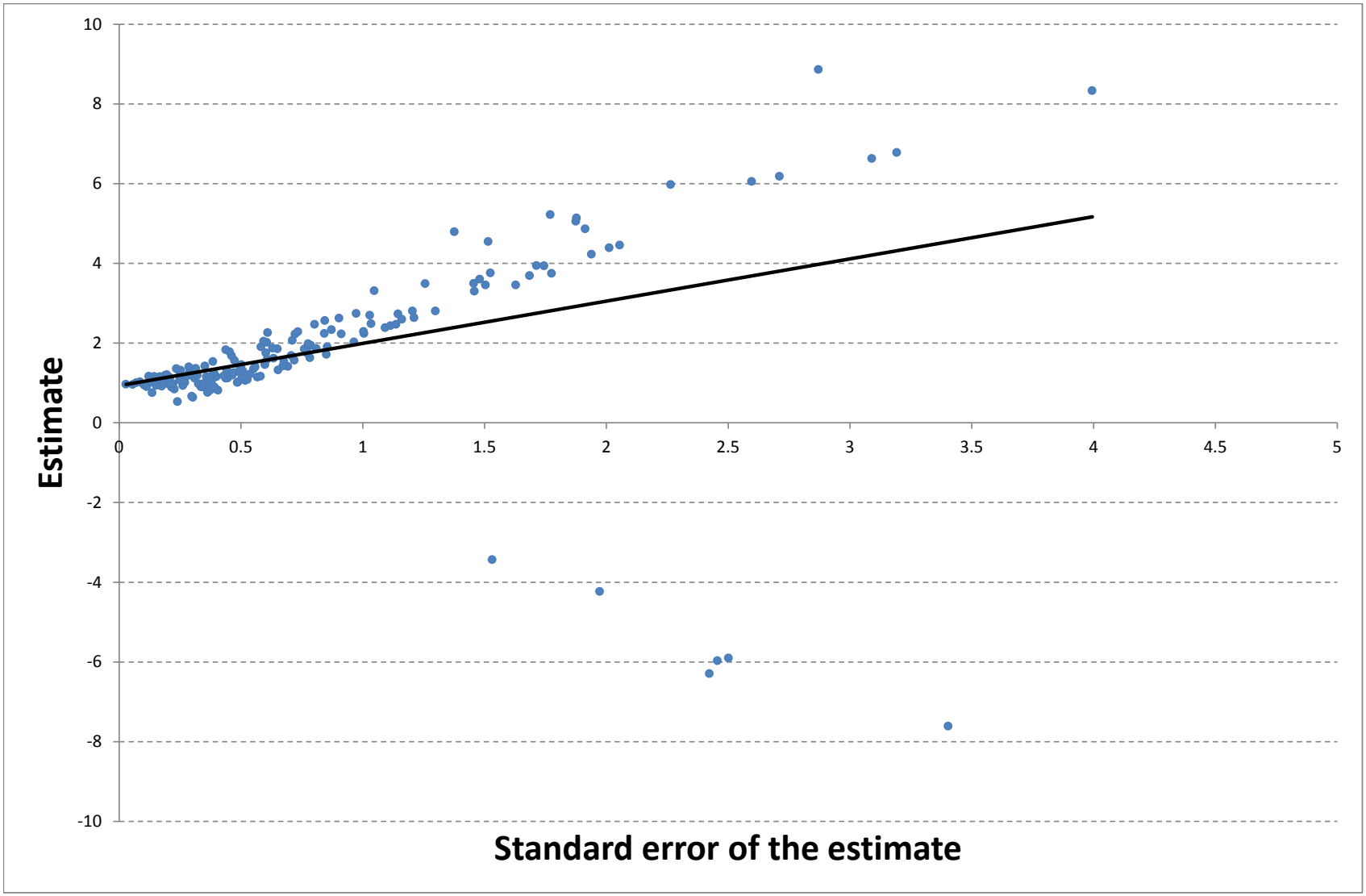
Simulation


Estimate
Standard error of the estimate
10
8
6
4
2
0
-2
-4
-6
-8
-10
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5

