Identification of and correction for publication bias

Isaiah Andrews   Maximilian Kasy

December 13, 2017

Introduction

- Fundamental requirement of science: replicability
- Different researchers should reach same conclusions
- Methodological conventions should ensure this (e.g., randomized experiments)
- Replicability often appears to fail, e.g.
  - Experimental economics (Camerer et al., 2016)
  - Experimental psychology (Open Science Collaboration, 2015)
  - Medicine (Ionnidias, 2005)
  - Cell Biology (Begley et al, 2012)
  - Neuroscience (Button et al, 2013)

Introduction

- Possible explanation: selective publication of results
- Due to:
  - Researcher decisions
  - Journal selectivity
- Possible selection criteria:
  - Statistically significant effects
  - Confirmation of prior beliefs
  - Novelty
- Consequences:
  - Conventional estimators are biased
  - Conventional inference does not control size

Introduction

Literature

Identification of publication bias:

- Good overview:
  Rothstein et al. (2006)
- Regression based:
  Egger et al. (1997)
- Symmetry of funnel plot ("trim and fill"):
  Duval and Tweedie (2000)
- Parametric selection models:
  Hedges (1992), Iyengar and Greenhouse (1988)
- Distribution of p-values, parametric distribution of true effects:
  Brodeur et al. (2016)

Corrected inference:

- McCrary et al. (2016)

Replication- and meta-studies for empirical part:

- Replication of econ experiments: Camerer et al. (2016)
- Replication of psych experiments: Open Science Collaboration (2015)
- Minimum wage: Wolfson and Belman (2015)
- Deworming: Croke et al. (2016)

Introduction

Our contributions

1. Nonparametric **identification of selectivity** in the publication process, using
   a) Replication studies: Absent selectivity, original and replication estimates should be symmetrically distributed
   b) Meta-studies: Absent selectivity, distribution of estimates for small sample sizes should be noised-up version of distribution for larger sample sizes
2. **Corrected inference** when selectivity is known
   a) Median unbiased estimators
   b) Confidence sets with correct coverage
   c) Allow for nuisance parameters and multiple dimensions of selection
   d) Bayesian inference accounting for selection
3. **Applications** to
   a) Experimental economics
   b) Experimental psychology
   c) Effects of minimum wages on employment
   d) Effects of de-worming

Outline

Setup

- Assume there is a population of latent studies indexed by i
- True parameter value in study i is Θ_i^*
  - Θ_i^* drawn from some population $\Rightarrow$ empirical Bayes perspective
  - Different studies may recover different parameters
- Each study reports findings X_i^*
  - Distribution of X_i^* given Θ_i^* known
- A given study may or may not be published
  - Determined by both researcher and journal: we don't try to disentangle
- Probability of publication $P(D_i = 1 | X_i^*, \Theta_i^*) = p(X_i^*)$
- Published studies are indexed by j

Setup

Definition (General sampling process)

Latent (unobserved) variables: (D_i, X_i^*, Θ_i^*)*, jointly i.i.d. across* i

$$\Theta_i^* \sim \mu$$
$$X_i^* | \Theta_i^* \sim f_{X^*|\Theta^*}(x|\Theta_i^*)$$
$$D_i | X_i^*, \Theta_i^* \sim Ber(p(X_i^*))$$

Truncation: We observe i.i.d. draws of X_j*, where*

$$I_j = \min\{i : \; D_i = 1, \; i > I_{j-1}\}$$
$$\Theta_j = \Theta_{I_j}^*$$
$$X_j = X_{I_j}^*$$

