Identification of and correction for publication bias*

Isaiah Andrews† Maximilian Kasy‡

October 4, 2018

Abstract

Some empirical results are more likely to be published than others. Selective publication leads to biased estimates and distorted inference. We propose two approaches for identifying the conditional probability of publication as a function of a study's results, the first based on systematic replication studies and the second on meta-studies. For known conditional publication probabilities, we propose bias-corrected estimators and confidence sets. We apply our methods to recent replication studies in experimental economics and psychology, and to a meta-study on the effect of the minimum wage. When replication and meta-study data are available, we find similar results from both.

Keywords: Publication bias, replication, meta-studies, identification, experimental economics, minimum wage

JEL Codes: C18, C12, C13

1 Introduction

Despite following the same protocols, replications of published experiments frequently find effects of smaller magnitude or opposite sign than those in the initial studies (cf. Open

*We thank Josh Angrist, Ellora Derenoncourt, Gary Chamberlain, Xavier D'Haultfoeuille, Andrew Gelman, Gary King, Jesse Shapiro, Jann Spiess, and seminar participants at Brown, Carlos III, CEMFI, Columbia, CREST, Harvard/MIT, the Harvard applied statistics seminar, the Harvard development retreat, Microsoft Research, the Montreal econometrics seminar, Penn, Penn State, Princeton, TSE, University of Connecticut, and Yale for many helpful comments and suggestions. We also thank Paul Wolfson and Dale Belman as well as Michael Kremer for sharing their data. This research was funded in part by the Silverman (1968) Family Career Development Chair at MIT, and the National Science Foundation (under grant number 1654234), and the Alfred P. Sloan Foundation (under the grant "Publication bias and specification searching. Identification, correction, and reform proposals").

†Department of Economics, Harvard University, iandrews@fas.harvard.edu

‡Department of Economics, Harvard University, maximiliankasy@fas.harvard.edu

Science Collaboration, 2015; Camerer et al., 2016). One leading explanation for replication failure is publication bias (cf. Ioannidis, 2005, 2008; McCrary et al., 2016; Christensen and Miguel, 2016). Journal editors and referees may be more likely to publish results that are statistically significant, that confirm some prior belief or, conversely, that are surprising. Researchers in turn face strong incentives to select which findings to write up and submit to journals based on the likelihood of ultimate publication, leading to what is sometimes called the file drawer problem (Rosenthal, 1979). We refer to these behaviors collectively as selective publication or publication bias. Left unaddressed, such selectivity can lead to severe bias in published estimates and confidence sets.

We first show how bias from selective publication can be corrected if the conditional publication probability (i.e. the probability of publication as a function of a study's results) is known. We then show how the conditional publication probability can be nonparametrically identified. Finally, we apply the proposed methods to several empirical literatures.

Correcting for publication bias After introducing our setup, Section 2 discusses the consequences of selective publication for statistical inference. When selectivity is known (at least up to scale), we propose median unbiased estimators and valid confidence sets for scalar parameters.[1]

Identification of publication bias Section 3 considers two approaches to identification. The first uses data from systematic replications of a collection of original studies. Following e.g. Camerer et al. (2016), by a replication we mean a study which applies the same experimental protocol to a new sample from the same population as the corresponding original study.[2] If the original and replication have the same sample size, then if there is no selectivity the joint distribution of initial and replication estimates across studies is symmetric. Asymmetries in this joint distribution nonparametrically identify conditional publication probabilities, assuming the latter depend only on the initial estimate. While replication sample sizes often differ from those in the initial study, we show that nonparametric identification extends to this case as well.

Our second identification approach uses data from meta-studies, by which we mean

[1] While our corrections eliminate bias due to selective publication, they cannot correct for problems with the underlying studies. If a study suffers from omitted variables bias (cf. Bruns and Ioannidis, 2016; Bruns, 2017), for instance, our corrections provide median unbiased estimates for the sum of the parameter of interest and the omitted variables bias. See Section 5.5 below for further discussion.

[2] Clemens (2017) terms such studies "reproductions," to distinguish them from "verifications" (cf. Chang and Li, 2018; Gertler et al., 2018) which try to reproduce the same results as the original paper based on the original sample.

studies that collect estimates and standard errors from multiple (published) studies. Under an independence assumption common in the meta-studies literature, if there is no selectivity then we can write the distribution of estimates for high variance studies as the distribution for low variance studies plus noise. Deviations from this prediction again identify conditional publication probabilities.

Both approaches identify conditional publication probabilities up to scale. Multiplying publication probabilities by a constant factor does not change the distribution of published estimates, and likewise does not affect publication bias and size distortions. Hence, identification up to scale is sufficient to apply our bias corrections.

Applications Section 4 applies the theory developed in this paper to three empirical literatures. Our first two applications use data from the experimental economics and psychology replication studies of Camerer et al. (2016) and Open Science Collaboration (2015), respectively. Estimates based on our replication approach suggest that results significant at the 5% level are over 30 times more likely to be published than are insignificant results, providing strong evidence of selectivity. Estimation based on our meta-study approach, which uses only the originally published results, yields similar conclusions.

Our third application considers the literature on the impact of minimum wages on employment, where no replication estimates are available. Estimates based on data from the meta-study by Wolfson and Belman (2015) suggest that results corresponding to a negative and significant effect of the minimum wage on employment are about 3 times more likely to be published than are insignificant results. Our point estimates suggest that results showing a positive and significant effect of minimum wages on employment are less likely to be published than negative and significant results, consistent with prior findings of Card and Krueger (1995) and Wolfson and Belman (2015), but we cannot reject that selection depends only on significance and not on sign. In the supplement we discuss two additional applications of our methods, using data from Croke et al. (2016) and Camerer et al. (2018).

Alternative approaches There is a large prior literature on publication bias. Section 5 discusses some of the alternatives from this literature, including meta-regression and approaches based on the distribution of p-values or z-statistics, and relates them to our framework. We further discuss the implications of "p-hacking" as discussed in e.g. Simonsohn et al. (2014) and Bruns and Ioannidis (2016) for our results.

Supplement A variety of supporting materials and extensions of our results are provided in the online supplement. Section A contains proofs for all results discussed in the

main text. Section B provides additional discussion of the data and methods used in our empirical applications, as well as a range of robustness checks. Section C contains further empirical results, including estimates based on alternative GMM estimation approaches and results for the Croke et al. (2016) and Camerer et al. (2018) applications. Finally, Section D discusses additional theoretical results, including results on inference with multidimensional selection and the impact of selection on Bayesian inference.

Notation Throughout the paper, upper case letters denote random variables and lower case letters denote realizations. We observe normally distributed estimates X with mean Θ and standard deviation Σ, where Θ and Σ may vary across studies.[3] We condition on Θ and Σ whenever frequentist objects are considered, while unconditional expectations, probabilities, and densities integrate over the population distribution of Θ and Σ. Estimates normalized by their standard deviation Σ are denoted by Z, and parameters Θ normalized by Σ are denoted by Ω. Latent studies (published or unpublished) are marked by a superscript $*$, while published studies have no superscript.

2 Setting

Throughout this paper we consider variants of the following data generating process. Within an empirical literature of interest, there is a population of latent studies i. The true effect Θ_i^* in study i is drawn from distribution μ_Θ. Thus, different latent studies may estimate different true parameters.[4] Conditional on the true effect Θ_i^* and the standard deviation Σ_i^* (which may also vary across studies), the result X_i^* in latent study i is drawn from the normal distribution $N(\Theta_i^*,\Sigma_i^{*2})$. For simplicity of notation we suppress the subscript i when possible.

Studies are published if $D=1$, which occurs with probability $p(Z^*)$, where $Z^*=X^*/\Sigma^*$. We observe the truncated sample of published studies (that is, we observe draws from the conditional distribution of (X^*,Σ^*) given $D=1$) and denote observations in this sample by (X,Σ). Publication decisions reflect both researcher and journal decisions; we do not attempt to disentangle the two. We obtain the following model:

Definition 1 (Truncated sampling process)

[3] Note that we use Σ to denote the (scalar) standard deviation rather than a variance matrix.

[4] The case where all latent studies estimate the same parameter is nested by taking the distribution μ_Θ to be degenerate.

