Which findings get published?
Which findings should be published?

Maximilian Kasy

September 4, 2018

Introduction

- Replicability is a fundamental requirement of science.
  Different researchers should reach the same conclusions.
  Methodological conventions should ensure this.
- Replications of published experiments frequently find effects which are of smaller magnitude or opposite sign.
- One explanation: Selective publication based on findings.
  1. Publication bias
     - Journal editor and referee decisions.
     - Statistical significance, surprisingness, or confirmation of prior beliefs.
  2. P-hacking and specification searching
     - Researcher decisions.
     - Incentives to select which findings to submit based on the likelihood of publication.

Two questions

1. <u>Which findings get published?</u>
   - How much and based on what criteria are findings selected?
   - How can we correct for such selection?
   - Existing approaches test whether publication is selective, but do not estimate the amount and form of selection.
2. <u>Which findings should be published?</u>
   - Replicability is not the only goal of research.
   - Relevance for policy (and other) decisions is important, as well.
   - These two goals might potentially stand in conflict.
   - Existing reform proposals focus on replicability and aim to eliminate selection, ignoring the role of relevance.

Andrews, I. and Kasy, M. (2018). Identification of and correction for publication bias

Frankel, A. and Kasy, M. (2018). Which findings should be published?

Roadmap

- I **Which findings get published?**
  1. Setup and bias-corrected inference
  2. Identification
     - a Replication studies
     - b Meta-studies
  3. Application: Lab experiments in economics
- II Which findings should be published?
  1. Setup and optimal publication rules
  2. Selective publication and statistical inference.
  3. Extensions
     - a Dynamic model
     - b Researcher incentives

  Conclusion

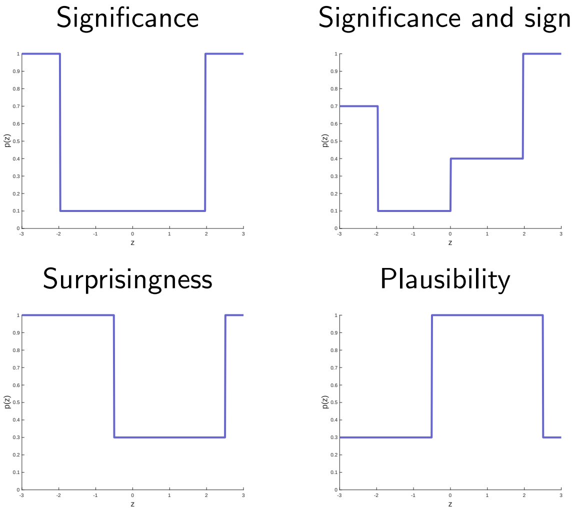
Part I: Which findings get published?

Key results

1. If form and magnitude of selection are known,
   we **can correct published findings**.
   - Unbiased estimates, confidence sets that control size.
   - Using "quantile inversion."
2. Form and magnitude of **selection** are **nonparametrically identified**.
   - Using systematic replication studies.
   - Using meta-studies.
3. **Published research is selected**:
   - Lab experiments in economics and psychology:
     Statistical significance
   - Effect of minimum wages on employment:
     Statistical significance, sign.
   - Deworming:
     Inconclusive.

Setup

Examples: Possible forms of selection $p(Z)$



- $p(Z)$: Probability that an estimate Z is published.

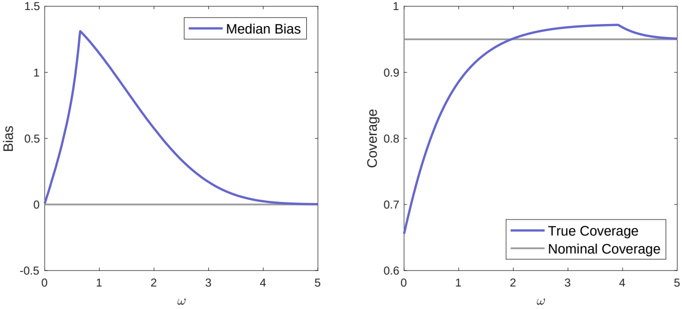
Setup

Assumptions and notation

Latent studies	
True parameter value	Θ^*
Standard error	σ^*
Distribution across studies	$(\Theta^*, \sigma^*) \sim \pi_{\Theta,\sigma}$
Reported estimate	
Distribution	$X^* \vert \Theta^*, \sigma^* \sim N(\Theta^*, \sigma^{*2})$
z-statistic, normalized parameter	$Z^* = \frac{X^*}{\sigma^*}$, $\Omega^* = \frac{\Theta^*}{\sigma^*}$
Publication decision	
Publication probability	$p(Z)$
Publication event	$D \vert X^*, \Theta^*, \sigma^* \sim Ber(p(Z^*))$
Observed sample	i.i.d. given $D = 1$
Observed variables	(X, σ)