Setup

Example: treatment effects

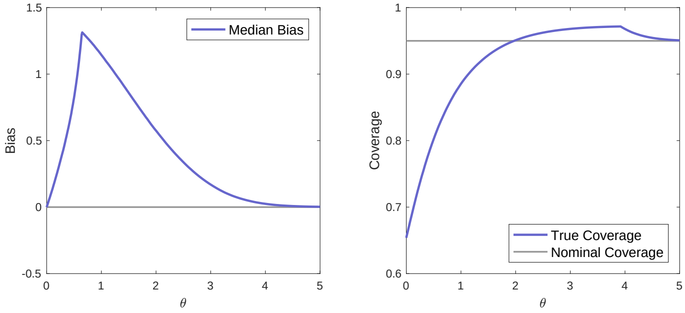
- Journal receives a stream of studies $i = 1, 2, \ldots$
- Each reporting experimental estimates X_i^* of treatment effects Θ_i^*
- Distribution of Θ_i^*: μ
- Suppose that $X_i^* | \Theta_i^* \sim N(\Theta_i^*, 1)$
- Publication probability: "significance testing,"

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

- Published studies: report estimate X_j of treatment effect Θ_j

Setup

Example continued – Publication bias



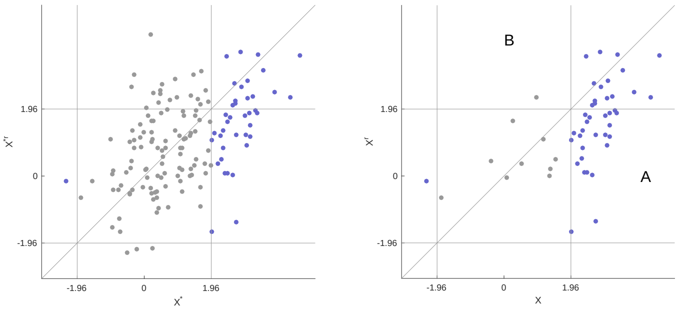
- Left: median bias of $\hat{\theta}_j = X_j$
- Right: true coverage of conventional 95% confidence interval

Outline

- Key unknown object in model: publication probability $p(\cdot)$
- We propose two approaches for identification:
  1. Replication experiments:
     - replication estimate X^r for the same parameter Θ
     - selectivity operates only on X, but not on X^r
  2. Meta-studies:
     - Variation in σ^*, where $X^* \sim N(\Theta^*, \sigma^{*2})$
     - Assume σ^* is (conditionally) independent of Θ^* across latent studies i
     - 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: identification using replication studies



- Left: no truncation
  $\Rightarrow$ areas A and B have same probability
- Right: $p(Z) = 0.1 + 0.9 \cdot \mathbf{1}(|Z| > 1.96)$
  $\Rightarrow$ A more likely then B

Identification

Approach 1: Replication studies

Definition (Replication sampling process)

- *Latent variables: as before,*

$$\begin{aligned} \Theta_i^* &\sim \mu \\ X_i^* | \Theta_i^* &\sim f_{X^*|\Theta^*}(x|\Theta_i^*) \\ D_i | X_i^*, \Theta_i^* &\sim Ber(p(X_i^*)) \end{aligned}$$

- *Additionally: replication draws,*

$$X_i^{*r} | X_i^*, D_i, \Theta_i^* \sim f_{X^*|\Theta^*}(x|\Theta_i^*)$$

- *Observability: as before,*

$$\begin{aligned} I_j &= \min\{i: \ D_i = 1, \ i > I_{j-1}\} \\ \Theta_j &= \Theta_{I_j} \\ (X_j, X_j^r) &= (X_{I_j}^*, X_{I_j}^{*r}) \end{aligned}$$

Identification

Theorem (Identification using replication experiments)

Assume that the support of $f_{X_i^*, X_i^{*r}}$ is of the form $A \times A$ for some set A. Then $p(\cdot)$ is identified on A up to scale.