$(\Theta^*,\Sigma^*,X^*,D)$ *are jointly i.i.d. across latent studies, with*

$$\begin{aligned}(\Theta^*,\Sigma^*) &\sim \mu_{\Theta,\Sigma}\\ X^*|\Theta^*,\Sigma^* &\sim N(\Theta^*,\Sigma^{*2})\\ D|X^*,\Theta^*,\Sigma^* &\sim Ber(p(Z^*)),\end{aligned}$$

where $Z^*=X^*/\Sigma^*$. *We observe i.i.d. draws* (X,Σ) *from the conditional distribution of* (X^*,Σ^*) *given* $D=1$. *Define* $Z=\frac{X}{\Sigma}$, $\Omega^*=\frac{\Theta^*}{\Sigma^*}$, $\Omega=\frac{\Theta}{\Sigma}$, *and denote the marginal distribution of* Θ^* *by* μ_Θ.

As we discuss in the proofs, many of our results can be extended to the case where X^* is non-normal. Our focus on the normal case is motivated by the fact that that X^* represents the final estimate in each study. Such estimates are approximately normal with a consistently estimable variance under mild conditions. Moreover, approximate normality of estimates is widely assumed in practice (for example to justify reporting standard errors), including in all the papers discussed in our applications.

The truncated sampling process of Definition 1 implies the likelihood.

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

for $\varphi(\cdot)$ the standard normal density. Note that $f_{Z|\Omega,\Sigma}(z|\omega,\sigma)=f_{Z|\Omega}(z|\omega)$. Moreover, the scale of the publication probability does not affect the distribution of published results, since for $c>0$, $p(\cdot)$ and $c\cdot p(\cdot)$ imply the same $f_{Z|\Omega}(z|\omega)$.

2.1 Illustrative example: Selection on statistical significance

To illustrate our setting we consider a simple example to which we will return throughout the paper. A journal receives a stream of studies reporting experimental estimates $X^*\sim N(\Theta^*,\Sigma^{*2})$ of treatment effects Θ^*, where each experiment examines a different treatment. The journal publishes studies with $Z^*=X^*/\Sigma^*$ in the interval $[-1.96,1.96]$ with probability $p(Z^*)=.1$, while results outside this interval are published with probability $p(Z^*)=1$. This publication policy reflects a preference for "significant results," where a two-sided z-test rejects the null hypothesis $\Theta^*=0$ (or equivalently, $\Omega^*=0$) at the 5% level. This journal is ten times more likely to publish significant results than insignificant ones. As a result, published results tend to over-estimate the magnitude of the treatment

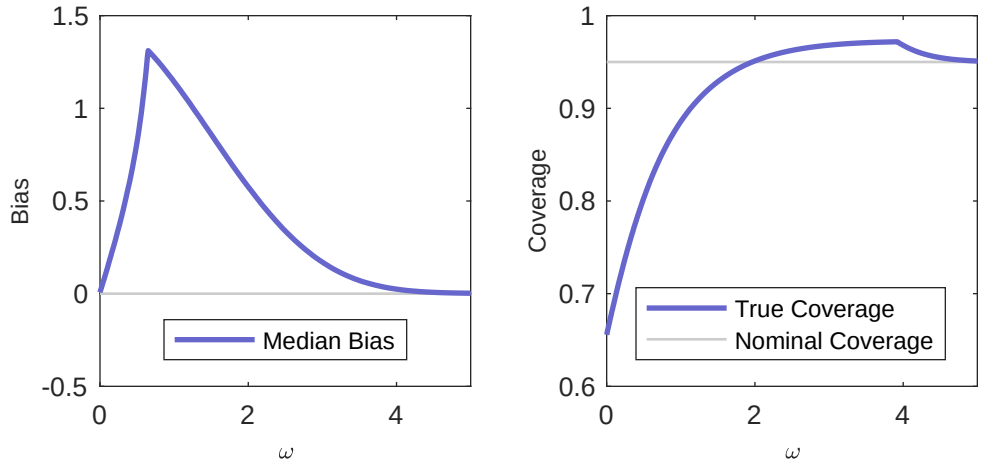

Figure 1: The left panel plots the median bias of the conventional estimator $\hat{\Theta}_j = Z_j$, while the right panel plots the true coverage of the conventional 95% confidence interval, both for $p(z) = .1 + .9 \cdot \mathbf{1}(|Z| > 1.96)$.

effect.[5] Published confidence intervals also under-cover the true parameter value for small values of Ω and over-cover for somewhat larger values. This is demonstrated by Figure 1, which plots the median bias, $med(\hat{\Omega}|\Omega = \omega) - \omega$, of the usual estimator $\hat{\Omega} = Z$, as well as the coverage of the conventional 95% confidence interval $[Z - 1.96, Z + 1.96]$.[6] While we have described this example in terms of selection by the journal, it could equivalently be interpreted as reflecting selection by researchers, or by both researchers and journals.

2.2 Corrected inference

If we know the form of selectivity we can correct the bias from selective publication. This section derives median unbiased estimators and valid confidence sets for Ω, which can immediately be turned into estimators and confidence sets for Θ via multiplication by Σ. These results ensure unbiasedness and correct coverage conditional on (Θ, Σ) for all (Θ, Σ), rather than just on average across the distribution of (Θ, Σ). For now we assume $p(\cdot)$ is known; corrections accounting for estimation error in $p(\cdot)$ are discussed in Section B.1 of the supplement.

Selective publication reweights the distribution of Z by $p(\cdot)$. To obtain valid estimators and confidence sets, we need to correct for this reweighting. To define these corrections, denote the distribution function for published results Z given true effect Ω

[5] See Ioannidis (2008) and Gelman (2018) for more discussion of this point.

[6] Note that $med(\hat{\Omega}|\Omega = \omega) - \omega = \frac{med(\hat{\Theta}|\Theta = \theta) - \theta}{\Sigma}$ so the median bias of $\hat{\Omega}$ can be interpreted as the median bias of X for θ, scaled by the standard error.

by $F_{Z|\Omega} = \int_{-\infty}^{z} f_{Z|\Omega}(\tilde{z}|\omega)d\tilde{z}$, for $f_{Z|\Omega}(z|\omega)$ as in Equation (1). We adapt an approach previously applied by, among others, D. Andrews (1993) and Stock and Watson (1998), and invert the distribution function as a function of ω to construct a quantile-unbiased estimator. Let us define $\hat{\omega}_\alpha(z)$ as the solution to

$$F_{Z|\Omega}(z|\hat{\omega}_\alpha(z)) = \alpha \in (0,1), \tag{2}$$

so z lies at the α-quantile of the distribution implied by $\hat{\omega}_\alpha(z)$. Using the monotonicity properties of $F_{Z|\Omega}$, we prove that $\hat{\omega}_\alpha(Z)$ is an α-quantile unbiased estimator for Ω.

Theorem 1

Suppose that $p(z) > 0$ for all z, and $p(\cdot)$ is almost everywhere continuous. Then $\hat{\omega}_\alpha(z)$ as defined in (2) exists, is unique, and is continuous and strictly increasing for all z. Furthermore, $\hat{\omega}_\alpha(Z)$ is α-quantile unbiased for Ω under the truncated sampling setup of Definition 1,

$$P(\hat{\omega}_\alpha(Z) \leq \omega | \Omega = \omega, \Sigma = \sigma) = \alpha \text{ for all } \omega.$$

These results allow straightforward frequentist inference that corrects for selective publication. In particular, using Theorem 1 we can consider the median-unbiased estimator $\hat{\omega}_{\frac{1}{2}}(z)$ for ω, as well as the equal-tailed level $1-\alpha$ confidence interval $\left[\hat{\omega}_{\frac{\alpha}{2}}(Z), \hat{\omega}_{1-\frac{\alpha}{2}}(Z)\right]$. This estimator and confidence set fully correct the bias and coverage distortions induced by selective publication. In the special case where insignificant results are published with probability zero while significant results are published with probability one, our corrected confidence sets exclude zero if and only if the test of McCrary et al. (2016) rejects.