Publication bias

Example: Selection on significance



- Publication probability: "significance testing,"

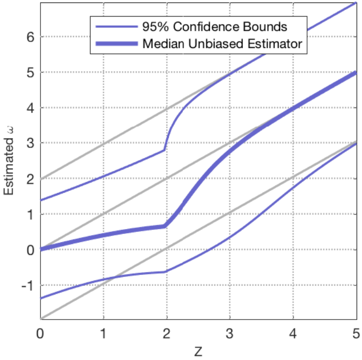
$$p(z) = \begin{cases} 0.1 & |z| < 1.96 \\ 1 & |z| \geq 1.96 \end{cases}$$

left median bias of $\hat{\omega} = Z$

right true coverage of conventional 95% confidence interval

Bias-corrected inference

Example: selection on significance



- If we know $p(\cdot)$, can we correct for bias and size distortions?
- Publication probability: "significance testing,"

$$p(z) = \begin{cases} 0.1 & |z| < 1.96 \\ 1 & |z| \geq 1.96 \end{cases}$$

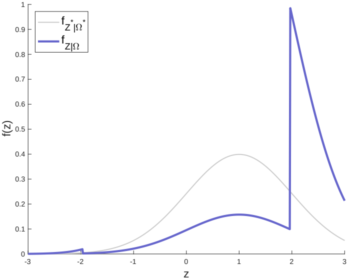
Bias-corrected inference

Density of published estimates

- How was this figure constructed?
- Density of published Z given Ω:

$$f_{Z|\Omega,\sigma}(z|\omega,\sigma) = f_{Z^*|\Omega^*,\sigma^*,D}(z|\omega,\sigma,1)$$
$$= \frac{p(z)}{E\left[p(Z^*)|\Omega^* = \omega\right]}\varphi(z-\omega).$$

- Example: Selection on significance.



- Corresponding cumulative distribution function: $F_{Z|\Omega}(z|\omega)$

Bias-corrected inference

Corrected frequentist estimators and confidence sets

- Define $\hat{\omega}_\alpha(z)$ via

$$F_{Z|\Omega}(z|\hat{\omega}_\alpha(z)) = \alpha.$$

- This definition implies that

$$P(\hat{\omega}_\alpha(Z) \leq \omega | \Omega = \omega) = \alpha \quad \forall \omega.$$

- Median-unbiased estimator: $\hat{\omega}_{\frac{1}{2}}(Z)$ for ω.
- Equal-tailed level $1-\alpha$ confidence interval:

$$\left[\hat{\omega}_{\frac{\alpha}{2}}(Z), \hat{\omega}_{1-\frac{\alpha}{2}}(Z)\right]$$

Roadmap

I Which findings get published?
  1 Setup and bias-corrected inference
  2 **Identification**
    a Replication studies
    b Meta-studies
  3 Application: Lab experiments in economics

II Which findings should be published?
  1 Setup and optimal publication rules
  2 Selective publication and statistical inference
  3 Extensions
    a Dynamic model.
    b Researcher incentives

Conclusion

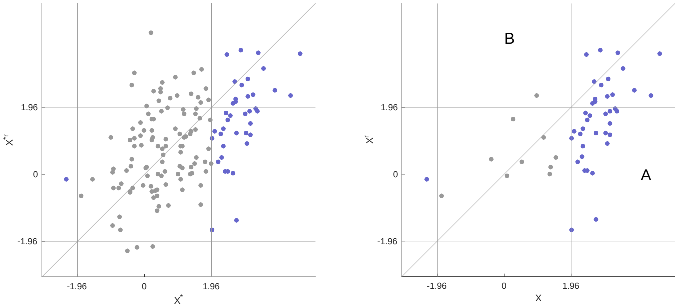
Identification

Identification of the selection mechanism $p(\cdot)$

- We propose two approaches for identification of $p(\cdot)$:
  1. Systematic replication experiments:
     - Replication estimates for the same parameters.
     - Selectivity operates only on original estimate, but not on replication estimate.
  2. Meta-studies:
     - Leveraging variation in σ^*.
     - Assume σ^* is (conditionally) independent of Θ^* across latent studies.
     - Standard assumption in the meta-studies literature; validated in our applications by comparison to replications.
- Advantages:
  1. Replications: Very credible
  2. Meta-studies: Widely applicable