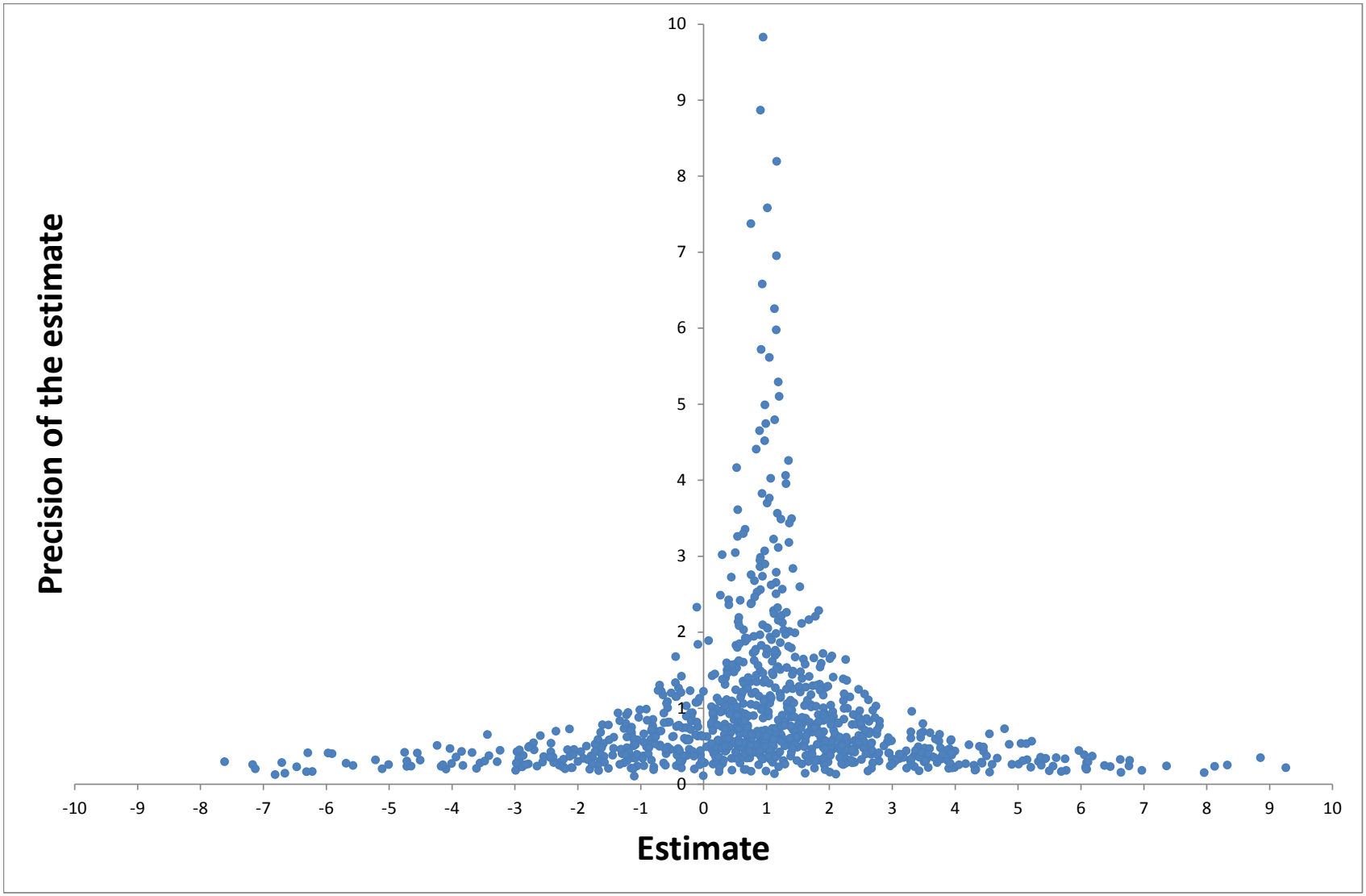
Simulation


10
8
6
4
2
0
-2
-4
-6
-8
-10
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Estimate
Standard error of the estimate