Intuition of proof:

- Marginal density of (X, X^r) is

$$f_{X,X^r}(x, x^r) = \frac{p(x)}{E[p(X_i^*)]} \int f_{X^*|\Theta^*}(x|\theta_i^*)\, f_{X^*|\Theta^*}(x^r|\theta_i^*)\, d\mu(\theta_i^*)$$

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

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

Identification

Practical complication

- Replication experiments follow the same protocol
  $\Rightarrow$ estimate same effect Θ
- But often different sample size
  $\Rightarrow$ different variance $\Rightarrow$ symmetry breaks down
- Additionally: replication sample size often determined based on power calculations given initial estimate
- $p(\cdot)$ is still identified (up to scale):
  - Assume X normally distributed
  - Intuition: Conditional on X, σ, (de-)convolve X^r with normal noise to get symmetry back
  - μ is identified as well

Identification

Further complication

- What if selectivity is based not only on observed X, but also on unobserved W?
- Would imply general selectivity of the form

$$D_i | X_i^*, \Theta_i^* \sim Ber(p(X_i^*, \Theta_i^*))$$

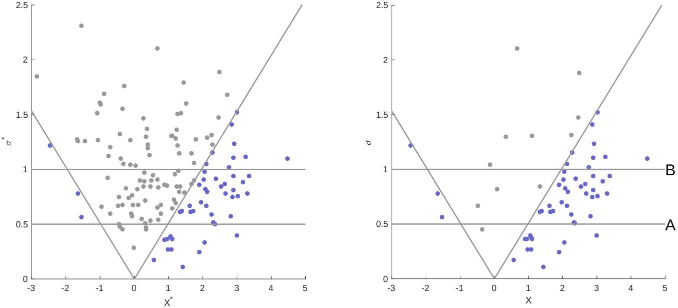
- Again assume normality,

$$X_i^{*r} | \sigma_i, D_i, X_i^*, \Theta_i^* \sim N(\Theta_i^*, \sigma_i^2)$$

- $\Rightarrow$ Solution:
  - Identify $\mu_{\Theta|X}$ from $f_{X^r|X}$ by deconvolution
  - Recover $f_{X|\Theta}$ by Bayes' rule (f_X is observed)
  - This density is all we need for bias corrected inference
- We use this to construct specification tests for our baseline model

Identification

Intuition: identification using meta-studies



- Left: no truncation
  $\Rightarrow$ dist for higher σ noised up version of dist for lower σ
- Right: $p(Z) = 0.1 + 0.9 \cdot \mathbf{1}(|Z| > 1.96)$
  $\Rightarrow$ "missing data" inside the cone

Identification

Approach 2: meta-studies

Definition (Independent σ sampling process)

$$\sigma_i^* \sim \mu_\sigma$$
$$\Theta_i^* | \sigma_i^* \sim \mu_\Theta$$
$$X_i^* | \Theta_i^*, \sigma_i^* \sim N(\Theta_i^*, \sigma_i^{*2})$$
$$D_i | X_i^*, \Theta_i^*, \sigma_i^* \sim Ber(p(X_i^*/\sigma_i^*))$$

We observe i.i.d. draws of (X_j, σ_j), where

$$I_j = \min\{i : \; D_i = 1, \; i > I_{j-1}\}$$
$$(X_j, \sigma_j) = (X_{I_j}^*, \sigma_{I_j}^*)$$

Define $Z^* = \frac{X^*}{\sigma^*}$ and $Z = \frac{X}{\sigma}$

Identification

Theorem (Nonparametric identification using variation in σ)

Suppose that the support of σ contains a neighborhood of some point σ_0. Then $p(\cdot)$ is identified up to scale.