Identification

Intuition for approach 1: Identification using replication studies



left No truncation
$\Rightarrow$ Areas A and B have same probability.

right A more likely then B.

$$p(z) = \begin{cases} 0.1 & |z| < 1.96 \\ 1 & |z| \geq 1.96 \end{cases}$$

Identification

Approach 1: Replication studies

- Consider the general setup introduced above.
- Assume that for each published estimate we **additionally observe a replication draw** X^r as well as σ^{r2} such that

$$X^{*r}|\Theta^*, \sigma^{*r}, \sigma^*, D, X^* \sim N(\Theta^*, \sigma^{*r2}).$$

- **Then** $p(\cdot)$ **is identified** up to scale, and π_Θ is identified as well.

Sketch of proof:

- Consider the special case $\sigma^{*r} = \sigma^*$.
- Marginal density of (X, X^r) is

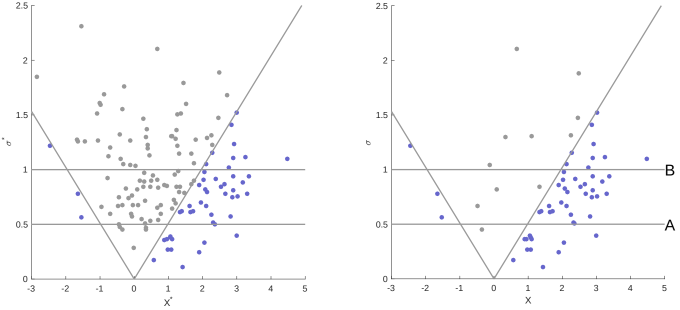
$$f_{Z,Z^r}(z, z^r) = \frac{p(z)}{E[p(Z^*)]} \int \varphi(z-\omega)\varphi(z^r-\omega)d\pi_\Omega(\omega).$$

- Thus, for all a, b, if $p(a) > 0$,

$$\frac{f_{Z,Z^r}(b,a)}{f_{Z,Z^r}(a,b)} = \frac{p(b)}{p(a)}.$$

Identification

Intuition for Approach 2: Identification using meta-studies



left No truncation

$\Rightarrow$ Dist for higher σ noised up version of dist for lower σ.

right "Missing data" inside the cone.

$$p(z) = \begin{cases} 0.1 & |z| < 1.96 \\ 1 & |z| \geq 1.96 \end{cases}$$

Identification

Approach 2: Meta-studies

- Consider the general setup introduced above.
- Assume additionally that σ^* **and** Θ^* **are independent**, and suppose that the support of σ contains an open interval.
- **Then** $p(\cdot)$ **is identified** up to scale, and π_Θ is identified as well.

Sketch of proof:

- Conditional density of Z given σ is

$$f_{Z|\sigma}(z|\sigma) = \frac{p(z)}{E[p(Z^*)|\sigma]} \int \varphi(z - \theta/\sigma) d\pi(\theta).$$

- Thus

$$\frac{f_{Z|\sigma}(z|\sigma_2)}{f_{Z|\sigma}(z|\sigma_1)} = \frac{E[p(Z^*)|\sigma = \sigma_1]}{E[p(Z^*)|\sigma = \sigma_2]} \cdot \frac{\int \varphi(z - \theta/\sigma_2) d\pi(\theta)}{\int \varphi(z - \theta/\sigma_1) d\pi(\theta)}.$$

- Recover π from right hand side,
  then recover $p(\cdot)$ from first equation.

Roadmap

I Which findings get published?
  1 Setup and bias-corrected inference
  2 Identification
    a Replication studies
    b Meta-studies
  3 **Application: Lab experiments in economics**

II Which findings should be published?
  1 Setup and optimal publication rules
  2 Selective publication and statistical inference
  3 Extensions
    a Dynamic model
    b Researcher incentives