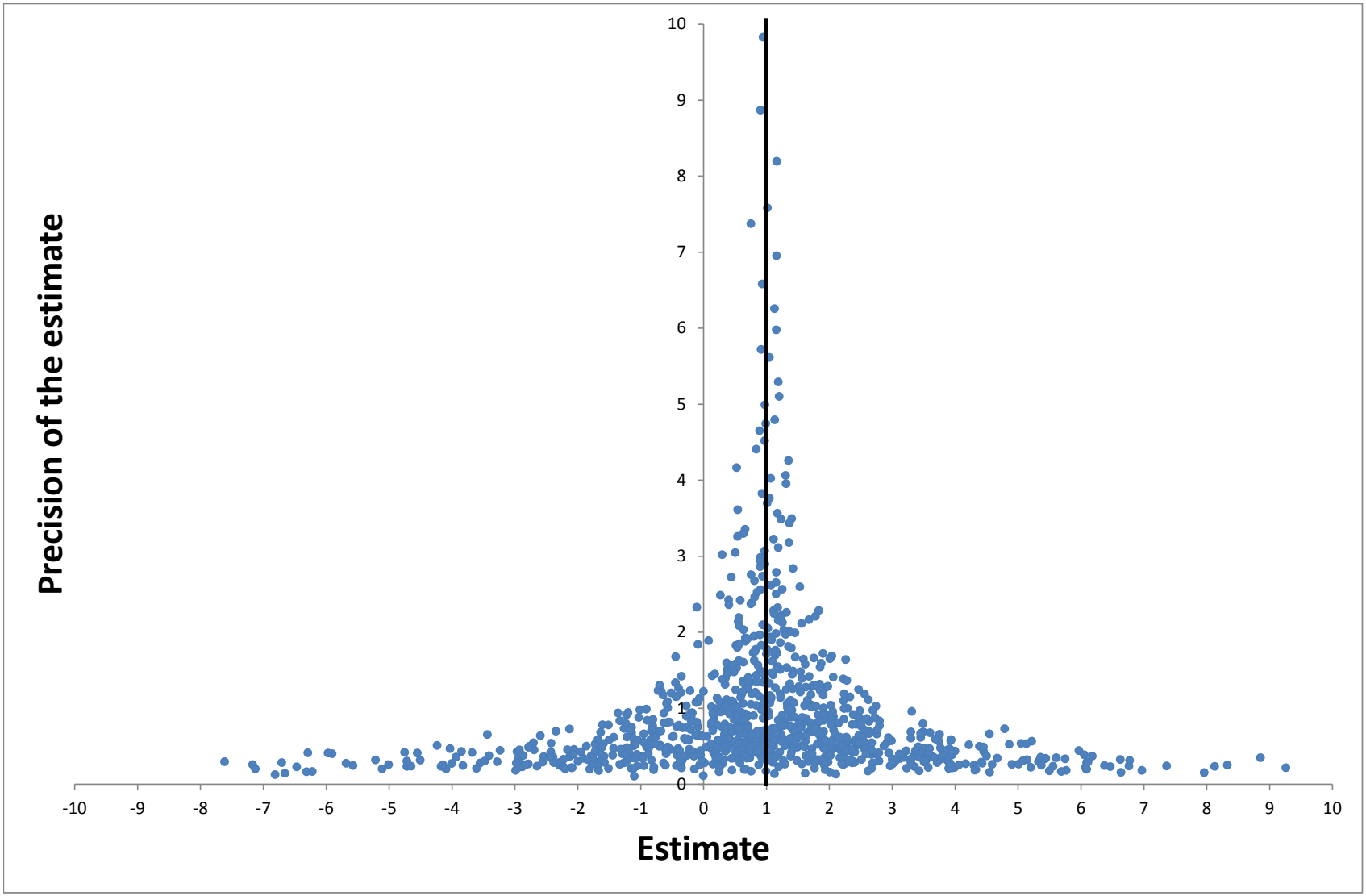
Simulation


Estimate
Standard error of the estimate
10
8
6
4
2
0
-2
-4
-6
-8
-10
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5