Intuition of proof:

- Conditional density of Z given σ is

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

- Thus

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

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

Outline

Bias-corrected inference

- Once we know $p(\cdot)$, can correct inference for selection
- For simplicity, here assume X, Θ both 1-dimensional
- Density of published X given Θ:

$$f_{X|\Theta}(x|\theta) = \frac{p(x)}{E[p(X^*)|\Theta^* = \theta]} \cdot f_{X^*|\Theta^*}(x|\theta)$$

- Corresponding cumulative distribution function: $F_{X|\Theta}(x|\theta)$

Bias-corrected inference

Corrected frequentist estimators and confidence sets

- We are interested in bias, and the coverage of confidence sets
  - Condition on θ: standard frequentist analysis
- Define $\hat{\theta}_\alpha(x)$ via

$$F_{X|\Theta}\left(x|\hat{\theta}_\alpha(x)\right) = \alpha$$

- Under mild conditions, can show that

$$P\left(\hat{\theta}_\alpha(X) \leq \theta|\theta\right) = \alpha \ \ \forall\theta$$

- Median-unbiased estimator: $\hat{\theta}_{\frac{1}{2}}(X)$ for θ
- Equal-tailed level $1-\alpha$ confidence interval:

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

Bias-corrected inference

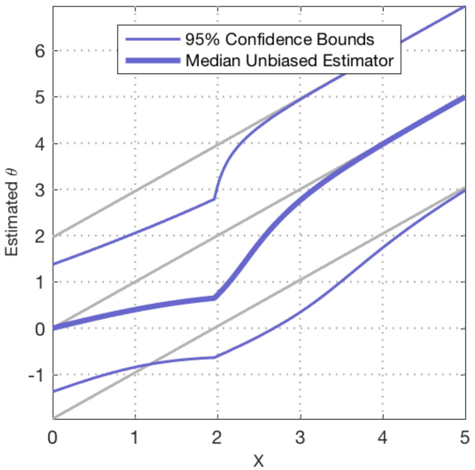
Example: treatment effects

- Let us return to the treatment effect example discussed above
- Again assume $X^*|\Theta^* \sim N(\Theta^*, 1)$ and

$$p(X) = 0.1 + 0.9 \cdot \mathbf{1}(|X| > 1.96)$$

Bias-corrected inference

Example continued – corrected confidence sets for $\beta_p = 0.1$



Outline

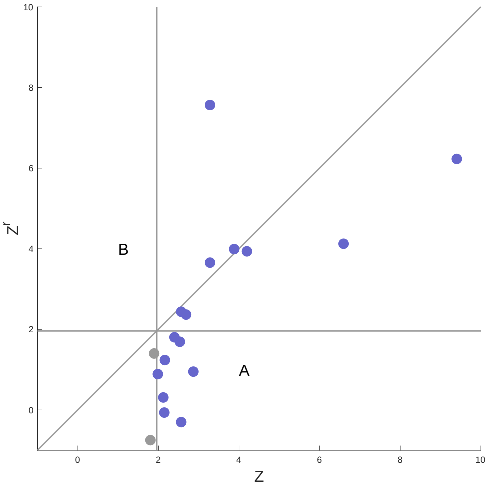
Applications

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

Applications

Economics Lab Experiments: Original and Replication Z Statistics



Applications

Economics Lab Experiments: Estimates of Selection model

- Model:

$$\Theta^* \sim N(0, \tau^2)$$
$$p(Z) \propto \begin{cases} \beta_p & |Z| < 1.96 \\ 1 & |Z| \geq 1.96 \end{cases}$$

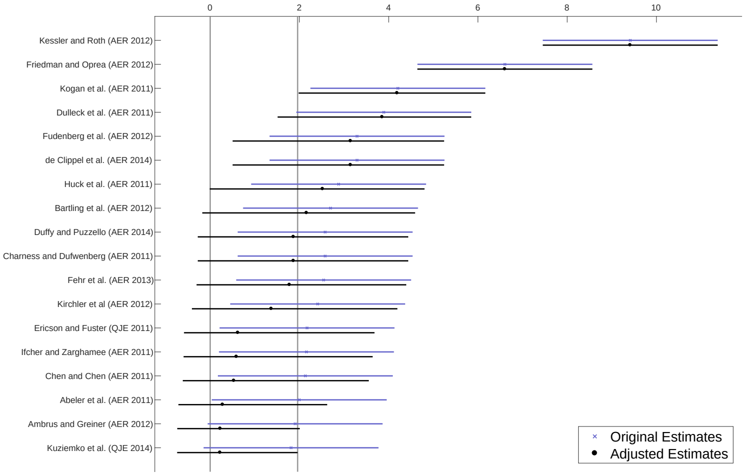
- Estimates:

τ	β_p
2.354	0.100
(0.750)	(0.091)

- Interpretation: insignificant (at the 5 % level) results about 10% as likely to be published as significant results

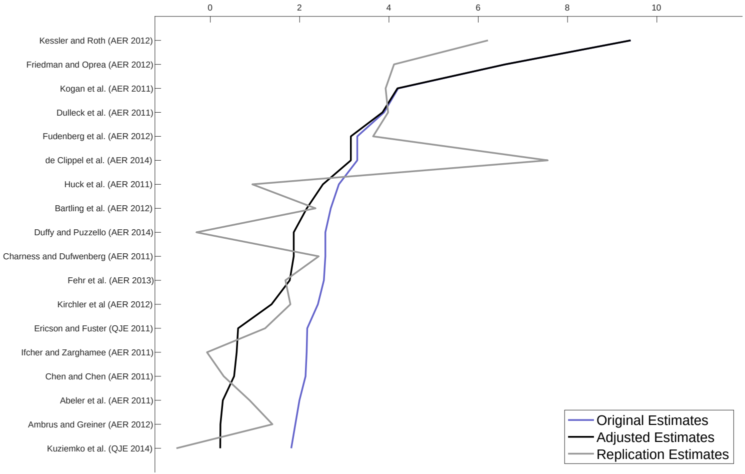
Applications

Economics Lab Experiments: Adjusted Estimates



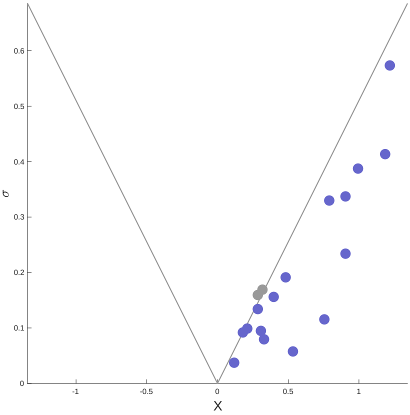
Applications

Economics Lab Experiments: Adjusted Estimates



Applications

Economics Lab Experiments: Meta-study Approach



Applications

Economics Lab Experiments: Meta-study Results

- Model:

$$\Theta^* \sim N(0, \tilde{\tau}^2)$$

$$p(X/\sigma) \propto \begin{cases} \beta_p & |X/\sigma| < 1.96 \\ 1 & |X/\sigma| \geq 1.96 \end{cases}$$

- Recall replication-based estimates:

τ	β_p
2.354	0.100
(0.750)	(0.091)

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

$\tilde{\tau}$	β_p
0.299	0.045
(0.073)	(0.045)

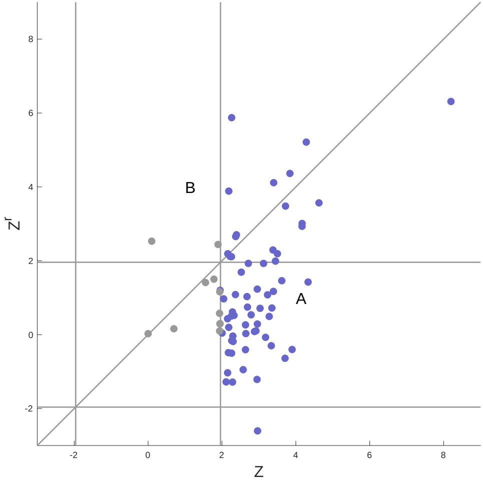
Applications

Replications of Lab Experiments in Psychology

- Open Science Collaboration (2015)
- 270 contributing authors
- Sample: 100 out of 488 articles published 2008 in
  - Psychological Science
  - Journal of Personality and Social Psychology
  - Journal of Experimental Psychology: Learning, Memory, and Cognition
- Some critiques by Gilbert et al. (2016):
  - statistical misinterpretation,
  - not all replication protocols endorsed by original authors
    $\Rightarrow$ we re-run estimators on subset of approved replications