Conclusion

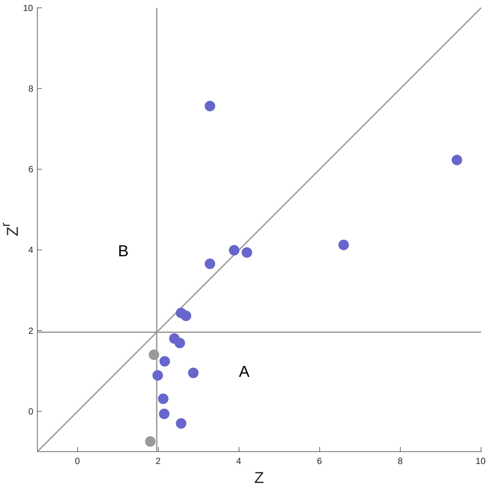
Application

Replications of Lab Experiments in Economics

- Camerer et al. (2016)
- Sample: all 18 between-subject laboratory experimental papers published in AER and QJE between 2011 and 2014.
- Scatterplot next slide:
  - $Z = X/\sigma$: normalized initial estimate.
  - $Z^r = X^r/\sigma$: replicate estimate.
  - Initial estimates normalized to be positive.

Application

Economics Lab Experiments: Original and Replication Z Statistics



Application

Economics Lab Experiments: Estimates of Selection model

- Model:

$$|\Omega^*| \sim \Gamma(\kappa, \lambda)$$
$$p(Z) \propto \begin{cases} \beta_p & |Z| < 1.96 \\ 1 & |Z| \geq 1.96 \end{cases}$$

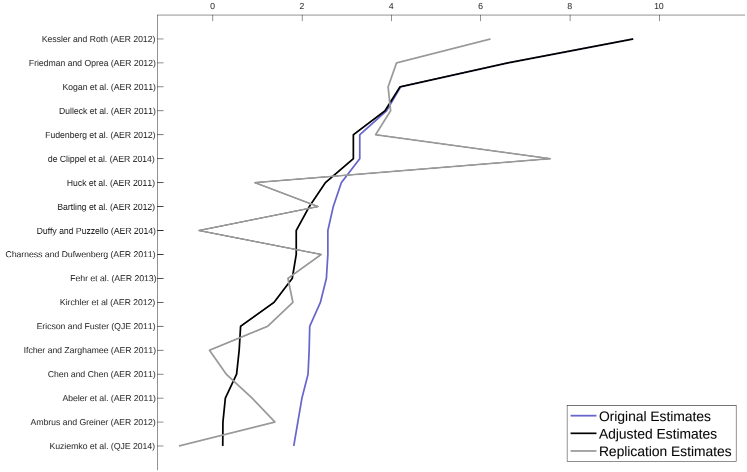
- Estimates:

κ	λ	β_p
0.373	2.153	0.029
(0.266)	(1.024)	(0.027)

- Interpretation: Insignificant (at the 5 % level) results about 3% as likely to be published as significant results.

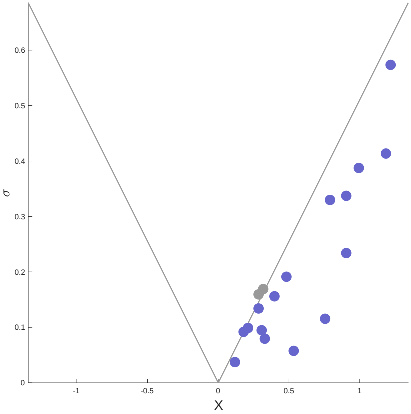
Application

Economics Lab Experiments: Adjusted Estimates



Application

Economics Lab Experiments: Meta-study Approach



Application

Economics Lab Experiments: Meta-study Results

- Model:

$$|\Theta^*| \sim \Gamma(\tilde{\kappa}, \tilde{\lambda})$$
$$p(Z) \propto \begin{cases} \beta_p & |Z| < 1.96 \\ 1 & |Z| \geq 1.96 \end{cases}$$

- Recall replication-based estimates:

κ	λ	β_p
0.373	2.153	0.029
(0.266)	(1.024)	(0.027)

- Meta-study based estimates (only β_p comparable):

$\tilde{\kappa}$	$\tilde{\lambda}$	β_p
1.343	0.157	0.038
(1.310)	(0.076)	(0.051)

Roadmap

I Which findings get published?
  1 Setup and bias-corrected inference
  2 Identification
    a Replication studies
    b Meta-studies
  3 Application: Lab experiments in economics

II Which findings should be published?
  1 Setup and optimal publication rules
  2 Selective publication and statistical inference
  3 Extensions
    a Dynamic model
    b Researcher incentives

Conclusion

Part II: Which findings should be published?