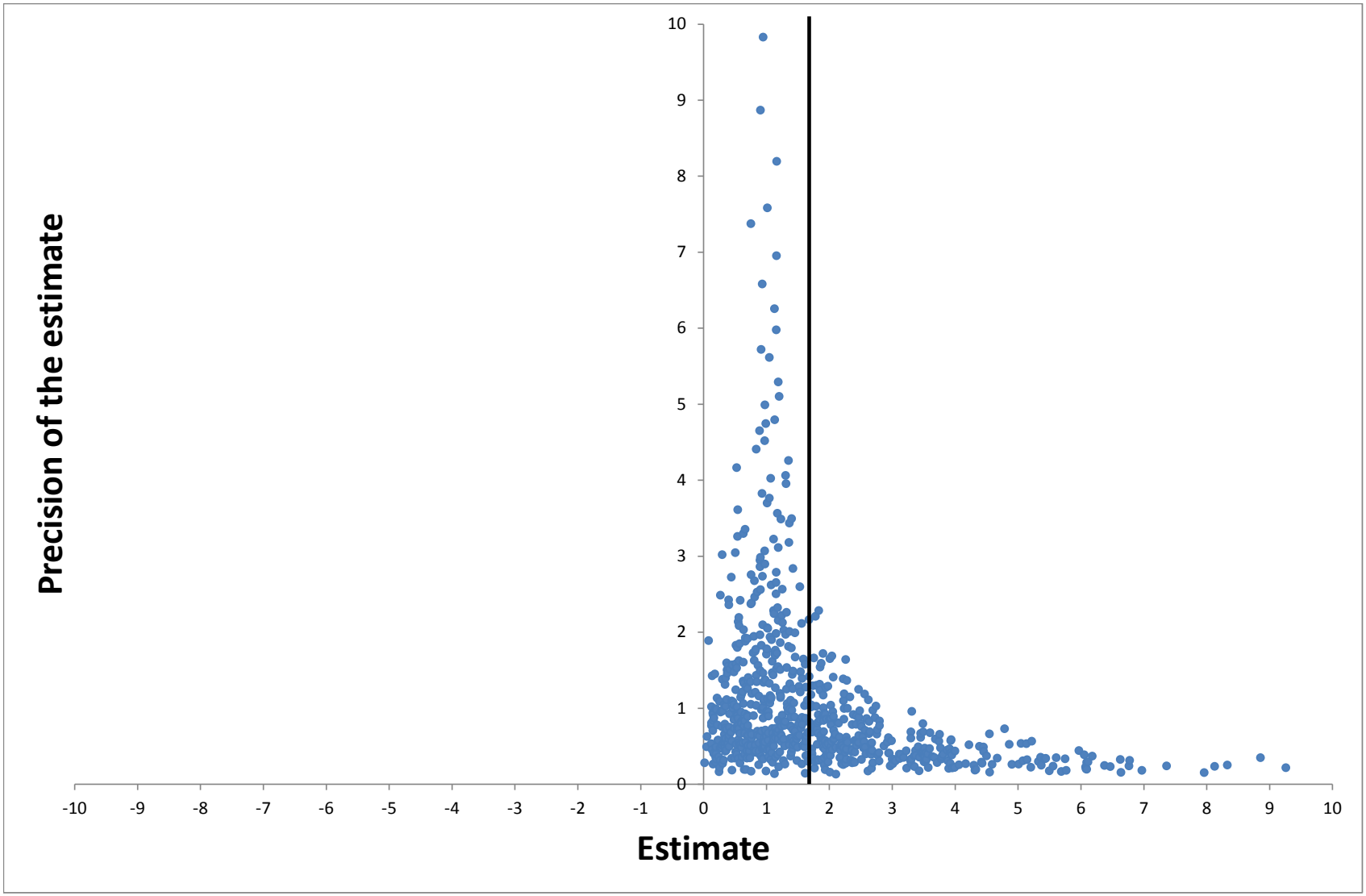
Simulation


Precision of the estimate
10
9
8
7
6
5
4
3
2
1
0
-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10
Estimate