Illustrative example (continued) To illustrate these results, we return to the treatment effect example discussed above. Figure 2 plots the median unbiased estimator, as well as upper and lower 95% confidence bounds, as a function of Z, again for the case with $p(Z^*) = 1$ when $|Z^*| > 1.96$ and $p(Z^*) = .1$ otherwise. We see that the median unbiased estimator lies below the usual estimator $\hat{\omega} = Z$ for small positive Z but that the difference is eventually decreasing in Z. The truncation-corrected confidence interval shown in Figure 2 has exactly correct coverage, is smaller than the usual interval for small Z, wider for moderate values Z, and essentially the same for $Z \geq 5$.

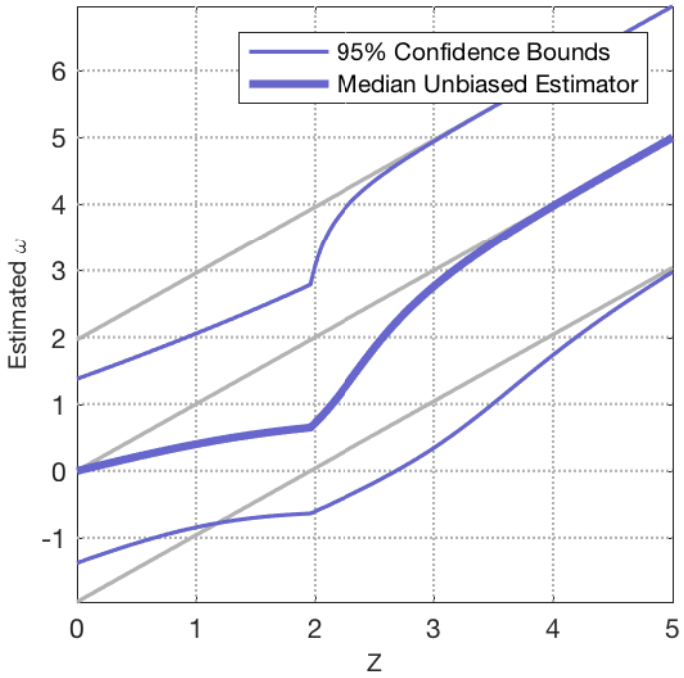

Figure 2: This figure plots 95% confidence bounds and the median unbiased estimator for the normal model where results that are significant at the 5% level are ten times more likely to be published than are insignificant results. The usual (uncorrected) estimator and confidence bounds are plotted in grey for comparison.

3 Identifying selection

This section proposes two approaches for identifying $p(\cdot)$. The first uses systematic replication studies, while the second uses meta-studies.

3.1 Systematic replication studies

The following theorem extends the model in Definition 1 above to incorporate a conditionally independent replication draw X^{*r} which is observed whenever X^* is. The key assumption for this theorem is that selectivity of publication operates only on X^* and not on X^{*r}. This assumption is plausible for systematic replication studies such as Open Science Collaboration (2015) and Camerer et al. (2016), but may fail in non-systematic replication settings, for instance if replication studies are published only when they "debunk" prior published results.

Theorem 2 (Nonparametric identification using replication experiments)

Consider the data generating process of Definition 1. Assume that for each latent study there exist a replication estimate and standard error (X^{*r},Σ^{*r}) *with*

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

where we again observe the replication estimate and standard error only for published studies. Then $p(\cdot)$ *is identified up to scale, and* μ_Θ *is identified as well.*

Intuition Consider the setup of Theorem 2, and define $Z^r = X^r/\Sigma$, that is as the *replication* estimate normalized by the *original* standard error. Assume for the moment that $\Sigma^{*r} = \Sigma^*$, so that the replication estimate X^{*r} has the same variance as X^*. Under these assumptions, the marginal density of (Z,Z^r) is

$$f_{Z,Z^r}(z,z^r) = \frac{p(z)}{E[p(Z^*)]} \int \varphi(z-\omega)\varphi(z^r-\omega)d\mu_\Omega(\omega). \tag{3}$$

This expression immediately implies that any asymmetries in the joint distribution of (Z,Z^r) must be due to the publication probability $p(\cdot)$. In particular,

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

whenever the denominators on either side are non-zero. Theorem 2 uses this identity to show that $p(\cdot)$ is nonparametrically identified up to scale.[7] That $p(\cdot)$ is only identified up to scale is intuitive: Equation (1) above shows that the scale of $p(\cdot)$ does not affect the distribution of published results, Equation (3) shows that the same remains true once we add replication results. Hence, the scale is both unnecessary for bias corrections and unidentified without data on unpublished results.

In general the replication standard error Σ^{r*} will differ from the original variance Σ^*, which takes us out of the symmetric framework. Additionally, the distribution of Σ^{r*} might depend on Z^*. Such dependence is present if power calculations are used to determine replication sample sizes, as in both Open Science Collaboration (2015) and Camerer et al. (2016). In that case, Σ^{*r} is positively related to the magnitude of Z^*, but conditionally unrelated to Θ^*. Identification carries over to this setting. To show this, the proof of Theorem 2 notes that we can recover the symmetric setting by (de)convolution of Z^r with normal noise.

[7]Note that this argument does not use normality of Z and Z^r, and thus generalizes to other estimator distributions.

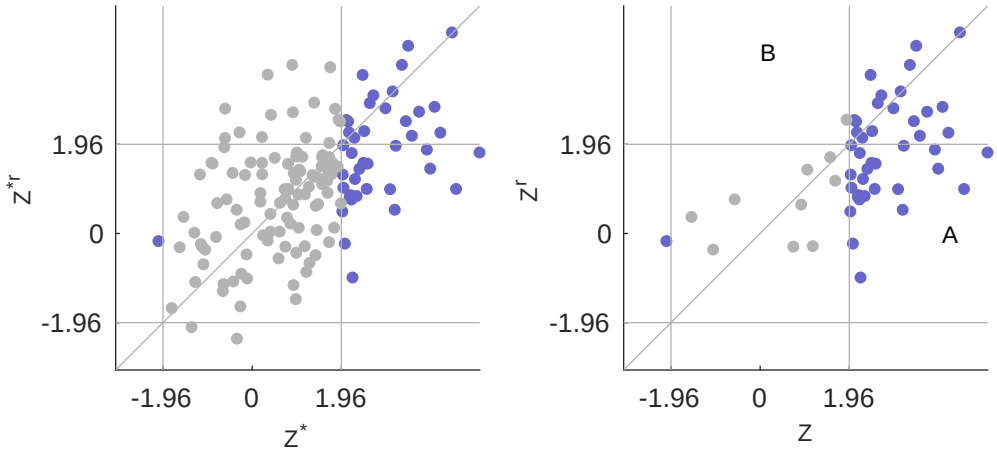

Figure 3: This figure illustrates the effect of selective publication in the replication experiments setting using simulated data, where selection is on statistical significance, as described in the text. The left panel shows the joint distribution of a random sample of latent estimates and replications; the right panel shows the subset which are published. Results where the original estimates are significantly different from zero at the 5% level are plotted in blue, while insignificant results are plotted in grey.

Illustrative example (continued) To illustrate our identification approach using replication studies, we return to the illustrative example introduced in Section 2. In this setting, suppose that the normalized true effect Ω^* is distributed $N(1,1)$ across latent studies. As before, assume that $p(Z^*)=1$ when $|Z^*|>1.96$, and that $p(Z^*)=.1$ otherwise. Assume finally that $\Sigma^{*r}=\Sigma^*=1$, so original and replication estimates both have variance one.

This setting is illustrated in Figure 3. The left panel of this figure shows 100 random draws (Z^*,Z^{*r}); draws where $|Z^*|\leq 1.96$ are marked in grey, while draws where $|Z^*|>1.96$ are marked in blue. The right panel shows the subset of draws (Z,Z^r) which are published. These are the same draws as (Z^*,Z^{*r}), except that 90% of the draws for which Z^* is statistically insignificant are deleted.

Our identification argument in this case proceeds by considering deviations from symmetry around the diagonal $Z=Z^r$. Let us compare what happens in the regions marked A and B. In A, Z is statistically significant but Z^r is not; in B it is the other way around. By symmetry of the data generating process, the latent (Z^*,Z^{*r}) fall in either area with equal probability. The fact that the observed (Z,Z^r) lie in region A substantially more often than in region B thus provides evidence of selective publication, and the exact deviation of the distribution of (Z,Z^r) from symmetry identifies $p(\cdot)$ up to scale.