Reforming scientific publishing

- Publication bias motivates calls for reform:
  Publication should not select on findings.
  - De-emphasize statistical significance, ban "stars."
  - Pre-analysis plans to avoid selective reporting of findings.
  - Registered reports reviewed and accepted prior to data collection.
- But: Is eliminating bias the right objective?
  How does it relate to informing decision makers?
- We characterize **optimal publication rules from an instrumental perspective**:
  - Study might inform the public about some state of the world.
  - Then the public chooses a policy action.
  - Take as given that not all findings get published (prominently).

Which findings should be published?

Key results

1. **Optimal** rules selectively **publish surprising findings**.
   In leading examples: Similar to two-sided or one sided tests.
2. But: Selective publication **always distorts inference**.
   There is a trade-off policy relevance vs. statistical credibility.
3. With **dynamics**: Additionally publish **precise null** results.
4. With **incentives**: Modify publication rule to **encourage more precise** studies.

Roadmap

- I Which findings get published?
  - 1 Setup and bias-corrected inference
  - 2 Identification
    - a Replication studies
    - b Meta-studies
  - 3 Application: Lab experiments in economics
- II Which findings should be published?
  - 1 **Setup and optimal publication rules**
  - 2 Selective publication and statistical inference
  - 3 Extensions
    - a Dynamic model
    - b Researcher incentives
- Conclusion

Setup

Timeline and notation

State of the world (parameter)	θ
Common prior	$\theta \sim \pi_0$
Study might be submitted	
Exogenous submission probability	q
Design (e.g., standard error)	$S \perp \theta$
Findings (estimate)	$X\|\theta, S^2 \sim f_{X\|\theta,S}$
Journal decides whether to publish	$D \in \{0,1\}$
Publication probability	$p(X,S)$
Publication cost	c
Public updates beliefs	$\pi_1 = \pi_1^{(X,S)}$ if $D=1$
	$\pi_1 = \pi_1^0$ if $D=0$
Public chooses policy action	$a = a^*(\pi_1) \in \mathbb{R}$
Utility	$U(a,\theta)$
Social welfare	$U(a,\theta) - Dc$

Setup

Comparison to setup of Part I

- New layer: Belief updating and policy decision by the public.
- Consider only one study here, that might or might not be published. $\Rightarrow$
  - Omit $*$ notation.
  - θ is like previous Θ^*.
  - S is like previous σ^*.
- Changes of interpretation:
  - π_0 is a prior, rather than a distribution across studies.
  - S might include sources of uncertainty in addition to sampling variation. Internal and external validity!
  - A special case of $f_{X|\theta,S}$ is $N(\theta, S^2)$.

Baseline model

Belief updating and policy decision

- Public belief when study is published: $\pi_1^{(X,S)}$.
  - Bayes posterior after observing (X, S).
  - Same as journal's belief when study is submitted.
- Public belief when no study is published: π_1^0.
  Two alternative scenarios:
  1. Naive updating: $\pi_1^0 = \pi_0$.
  2. Bayes updating: π_1^0 is Bayes posterior given no publication.
- Public action $a = a^*(\pi_1)$
  maximizes posterior expected welfare, $\mathbb{E}_{\theta \sim \pi_1}[U(a, \theta)]$.
  Default action $a^0 = a^*(\pi_1^0)$.

Optimal publication rules

- Coming next: We show that

ex-ante optimal rules, maximizing expected welfare, are those which
ex-post publish findings that have a big impact on policy.

- **Interim gross benefit** $\Delta(\pi, a^0)$ of publishing equals
  - Expected welfare given publication, $\mathbb{E}_{\theta\sim\pi}[U(a^*(\pi), \theta)]$,
  - minus expected welfare of default action, $\mathbb{E}_{\theta\sim\pi}[U(a^0, \theta)]$.
- **Interim optimal publication rule**:
  Publish if interim benefit exceeds cost c.
- Want to maximize **ex-ante expected welfare**:

$$EW(p, a^0) = \mathbb{E}[U(a^0, \theta)] + q \cdot \mathbb{E}\Big[p(X, S) \cdot (\Delta(\pi_1^{(X,S)}, a^0) - c)\Big].$$

- Immediate consequence:
  Optimal policy is interim optimal given a^0.