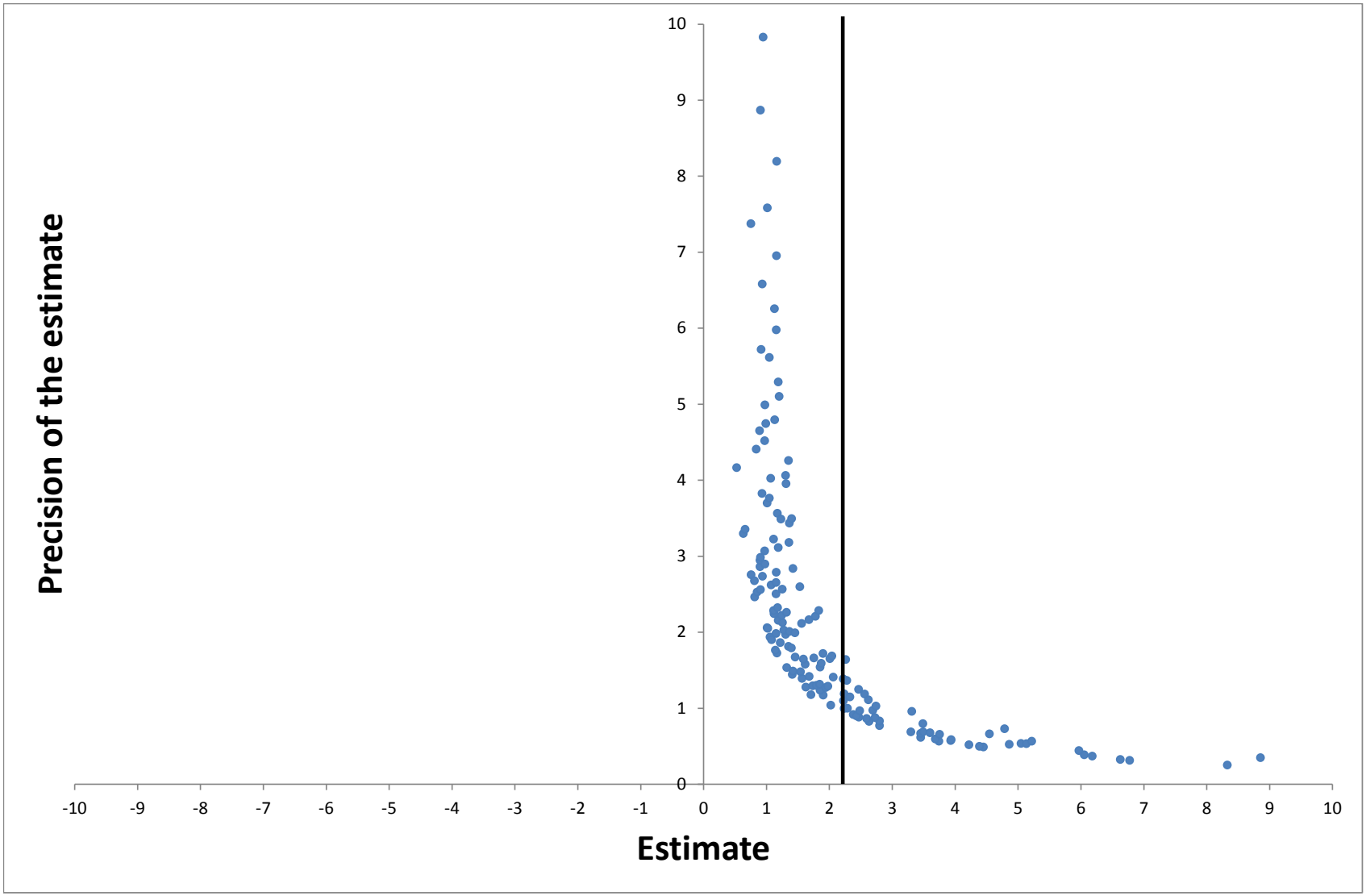
Simulation


Precision of the estimate
10
9
8
7
6
5
4
3
2
1
0
-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10
Estimate

Simulation


Precision of the estimate
10
9
8
7
6
5
4
3
2
1
0
-10
-9
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
9
10
Estimate

Simulation


Precision of the estimate
Estimate
10
9
8
7
6
5
4
3
2
1
0
-10
-9
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
9
10

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).

Precision effect test with standard error (PEESE):

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

The equation is heteroscedastic. Weighted least squares yield

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

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.

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.