3.2 Meta-studies

The following theorem extends the model in Definition 1 by assuming that Θ^* is statistically independent of Σ^* across latent studies, so studies with smaller standard errors do not have systematically different estimands. This is a strong assumption, but is imposed by many popular meta-analysis techniques including in meta-regression (see Section 5.2) and the "trim and fill" method (Duval and Tweedie, 2000). This assumption holds trivially if Θ^* is constant across latent studies. In our applications with replication data, estimates for $p(\cdot)$ based on this assumption are similar to those based on our replication approach, lending further support to this approach.

Theorem 3 (Nonparametric identification using meta-studies)
Consider the data generating process of Definition 1. Assume additionally that Σ^* and Θ^* are independent, and that the support of Σ contains an open interval. Then $p(\cdot)$ is identified up to scale, and μ_Θ is identified as well.

Intuition Consider the setup of Theorem 2. The conditional density of Z given Σ is

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

This implies that, for $\sigma_2 > \sigma_1$,

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

where the first term on the right hand side does not depend on z. Since $f_{Z|\Sigma}(z|\sigma_2)/f_{Z|\Sigma}(z|\sigma_1)$ is identified, this suggests we might be able to invert this equality to recover μ_Θ, which would then allow us to identify $p(\cdot)$. The proof of Theorem 3 builds on this idea.

Illustrative example (continued) As before, assume that Θ^* is $N(1,1)$ distributed, and that $p(Z^*)=1$ when $|Z^*|>1.96$, and that $p(Z^*)=.1$ otherwise. Suppose further that Σ^* is independent of Θ^* across latent studies. This setting is illustrated in Figure 4. The left panel of this figure shows 100 random draws (X^*,Σ^*); draws where $|X^*/\Sigma^*| \leq 1.96$ are marked in grey, while draws where $|X^*/\Sigma^*| > 1.96$ are marked in blue. The right panel shows the subset of draws (X,Σ) which are published, where 90% of statistically insignificant draws are deleted.

Compare what happens for two different values of the standard deviation Σ, marked by A and B in Figure 4. By the independence of Σ^* and Θ^*, the distribution of X^*

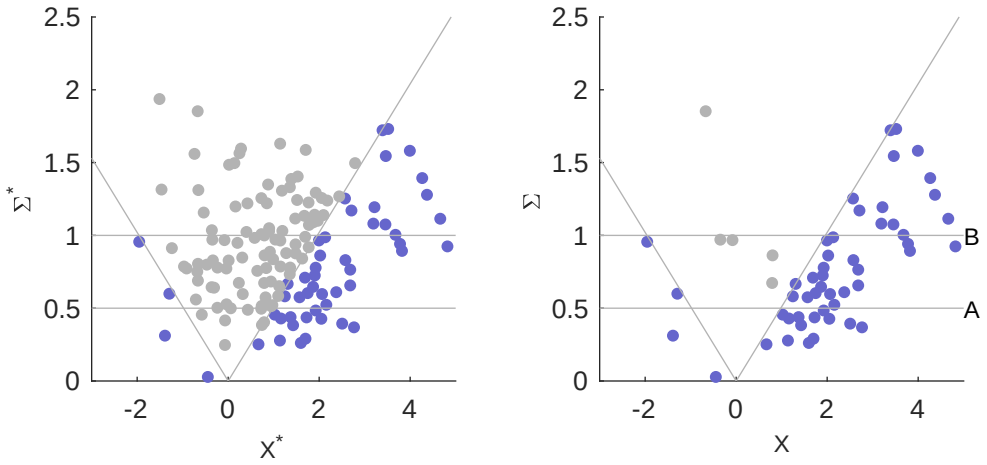

Figure 4: This figure illustrates the effect of selective publication in the meta-studies setting using simulated data, where selection is on statistical significance, as described in the text. The left panel shows a random sample of latent estimates; the right panel shows the subset of estimates which are published. Results which are significantly different from zero at the 5% level are plotted in blue, while insignificant results are plotted in grey.

for larger values of Σ^* is a noised up version of the distribution for smaller values of Σ^*. To the extent that the same does not hold for the distribution of published X given Σ, this must be due to selectivity in the publication process. In this example, statistically insignificant observations are "missing" for larger values Σ. Since publication is more likely when $|X^*/\Sigma^*| > 1.96$, the estimated values X tend to be larger on average for larger values of Σ, and the details of how the conditional distribution of X given Σ varies with Σ will again allow us to identify $p(\cdot)$ up to scale.

3.3 Estimation

The sample sizes in our applications are limited, which makes fully nonparameteric estimation impractical. In the supplement we build on our identification arguments to derive GMM estimators that assume a functional form for the conditional publication probability $p(\cdot)$ but are nonparametric in the distribution μ of true effects. For simplicity and ease of exposition, however, in the main text we specify parsimonious parametric models for both $p(\cdot)$ and μ which we fit by maximum likelihood, similar to Hedges (1992). Our nonparametric identification results suggest that there is hope for estimation robust to functional form assumptions, and this is borne out by the similarity of the maximum likelihood estimates reported here to the GMM results reported in the supplement.

We consider step function models for $p(\cdot)$, with jumps at conventional critical values, and possibly at zero. Since $p(\cdot)$ is only identified up to scale, we impose the normalization

$p(z)=1$ for $z>1.96$ throughout. This is without loss of generality, since $p(\cdot)$ is allowed to be larger than 1 for other cells. We assume different parametric models for the distribution of latent effects Θ^*, discussed case-by-case below. In our first two applications the sign of the original estimates is normalized to be positive.[8] We denote these normalized estimates by $W=|Z|$, and in these settings we impose that $p(\cdot)$ is symmetric.

4 Applications

This section applies the results developed above to estimate the degree of selectivity in three empirical literatures. We first consider data from the large scale replication studies Camerer et al. (2016) and Open Science Collaboration (2015), which examine experimental studies in economics and psychology, respectively. We then turn to the meta-study Wolfson and Belman (2015) on the effect of the minimum wage on employment. We consider two additional applications in the supplement, using replication data from Camerer et al. (2018) on social-science experiments and the meta-study data from Croke et al. (2016) on the effect of deworming.

Plausibility of identifying assumptions The results of Section 3 imply nonparametric identification of both $p(\cdot)$ and μ_Θ. Our approach using replication data is based on the assumption that selection for publication depends only on the original estimates and not on the replication estimates. This assumption is highly plausible by design in the two replication settings we consider, which use data from systematic replication studies. These studies pre-specify and replicate a large number of results published in a given time period and set of journals, and report all replication results together.

Our approach using meta-studies is based on the assumption that studies on a given topic with different standard errors do not have systematically different estimands. While we cannot guarantee validity of this assumption by design, its plausibility is enhanced by our finding that it yields estimates very similar to the approach based on replication studies in all applications we have analyzed (including Camerer et al. (2016), Open Science Collaboration (2015), and Camerer et al. (2018)). Variants of this assumption (or the strictly stronger assumption that Θ is constant) are common in existing meta-studies.

Finally, for both approaches we assume that conditional on (Θ^*,Σ^*) estimates are approximately normal, consistent with the inference methods used in the underlying studies.

[8] The studies in these datasets consider different outcomes, so the relative signs of effects across studies are arbitrary. Setting the sign of the initial estimate in each study to be positive ensures invariance to the sign normalization chosen by the authors of each study.

4.1 Economics laboratory experiments

Our first application uses data from a recent large-scale replication of experimental economics papers by Camerer et al. (2016). The authors replicated all 18 between-subject laboratory experiment papers published in the American Economic Review and Quarterly Journal of Economics between 2011 and 2014.[9] Further details on the selection and replication of results can be found in Camerer et al. (2016), while details on our handling of the data are discussed in the supplement.

A strength of this dataset for our purposes, beyond the availability of replication estimates, is the fact that it replicates results from all papers in a particular subfield published in two leading economics journals over a fixed period of time. This mitigates concerns about the selection of which studies to replicate. Moreover, since the authors replicate 18 such studies, it seems reasonable to think that they would have published their results regardless of what they found, consistent with our assumption that selection operates only on the initial studies and not on the replications.