Optimal publication rules

Optimality and interim optimality

- Under **naive updating**:
  - Default action $a^0 = a^*(\pi_0)$ does not depend on p.
  - **Interim optimal** rule given a^0 **is optimal**.
- Under **Bayes updating**:
  - a^0 maximizes $EW(p, a^0)$ given p.
  - p maximizes $EW(p, a^0)$ given a^0, when interim optimal.
  - These conditions are **necessary but not sufficient** for joint optimality.
- **Commitment does not matter** in our model.
  - Ex-ante optimal is interim optimal.
  - This changes once we consider researcher incentives (endogenous study submission).

Leading examples

- **Normal prior and signal**, normal posterior:

$$\theta \sim \pi_0 = N(\mu_0, \sigma_0^2)$$
$$X|\theta, S \sim N(\theta, S^2)$$

- **Canonical utility functions**:
  1. Quadratic loss utility, $\mathscr{A} = \mathbb{R}$:

     $$U(a,\theta) = -(a-\theta)^2$$

     Optimal policy action: $a =$ posterior mean.

  2. Binary action utility, $\mathscr{A} = \{0,1\}$:

     $$U(a,\theta) = a \cdot \theta$$

     Optimal policy action: $a = 1$ iff posterior mean is positive.

Leading examples

Interim optimal rules

- Quadratic loss utility: "**Two-sided test**." Publish if

$$\left|\mu_1^{(X,S)} - a^0\right| \geq \sqrt{c}.$$

- Binary action utility: "**One-sided test**." Publish if

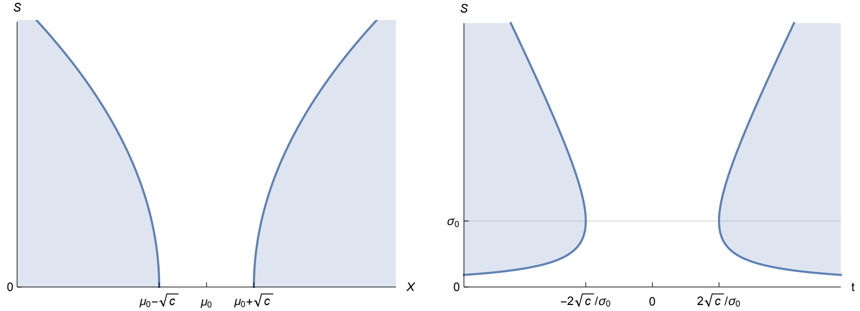
$$a^0 = 0 \text{ and } \mu_1^{(X,S)} \geq c, \quad \text{or}$$
$$a^0 = 1 \text{ and } \mu_1^{(X,S)} \leq -c.$$

- Normal prior and signals:

$$\mu_1^{(X,S)} = \frac{\sigma_0^2}{S^2+\sigma_0^2} X + \frac{S^2}{S^2+\sigma_0^2} \mu_0.$$

Leading examples

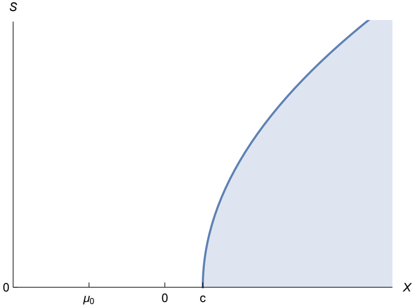
Quadratic loss utility, normal prior, normal signals

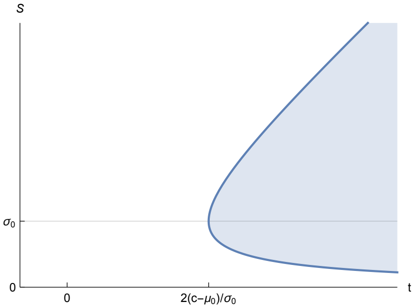


- Optimal publication region (shaded). Axes:
  - left: Estimate X, standard error S. (As in meta-studies plots!)
  - right: "t-statistic" $t = (X - \mu_0)/S$, standard error S.
- Note:
  - Given S, publish outside symmetric interval around μ_0.
  - Critical value for t-statistic is non-monotonic in S.

Leading examples

Binary action utility, normal prior, normal signals





- Optimal publication region (shaded). Axes:
  - left: Estimate X, standard error S.
  - right: "t-statistic" $t = (X - \mu_0)/S$, standard error S.
- Note:
  - When prior mean is negative, optimal rule publishes for large enough positive X.

Generalizing beyond these examples

Two key results that generalize:

- **Don't publish null results:**
  A finding that induces $a^*(\pi^I) = a^0 = a^*(\pi_1^0)$ always has 0 interim benefit and should never get published.