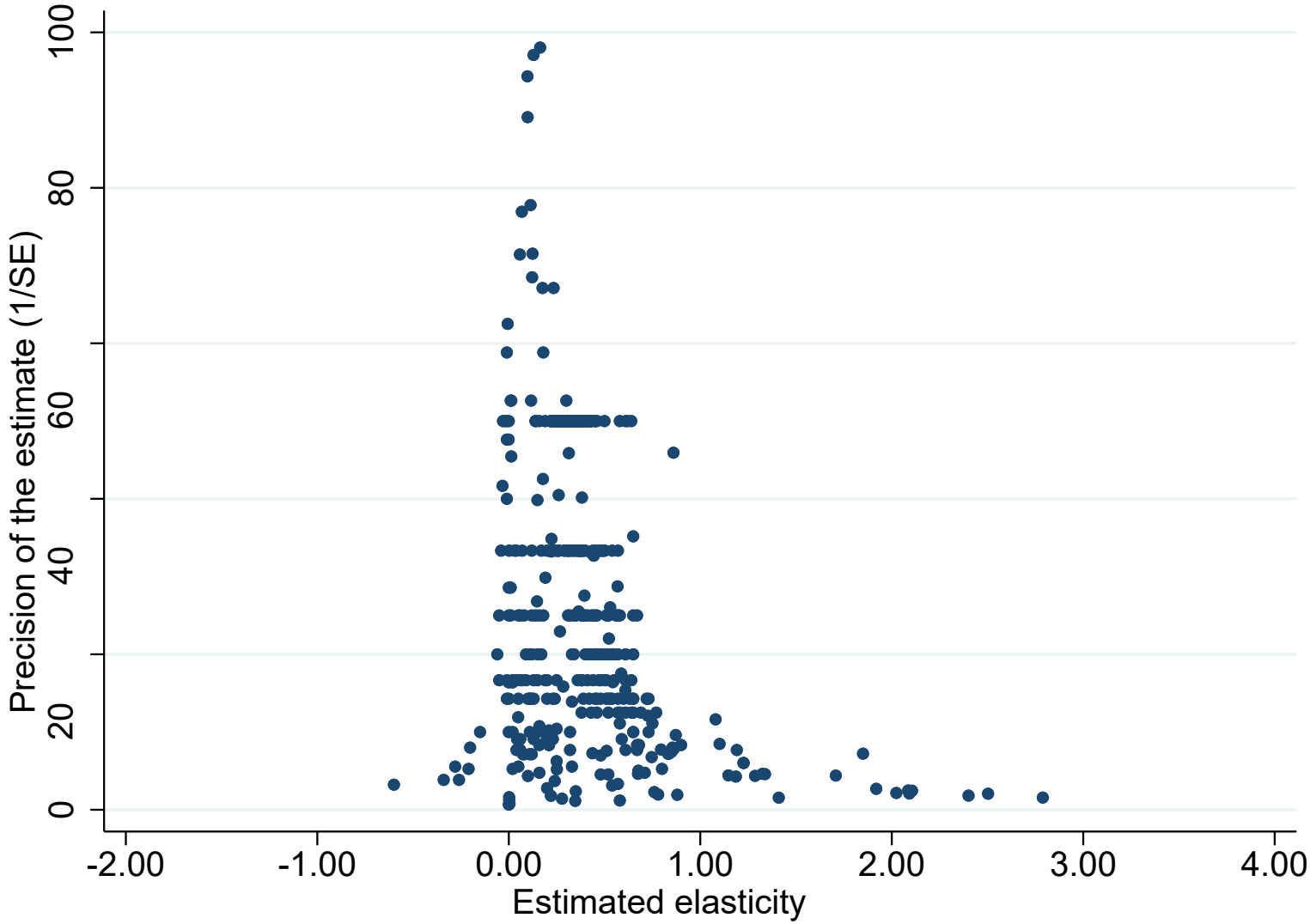
Key features:

- 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.

Example: labor supply elasticity



Results vary: RoBMA?

Panel A: Linear tests					
	OLS	FE	Precision	Study	IV
Standard error	1.595***	0.788***	2.222***	2.106***	2.706
(*publication bias*)	(0.262)	(0.283)	(0.484)	(0.258)	(2.103)
	[0.93, 2.20]	-	[1.23, 3.31]	[1.53, 2.56]	-
Constant	0.301***	0.370***	0.247***	0.252***	0.130*
(*mean beyond bias*)	(0.044)	(0.026)	(0.065)	(0.109)	(0.066)
	[0.12, 0.42]	-	[0.11, 0.31]	[0.13, 0.38]	-
Observations	723	723	723	723	522
Studies	36	36	36	36	23
Panel B: Nonlinear tests					
	Ioannidis *et al.* (2017)	Andrews & Kasy (2019)	Bom & Rachinger (2019)	Furukawa (2021)	van Aert & van Assen (2021)
Effect beyond bias	0.260***	0.356***	0.207**	0.187*	0.374***
	(0.041)	(0.061)	(0.094)	(0.112)	(0.020)
Observations	723	723	723	723	723
Studies	36	36	36	36	36

Bias caused by preference for positive estimates