A caveat to the interpretation of our results is that Camerer et al. (2016) select the most important statistically significant finding from each paper, as emphasized by the original authors, for replication. This selection changes the interpretation of $p(\cdot)$, which has to be interpreted as the probability that a result was published *and* selected for replication. In this setting, our corrected estimates and confidence intervals provide guidance for interpreting the headline results of published studies. For consistency with the rest of the paper, however, we continue to discuss $p(\cdot)$ as the publication probability.

Histogram Before we discuss our formal estimation results, consider the distribution of originally published estimates $W=|Z|$, shown by the histogram in the left panel of Figure 5. This histogram suggests a large jump in the density $f_W(\cdot)$ at the cutoff 1.96, and thus a corresponding jump in the publication probability $p(\cdot)$ at the same cutoff; see Section 5.3 below. Such a jump is confirmed by both our replication and meta-study approaches.

Results from replication specifications The middle panel of Figure 5 plots the joint distribution of $(W,W^r)=\text{sign}(Z)\cdot(Z,Z^r)$ in the replication data of Camerer et al. (2016).

[9] In their supplementary materials, Camerer et al. (2016) state that "To be part of the study a published paper needed to report at least one significant between subject treatment effect that was referred to as statistically significant in the paper." However, we have reviewed the issues of the American Economic Review and Quarterly Journal of Economics from the relevant period, and confirmed that no studies were excluded due to this restriction.

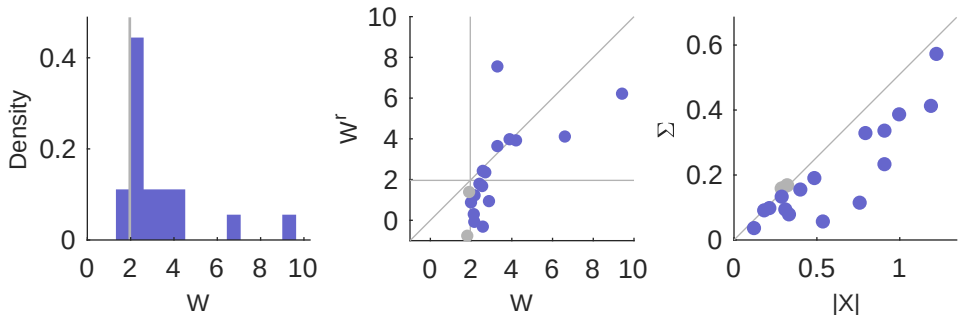

Figure 5: The left panel shows a binned density plot for the normalized z-statistics $W = |X|/\Sigma$ using data from Camerer et al. (2016). The grey line marks $W = 1.96$. The middle panel plots the z-statistics W from the initial study against the estimate W^r from the replication study. The grey lines mark W and $W^r = 1.96$, as well as $W = W^r$. The right panel plots the initial estimate $|X| = W \cdot \Sigma$ against its standard error Σ. The grey line marks $|X|/\Sigma = 1.96$.

To estimate the degree of selection in these data we consider the model

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

This assumes that the absolute value of the normalized true effect Ω^* follows a gamma distribution with shape parameter κ and scale parameter λ. This nests a wide range of cases, including χ^2 and exponential distributions, while keeping the number of parameters low. Our model for $p(\cdot)$ allows a discontinuity in the publication probability at $|Z| = 1.96$, the critical value for a 5% two-sided z-test. Fitting this model by maximum likelihood yields the estimates reported in the left panel of Table 1. Recall that β_p in this model can be interpreted as the publication probability for a result that is insignificant at the 5% level based on a two-sided z-test, relative to a result that is significant at the 5% level. These estimates therefore imply that significant results are more than thirty times more likely to be published than insignificant results. Moreover, we strongly reject the hypothesis of no selectivity, $H_0 : \beta_p = 1$.

Results from meta-study specifications While the Camerer et al. (2016) data include replication estimates, we can also apply our meta-study approach using just the initial estimates and standard errors. Since this approach relies on additional independence assumptions, comparing these results to those based on replication studies provides a useful check of the reliability of our meta-analysis estimates.

We begin by plotting the data used by our meta-analysis estimates in the right panel

Replication			Meta-study		
κ	λ	β_p	$\tilde{\kappa}$	$\tilde{\lambda}$	β_p
0.373	2.153	0.029	1.343	0.157	0.038
(0.267)	(1.026)	(0.027)	(1.310)	(0.076)	(0.051)

Table 1: Selection estimates from lab experiments in economics, with robust standard errors in parentheses. The left panel reports estimates from replication specifications, while the right panel reports results from meta-study specifications. Publication probability β_p is measured relative to the omitted category of studies significant at 5% level, so an estimate of 0.029 implies that results which are insignificant at the 5% level are 2.9% as likely to be published as significant results. The parameters (κ,λ) and $(\tilde{\kappa},\tilde{\lambda})$ are not comparable.

of Figure 5. We consider the model

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

noting that Θ^* is the mean of X^*, rather than Z^*, and thus that the interpretation of $(\tilde{\kappa},\tilde{\lambda})$ differs from that of (κ,λ) in our replication specifications. Fitting this model by maximum likelihood yields the estimates reported in the right panel of Table 1. Comparing these estimates to those in the left panel, we see that the estimates from the two approaches are similar, though the metastudy estimates suggest a somewhat smaller degree of selection. Hence, we find that in the Camerer et al. (2016) data we obtain similar results from our replication and meta-study specifications.

Bias correction To interpret our estimates, we calculate our median-unbiased estimator and confidence sets based on our replication estimate $\beta_p = .029$. Figure 6 plots the median unbiased estimator, as well as the original and adjusted confidence sets, for the 18 studies included in Camerer et al. (2016). Considering the first panel, which plots the median unbiased estimator along with the original and replication estimates, we see that the adjusted estimates track the replication estimates fairly well but are smaller than the original estimates in many cases. The second panel plots the original estimate and conventional 95% confidence set in blue, and the adjusted estimate and 95% confidence set in black. As we see from this figure, twelve of the adjusted confidence sets include zero, compared to just two of the original confidence sets. Hence, adjusting for the estimated degree of selection substantially changes the number of significant results in this setting.