- **Publish findings outside interval:**
  Suppose
  - U is supermodular.
  - $f_{X|\theta,S}$ satisfies monotone likelihood ratio property given $S = s$.
  - Updating is either naive or Bayes.

  Then there exists an interval $I^s \subseteq \mathbb{R}$ such that (X, s) is published under the optimal rule if and only if $X \notin I^s$.

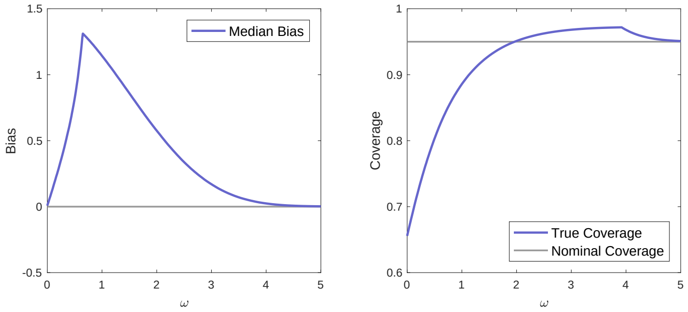
Roadmap

Selective publication and inference

- Just showed:
  Optimal publication rules select on findings.
- But: Selective publication rules can distort inference.
- We show a stronger result:
  Any selective publication rule distorts inference.
- Put differently:
  If we desire that standard inference be valid, then the publication rule must not select on findings at all.

Recall: Publication bias

Example: Selection on significance



- Publication probability: “significance testing,”

$$p(z) = \begin{cases} 0.1 & |z| < 1.96 \\ 1 & |z| \geq 1.96 \end{cases}$$

left median bias of $\hat{\omega} = Z$

right true coverage of conventional 95% confidence interval

Selective publication and inference

Validity of inference is equivalent to no selection

For normal signals and prior support with non-empty interior, **the following statements are equivalent**:

1. Non-selective publication.
   $p(x, s)$ is constant in x for each s.
2. Publication probability constant in state.
   $\mathbb{E}[p(X, S)|\theta, S = s]$ is constant over $\theta \in \Theta_0$ for each s.
3. Frequentist unbiasedness.
   $\mathbb{E}[X|\theta, S = s, D = 1] = \theta$ for $\theta \in \Theta_0$ and for all s.
4. Bayesian validity of naive updating.
   For all distributions F_S, the Bayesian default belief π_1^0 is equal to the prior π_0.

Selective publication and inference

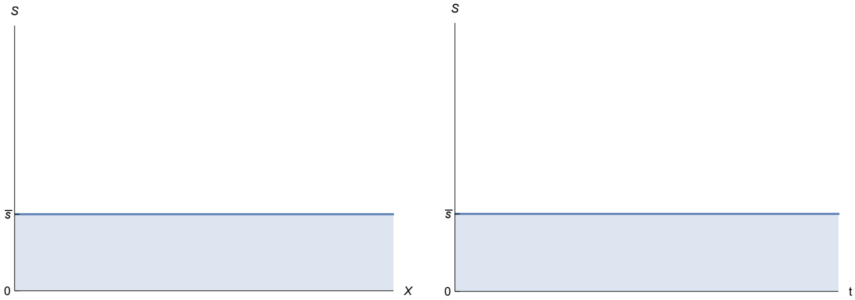
Intuition and implications

- Sketch of proof:
  - Non-selective publication $\Rightarrow$ the other conditions: immediate.
  - Constant publication probability $\Rightarrow$ non-selective publication: Completeness of the normal location family.
  - Unbiasedness $\Rightarrow$ constant publication probability: "Tweedie's formula" and integration.
- **Optimal publication if we require non-selectivity**?
- Suppose
  - There are normal signals.
  - Updating is either naive or Bayesian.
  - The publication rule is restricted to not select on X.

  Then there exists $\bar{s} \geq 0$ for which the optimal rule **publishes** a study **if and only if** $S \leq \bar{s}$.

Selective publication and inference

Optimal non-selective publication region



- For quadratic loss utility, normal prior, normal signals.
- Subject to the constraint that $p(x, s)$ is restricted to not depend on x.