Applications

Experiments in Psychology: Original and Replication Z Statistics



Applications

Experiments in Psychology: Estimates of Selection Model

- Model:

$$\Theta^* \sim N(0, \tau^2)$$

$$p(Z) \propto \begin{cases} \beta_{p1} & |Z| < 1.64 \\ \beta_{p2} & 1.64 \leq |Z| < 1.96 \\ 1 & |Z| \geq 1.96 \end{cases}$$

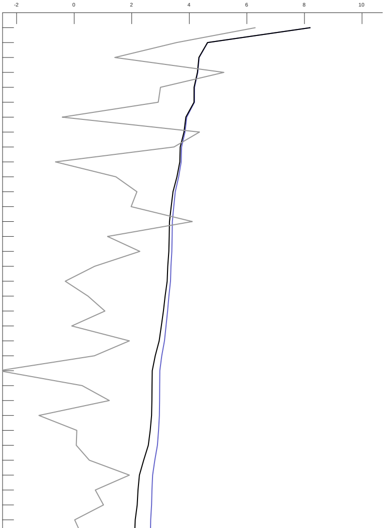
- Estimates:

τ	$\beta_{p,1}$	$\beta_{p,2}$
1.252	0.021	0.294
(0.195)	(0.012)	(0.128)

- Results insignificant at the 10% level 2% as likely to be published as results significant at 5% level
- Results significant at the 5% level over three times as likely to be published as results significant at 10% level

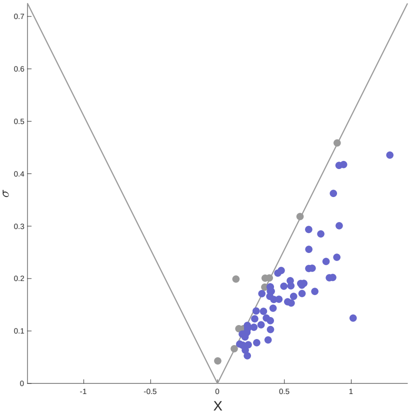
Applications

Original and Replication Z Statistics: Psychology Lab Experiments



Applications

Psychology Lab Experiments: Meta-studies Approach



Applications

Psychology Lab Experiments: Estimates of Meta-studies Selection Model

- Model:

$$\Theta^* \sim N(0, \tau^2)$$

$$p(Z) \propto \begin{cases} \beta_{p1} & |Z| < 1.64 \\ \beta_{p2} & 1.64 \leq |Z| < 1.96 \\ 1 & |Z| \geq 1.96 \end{cases}$$

- Recall replication-based estimates:

τ	$\beta_{p,1}$	$\beta_{p,2}$
1.252	0.021	0.294
(0.195)	(0.012)	(0.128)

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

$\tilde{\tau}$	$\beta_{p,1}$	$\beta_{p,2}$
0.252	0.025	0.375
(0.041)	(0.015)	(0.166)

Applications

Psychology Lab Experiments: Approved Replications

- 67 studies
- Replication-based estimates:

τ	$\beta_{p,1}$	$\beta_{p,2}$
1.385	0.038	0.512
(0.272)	(0.024)	(0.239)

- Meta-study based estimates:

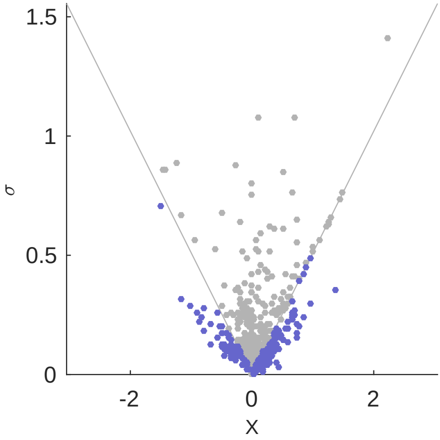
$\tilde{\tau}$	$\beta_{p,1}$	$\beta_{p,2}$
0.272	0.042	0.621
(0.055)	(0.027)	(0.300)

- β_p estimates systematically larger than those in full dataset

Applications

Meta-study of the Effect of Minimum Wages on Employment

- Wolfson and Belman (2015)
- Elasticity of employment w.r.t. the minimum wage
  $X > 0 \Leftrightarrow$ negative employment effect
- 1000 estimates from 37 studies using U.S. data that were circulated after 2000, either as articles in journals or as working papers
- For some: more than 1 estimate per study



1.5
1
0.5
0
σ
-2
0
2
X

Estimates of selection model

- Model:

$$\Theta^* \sim N(\bar{\theta}, \tau^2)$$

$$p(X/\sigma) \propto \begin{cases} \beta_{p1} & X/\sigma < -1.96 \\ \beta_{p2} & -1.96 \leq X/\sigma < 0 \\ \beta_{p3} & 0 \leq X/\sigma < 1.96 \\ 1 & X/\sigma \geq 1.96 \end{cases}$$

- Estimates:

$\bar{\theta}$	$\tilde{\tau}$	$\beta_{p,1}$	$\beta_{p,2}$	$\beta_{p,3}$
-0.024	0.122	0.225	0.424	0.738
(0.053)	(0.038)	(0.117)	(0.205)	(0.289)

- Bias in favor of estimates which find minimum wage reduces employment

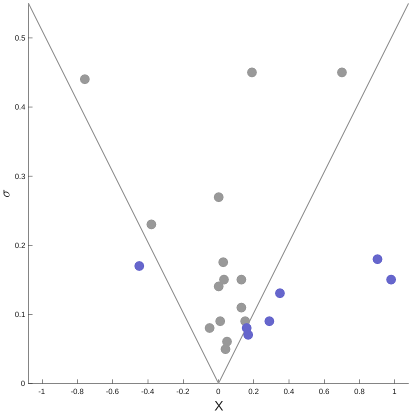
Applications

Meta-Study of the Effects of Deworming

- Croke et al. (2016)
- Follow procedures outlined in the “Cochrane Handbook for Systematic Reviews of Interventions”
- Randomized controlled trials of deworming that include child body weight as an outcome
- 22 estimates from 20 studies

Applications

Meta-Study of the Effects of Deworming



Applications

Deworming: Estimates of selection model

- Model:

$$\Theta^* \sim N(\bar{\theta}, \tau^2)$$

$$p(X) \propto \begin{cases} \beta_p & |X/\sigma| < 1.96 \\ 1 & |X/\sigma| \geq 1.96 \end{cases}$$

- Estimates:

$\bar{\theta}$	$\tilde{\tau}$	β_p
0.190	0.343	2.514
(0.120)	(0.128)	(1.872)

Conclusion

- Selectivity in the publication process is a potentially serious problem for statistical inference.
- We non-parametrically identify the form of selectivity:
  - Using replication studies:
    Original and replication estimates would be symmetrically distributed, absent selectivity
  - Using meta-studies:
    Under an independence assumption, higher-variance estimates distribution would be noised-up version of lower-variance estimate distribution, absent selectivity

Conclusion

- Easy correction for selectivity, if form is known:
  - Median unbiased estimators
  - Equal-tailed confidence sets with correct coverage
- Empirical findings:
  - Selectivity on significance in experimental economics, experimental psychology
  - Selectivity towards negative employment effects in minimum wage literature
  - Noisy estimates in meta-study for de-worming

Thank you!