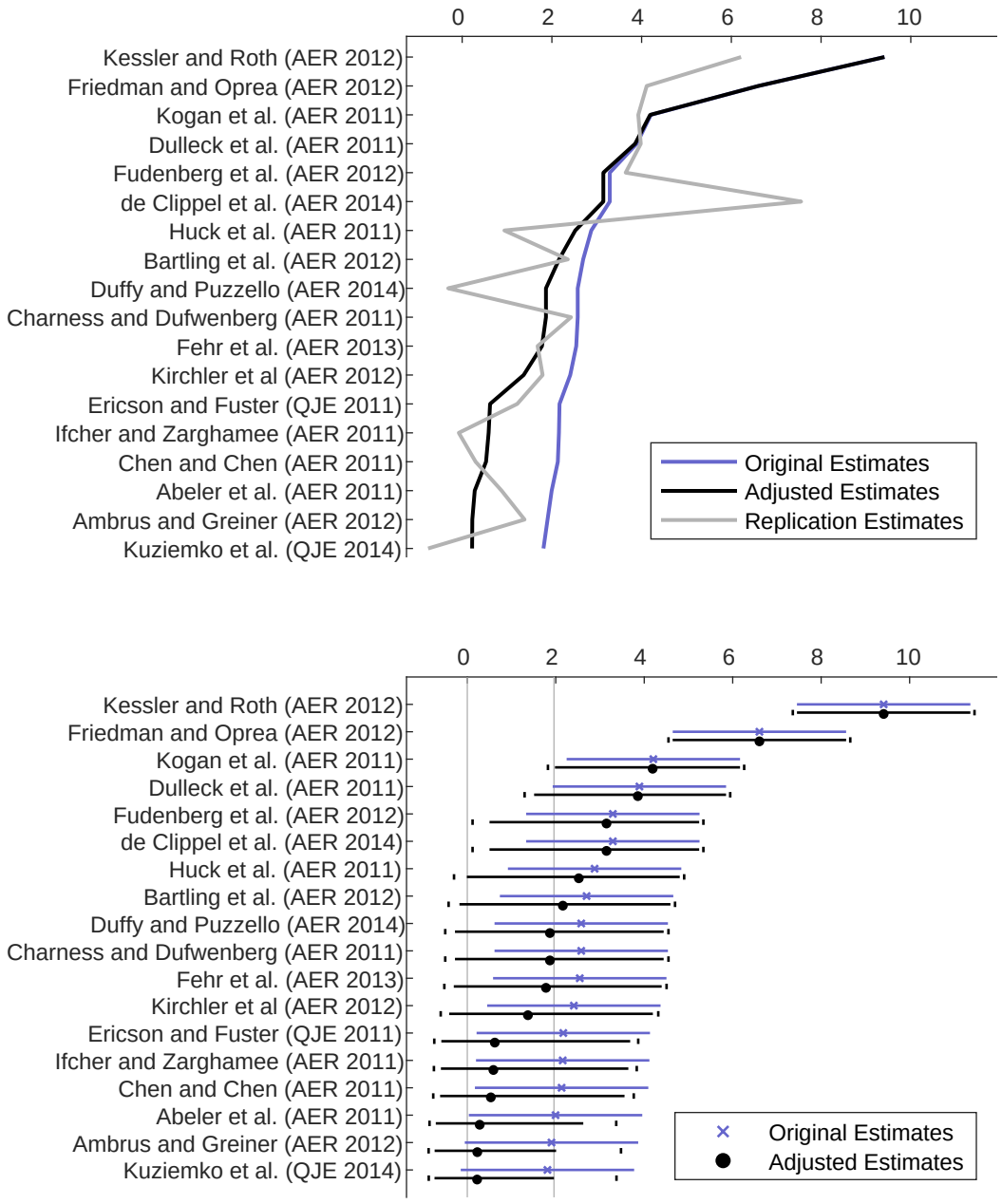

Figure 6: The top panel plots the estimates W and W^r from the original and replication studies in Camerer et al. (2016), along with the median unbiased estimate $\hat{\theta}_{\frac{1}{2}}$ based on the estimated selection model and the original estimate. The bottom panel plots the original estimate and 95% confidence interval, as well as the median unbiased estimate and adjusted 95% confidence interval $\left[\hat{\theta}_{0.025}(W),\hat{\theta}_{0.975}(W)\right]$ based on the estimated selection model. Adjusted intervals not accounting for estimation error in the selection model are plotted with solid lines, while endpoins for intervals accounting for estimation error are marked with "ı" – see Section B.1 of the supplement.

4.2 Psychology laboratory experiments

Our second application is to data from Open Science Collaboration (2015), who conducted a large-scale replication of experiments in psychology. The authors considered studies published in three leading psychology journals, Psychological Science, Journal of Personality and Social Psychology, and Journal of Experimental Psychology: Learning, Memory, and Cognition, in 2008. They assigned papers to replication teams on a rolling basis, with the set of available papers determined by publication date. Ultimately, 158 articles were made available for replication, 111 were assigned, and 100 of those replications were completed in time for inclusion in Open Science Collaboration (2015). Replication teams were instructed to replicate the final result in each article as a default, though deviations from this default were made based on feasibility and the recommendation of the authors of the original study. Ultimately, 84 of the 100 completed replications consider the final result of the original paper.

As with the economics replications above, the systematic selection of results for replication in Open Science Collaboration (2015) is an advantage from our perspective. A complication in this setting, however, is that not all of the test statistics used in the original and replication studies are well-approximated by z-statistics (for example, some of the studies use χ^2 test statistics with two or more degrees of freedom). To address this, we limit attention to the subset of studies which use z-statistics or close analogs thereof, leaving us with a sample of 73 studies. Specifically, we limit attention to studies using z- and t-statistics, or χ^2 and F-statistics with one degree of freedom (for the numerator, in the case of F-statistics), which can be viewed as the squares of z- and t-statistics, respectively. To explore sensitivity of our results to denominator degrees of freedom for t- and F-statistics, in the supplement we limit attention to the 52 observations with denominator degrees of freedom of at least 30 in the original study and find quite similar results.

Histogram The distribution of originally published estimates W is shown by the histogram in the left panel of Figure 7. This histogram suggests a large jump in the density $f_W(\cdot)$ at the cutoff 1.96, as well as possibly a jump at the cutoff 1.64, and thus of corresponding jumps of the publication probability $p(\cdot)$ at the same cutoffs. Such jumps are again confirmed by the estimates from both our replication and meta-study approaches.

Results from replication specifications The middle panel of Figure 7 plots the joint distribution of W, W^r in the replication data of Open Science Collaboration (2015). Relative to the plot for Camerer et al. (2016), we see a larger fraction of studies where

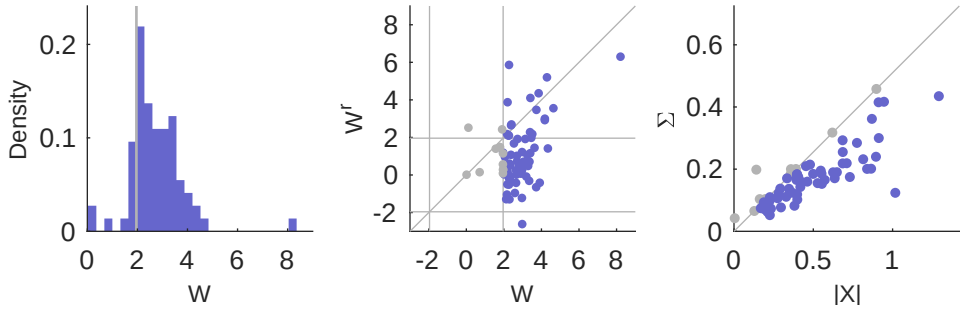

Figure 7: The left panel shows a binned density plot for the normalized z-statistics $W=|X|/\Sigma$ using data from Open Science Collaboration (2015). The grey line marks $W=1.96$. The middle panel plots the z-statistics W from the initial study against the estimate W^r from the replication study. The grey lines mark $|W|$ and $|W^r|=1.96$, as well as $W=W^r$. The right panel plots the initial estimate $|X|=W\cdot\Sigma$ against its standard error Σ. The grey line marks $|X|/\Sigma=1.96$.

Replication				Meta-study			
κ	λ	$\beta_{p,1}$	$\beta_{p,2}$	$\tilde{\kappa}$	$\tilde{\lambda}$	$\beta_{p,1}$	$\beta_{p,2}$
0.315	1.308	0.009	0.205	0.974	0.153	0.017	0.306
(0.141)	(0.331)	(0.005)	(0.088)	(0.549)	(0.053)	(0.009)	(0.135)

Table 2: Selection estimates from lab experiments in psychology, with robust standard errors in parentheses. The left panel reports estimates from replication specifications, while the right panel reports results from meta-study specifications. Publication probabilities β_p are measured relative to the omitted category of studies significant at 5% level. The parameters (κ,λ) and $(\tilde{\kappa},\tilde{\lambda})$ are not comparable.

$W>1.96$ for the original study while $W^r<1.96$ in the replication study (roughly 60% of studies in Open Science Collaboration (2015), compared to 44% in Camerer et al. (2016)). This could in principal be due to differences in selection, differences in the distribution of effects, or to other factors. To disentangle these issues, we fit the model

$$|\Omega^*|\sim\Gamma(\kappa,\lambda),\quad p(Z)\propto\begin{cases}\beta_{p,1} & |Z|<1.64\\ \beta_{p,2} & 1.64\leq|Z|<1.96\\ 1 & |Z|\geq 1.96.\end{cases}$$

This model again assumes that the absolute value of the normalized true effect $|\Omega^*|$ follows a gamma distribution across latent studies. Given the larger sample size, we consider a slightly more flexible model than before and allow discontinuities in the publication probability at the critical values for both 5% and 10% two-sided z-tests.