Roadmap

I Which findings get published?
  1 Setup and bias-corrected inference
  2 Identification
    a Replication studies
    b Meta-studies
  3 Application: Lab experiments in economics

II Which findings should be published?
  1 Setup and optimal publication rules
  2 Selective publication and statistical inference
  3 **Extensions**
    a Dynamic model
    b Researcher incentives

Conclusion

A dynamic two-period model

- Period 1 as before, with study (X_1, S_1), action $a_1 = a^*(\pi_1)$.
- Now additionally: Period 2 study, always published.
- Independent estimate

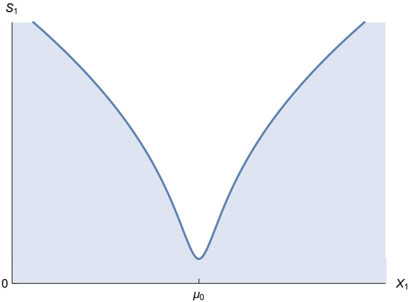
$$X_2|\theta, X_1, S_1 \sim F_{X_2|\theta}.$$

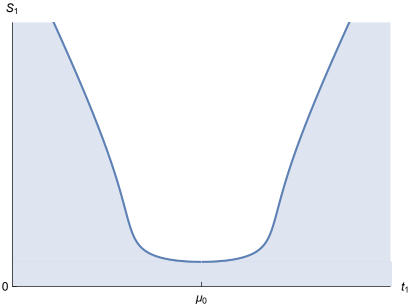
- Period 2 action $a_2 = a^*(\pi_2)$.
- Social welfare

$$\alpha U(a_1, \theta) - Dc + (1-\alpha) U(a_2, \theta).$$

A dynamic two-period model

Quadratic loss utility, normal prior, normal signals, naive updating





- Optimal publication region (shaded).
- Note:
  - For S small enough, publish even when $X = \mu_0$.

A dynamic two-period model

General implications

- More generally, for normal signals:
  1. The benefit of publication is strictly positive whenever $\pi_1^I \neq \pi_1^0$.
  2. The benefit goes to 0 as either $s_2 \to 0$ or $s_2 \to \infty$.
- Put differently:
  1. **Null results** (= prior mean) **that improve precision** are valuable to **prevent future mistakes.**
  2. This value disappears for
     a) very precise future information (won't make any mistakes either way), and
     b) very imprecise future information (effectively back to one-period case).

Researcher Incentives

- Thus far: study submission and design exogenous, random.
- Assume now that a researcher
  1. decides whether or not to submit a study,
  2. and picks a design S.
- Normal signals with standard error S.
- Researcher utility:
  1. Utility 1 from getting published,
  2. cost $\kappa(S)$ depending on design S.
- Expected researcher utility

$$E_{\theta \sim \pi_0, X \sim N(\theta, S^2)}[p(X,S)] - \kappa(S).$$

- Outside option with utility 0.
- Journal faces
  1. participation constraint (PC) and
  2. incentive compatibility constraint (ICC).

Researcher Incentives

Constrained optimal rule

- Journal objective as before, $U(a, \theta) - Dc$.
- Journal commits to publication rule $p(x, s)$ ex-ante. Commitment matters in this extension!
- Optimal publication rule subject to (PC) and (ICC)?
- Solution: Relative to baseline model, journal **distorts publication rule** in two ways
  - Reject imprecise studies (large S) – even if valuable ex post.
  - For low enough S, set interim benefit threshold for acceptance below c.

Conclusion

Key findings

I Which findings get published?
   1. If form and magnitude of selection are known, we can correct published findings.
   2. Form and magnitude of selection are nonparametrically identified.
   3. Published research is selected.

II Which findings should be published?
   1. Optimal rules selectively publish surprising findings. In leading examples: Similar to two-sided or one sided tests.
   2. But: Selective publication always distorts inference. There is a trade-off policy relevance vs. statistical credibility.
   3. With dynamics: Additionally publish precise null results.
   4. With incentives: Modify publication rule to encourage more precise studies.

Conclusion

Outlook

Different ways of thinking about statistics / econometrics:

1. Making decisions based on data.
    - Objective function?
    - Set of feasible actions?
    - Prior information?
2. Statistics as (optimal) communication.
    - Not just "you and the data."
    - What do we communicate to whom?
    - Subject to what costs and benefits?
      Why not publish everything? Attention?
3. Statistics / research as a social process.
    - Researchers, editors and referees, policymakers.
    - Incentives, information, strategic behavior.
    - Social learning, paradigm changes.

Much to be done!

A web-app for estimating publication bias in meta-studies is available at

`https://maxkasy.github.io/home/metastudy/`

Thank you!