Fitting this model by maximum likelihood yields the estimates reported in the left panel of Table 2. These estimates imply that results that are significantly different from zero at the 5% level are over a hundred times more likely to be published than results that are insignificant at the 10% level, and nearly five times more likely to be published than results that are significant at the 10% level but insignificant at the 5% level. We strongly reject the hypothesis of no selectivity.[10]

Our results for this setting are roughly consistent with those of Johnson et al. (2017), who independently consider the Open Science Collaboration (2015) data and likewise estimate a step function model for $p(\cdot)$, but allow a discontinuity only at the 5% significance level. Johnson et al. (2017) estimate that insignificant results are only about 0.5% as likely to be published as are significant results. The Johnson et al. specifications for μ allow the possibility that $Pr\{\Theta^* = 0\} > 0$ and they estimate that $\Theta^* = 0$ about 90% of the time. Similarly, our estimated gamma distribution has mode equal to zero.

Results from meta-study specifications As before, we re-estimate our model using our meta-study specifications, and plot the joint distribution of estimates and standard errors in the right panel of Figure 7. Fitting the model yields the estimates reported in the right panel of Table 2. As in the last section, we find that the meta-study and replication estimates are broadly similar, though the meta-study estimates again suggest a somewhat more limited degree of selection.[11]

Approved replications Gilbert et al. (2016) argue that the protocols in some of the Open Science Collaboration (2015) replications differed substantially from the initial studies. These arguments were disputed by many of the Open Science Collaboration (2015) authors in Anderson et al. (2016), who note that many of the replications used protocols approved in advance by the authors of the underlying papers. In Section B.6.2 of the supplement we report results based on the subset of approved replications and find roughly similar estimates, though the estimated degree of selection is smaller.

Bias corrections To interpret our results, we plot our median-unbiased estimates based on the Open Science Collaboration (2015) data in Figure 8. We see that our

[10] If, as in the Camerer et al. (2016) application, we instead estimate the model only with a discontinuity at the 5% level, we estimate that results significant at the 5% level are over forty times more like to be published than insignificant results. This is a larger degree of selection than we estimate in Camerer et al. (2016), though the difference is not significant at conventional levels.

[11] If we instead estimate the model only with a discontinuity at the 5% level, we estimate a somewhat smaller degree of selectivity than in the metastudy specifications for Camerer et al. (2016), $\beta_p = 5.71\%$, though the difference is again insignificant.

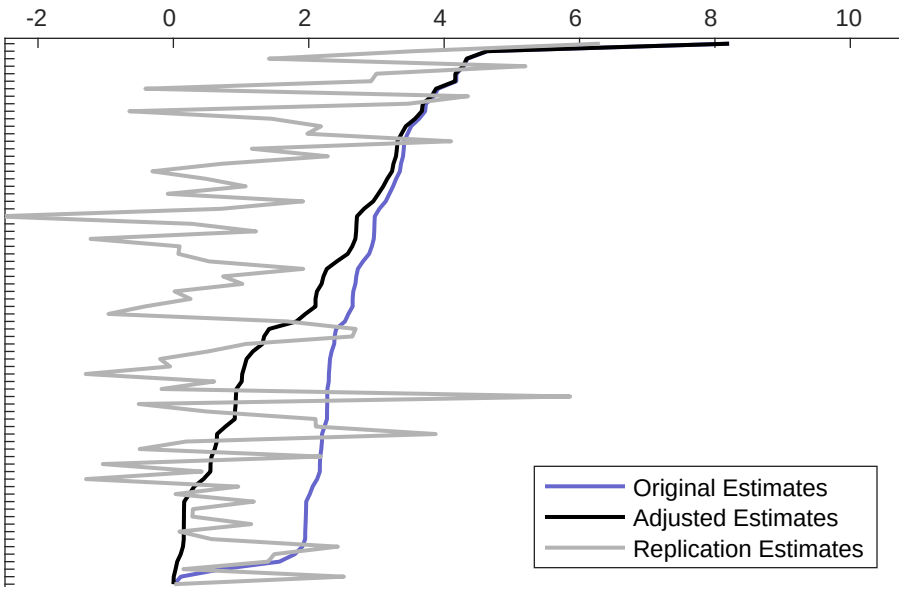

Figure 8: This figure plots the estimates W and W^r from the original and replication studies in Open Science Collaboration (2015), along with the median unbiased estimate $\hat{\theta}_{\frac{1}{2}}$ based on the estimated selection model and the original estimate.

adjusted estimates track the replication estimates fairly well for studies with small original z-statistics, though unlike in Figure 6 differences are larger for studies with larger original z-statistics. Note, however, that we have sorted on the original results, so patterns of this sort can arise purely from mean reversion.

Our adjustments again dramatically change the number of significant results, with 62 of the 73 original 95% confidence sets excluding zero, and only 28 of the adjusted confidence sets (not displayed) doing the same.

4.3 Effect of minimum wage on employment

Our final application uses data from Wolfson and Belman (2015), who conduct a meta-analysis of studies on the elasticity of employment with respect to the minimum wage. In particular, Wolfson and Belman (2015) collect analyses of the effect of minimum wages on employment that use US data and were published or circulated as working papers after the year 2000. They collect estimates from all studies fitting their criteria that report both estimated elasticities of employment with respect to the minimum wage and standard errors, resulting in a sample of a thousand estimates drawn from 37 studies, and

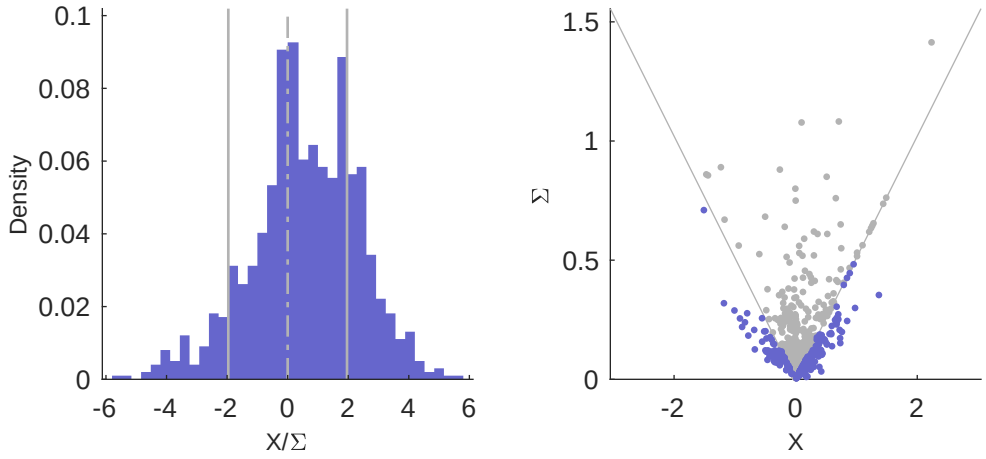

Figure 9: The left panel shows a binned density plot for the z-statistics X/Σ in the Wolfson and Belman (2015) data. The solid grey lines mark $|X|/\Sigma = 1.96$, while the dash-dotted grey line marks $X/\Sigma = 0$. The right panel plots the estimate X against its standard error Σ. The grey lines mark $|X|/\Sigma = 1.96$.

we use these estimates as the basis of our analysis. For further discussion of these data, see Wolfson and Belman (2015).

Since the Wolfson and Belman (2015) sample includes both published and unpublished papers, we evaluate our estimators based on both the full sample and the sub-sample of published estimates. We find qualitatively similar answers for the two samples, so we report results based on the full sample here and discuss results based on the subsample of published estimates in the supplement. We define X so that $X > 0$ indicates a negative effect of the minimum wage on employment.

Histogram Consider first the distribution of the normalized estimates Z, shown by the histogram in the left panel of Figure 9. This histogram is somewhat suggestive of jumps in the density $f_Z(\cdot)$ around the cutoffs -1.96, 0, and 1.96, and thus of corresponding jumps in the publication probability $p(\cdot)$ at the same cutoffs; these jumps seem less pronounced than in our previous applications, however.

Results from meta-study specifications For this application we do not have any replication estimates, and so move directly to our meta-study specifications. The right panel of Figure 9 plots the joint distribution of X, the estimated elasticity of employment with respect to decreases in the minimum wage, and the standard error Σ in the Wolfson and Belman (2015) data.

We next consider the model

$$\Theta^* \sim \bar{\theta} + t(\nu) \cdot \tilde{\tau}, \quad p(Z) \propto \begin{cases} \beta_{p,1} & Z < -1.96 \\ \beta_{p,2} & -1.96 \leq Z < 0 \\ \beta_{p,3} & 0 \leq Z < 1.96 \\ 1 & Z \geq 1.96. \end{cases}$$

Since the data are not sign-normalized, we model Θ^* using a t distribution with degrees of freedom $\tilde{\nu}$ and location and scale parameters $\bar{\theta}$ and $\tilde{\tau}$, respectively. Unlike in our previous applications, we allow the probability of publication to depend on the sign of the z-statistic Z rather than just on its absolute value. This is important, since it seems plausible that the publication prospects for a study could differ depending on whether it found a ($X<0$) positive or negative ($X>0$) effect of the minimum wage on employment.

Our estimates based on these data are reported in Table 3, where we find that results which are insignificant at the 5% level are about 30% as likely to be published as are significant estimates finding a negative effect of the minimum wage on employment. Our point estimates also suggest that studies finding a positive and significant effect of the minimum wage on employment may be less likely to be published, but this estimate is quite noisy and we cannot reject the hypothesis that selection depends only on significance and not on sign. Unlike our other results, this is sensitive to the details of the specification: if we instead restrict the distribution of true effects Θ^* to be normal, our estimate for $\beta_{p,1}$ drops to 0.225 with a standard error of 0.118. On the other hand, our GMM approach discussed in Section C.1 of the supplement returns a $\beta_{p,1}$ estimate of 1.174 with a standard error of 0.417.

$\bar{\theta}$	$\tilde{\tau}$	$\tilde{\nu}$	$\beta_{p,1}$	$\beta_{p,2}$	$\beta_{p,3}$
0.018	0.019	1.303	0.697	0.270	0.323
(0.009)	(0.011)	(0.279)	(0.350)	(0.111)	(0.094)

Table 3: Meta-study estimates from minimum wage data, with standard errors clustered by study in parentheses. Publication probabilities β_p measured relative to omitted category of estimates positive and significant at 5% level.

Since the studies in this application estimate related parameters, it is also interesting to consider the estimate $\bar{\theta}$ for the mean effect in the population of latent estimates. The point estimate is small but significantly different from zero at the 5% level, and suggests that the average latent study finds a small negative effect of the minimum wage on

employment. This effect is about half as large as the "naive" average effect $\bar{\theta}$ we would estimate by ignoring selectivity, .041 with a standard error of 0.011.

These results are consistent with the meta-analysis estimates of Wolfson and Belman (2015), who found evidence of some publication bias towards a negative employment effect, as well as the results of Card and Krueger (1995), who focused on an earlier, non-overlapping set of studies.

Multiple estimates A complication arises in this application, relative to those considered so far, due to the presence of multiple estimates per study. Since it is difficult to argue that a given estimate in each of these studies constitutes the "main" estimate, restricting attention to a single estimate per study would be arbitrary. This somewhat complicates inference and identification.

For inference, it is implausible that estimate standard-error pairs (X,Σ) are independent within study. To address this, we cluster our standard errors by study.

For identification, the problem is somewhat more subtle. Our model assumes that the latent parameters Θ_i^* and Σ_i^* are statistically independent across estimates i, and that D_i is independent of (Θ_i^*,Σ_i^*) conditional on Z_i^*. It is straightforward to relax the assumption of independence across i, provided the marginal distribution of $(\Theta_i^*,\Sigma_i^*,X_i^*,D_i)$ is such that D_i remains independent of (Θ_i^*,Σ_i^*) conditional on Z_i^*. This conditional independence assumption is justified if we believe that both researchers and referees consider the merits of each estimate on a case-by-case basis, and so decide whether or not to publish each estimate separately. Alternatively, it can also be justified if the estimands Θ^* within each study are statistically independent (relative to the population of estimands in the literature under consideration).

5 Alternative approaches

Many approaches to detecting selectivity and publication bias have been proposed in the literature. Good reviews are provided by Rothstein et al. (2006) and Christensen and Miguel (2016). In this section we analyze some of these approaches through the lens of our framework and relate them to our results.

5.1 Should results "replicate?"

The findings of recent systematic replication studies such as Open Science Collaboration (2015) and Camerer et al. (2016) are sometimes interpreted as indicating an inability to "replicate the results" of published research. In this setting, a "result" is understood to

"replicate" if both the original study and its replication find a statistically significant effect in the same direction. The share of results which replicate in this sense is prominently discussed in Camerer et al. (2016). Our framework shows that the probability of replication in this sense might be low even without selective publication or other sources of bias.

Consider the setup for replication experiments of Theorem 2, with constant publication probability $p(\cdot)$, so that publication is not selective and $f_{Z,Z^r} = f_{Z^*,Z^{r*}}$. For illustration, assume further that $\Sigma^* = \Sigma^{r*}$ with probability 1. For Φ the standard normal distribution function, the probability that a result replicates in the sense described above is

$$P(Z^{*r} \cdot sign\{Z^*\} > 1.96 || Z^*| > 1.96) = \frac{\int [\Phi(-1.96-\omega)^2 + \Phi(-1.96+\omega)^2] d\mu_\Omega(\omega)}{\int [\Phi(-1.96-\omega) + \Phi(-1.96+\omega)] d\mu_\Omega(\omega)}.$$

If the true effect is zero in all studies then this probability is 0.025. If the true effect in all studies is instead large, so that $|\Omega^*| > M$ with probability one for some large M, then the probability of replication is approximately one. Thus, any replication probability between 0.025 and one is consistent with no selection, and low replication frequencies are not necessarily indicative of selective publication; they could instead be due to a large share of small true effects. Strengths and weaknesses of alternative measures of replication are discussed in Simonsohn (2015) and Patil and Peng (2016).

5.2 Meta-regressions

A popular test for publication bias in meta-studies (cf. Card and Krueger, 1995; Egger et al., 1997) is based on meta-regression, which uses regressions of either of the following forms:

$$E^*[X|1,\Sigma] = \gamma_0 + \gamma_1 \cdot \Sigma, \quad E^*\left[Z|1,\tfrac{1}{\Sigma}\right] = \beta_0 + \beta_1 \cdot \tfrac{1}{\Sigma},$$

where we use E^* to denote best linear predictors. Under the assumptions of Theorem 3, if $p(\cdot)$ is constant then it follows immediately that

$$E^*[X|1,\Sigma] = E[\Theta^*], \quad E^*\left[Z|1,\tfrac{1}{\Sigma}\right] = E[\Theta^*] \cdot \tfrac{1}{\Sigma}.$$

Hence, testing that either $\gamma_1 = 0$ or $\beta_0 = 0$ delivers a valid test for the null hypothesis of no selectivity, though there are some forms of selectivity against which such tests have no power.

Absent publication bias, β_1 and γ_0 recover the average of Θ^* in the population of latent studies. These coefficients are sometimes interpreted as selection-corrected estimates of the mean effect across studies (cf. Doucouliagos and Stanley, 2009; Christensen and Miguel,

2016), but this interpretation is potentially misleading in the presence of publication bias. In particular, the conditional expectation $E[X|1,\Sigma]$ is nonlinear in both Σ and $1/\Sigma$, which implies that β_0, γ_1 are generally biased as estimates of $E[\Theta^*]$.[12] We discuss a simple example with one-sided significance testing in Section D.1 of the supplement.

A variety of generalizations to meta-regression have been proposed in the literature, including by Stanley and Doucouliagos (2014), who propose to use power-weighted meta-regressions to increase robustness to selective publication, and Stanley et al. (2017) who consider non-linear meta-regressions. Meta-regressions have also been widely used in applications, including by Carter et al. (2017), Havránek (2015), and Ioannidis et al. (2017).[13]

5.3 The distribution of p-values and z-statistics

Another approach in the literature considers the distribution of p-values, or the corresponding z-statistics, across published studies (cf. De Long and Lang, 1992; Schuemie et al., 2014; Simonsohn et al., 2014; Brodeur et al., 2016, 2018). Assuming normality, there is a one-to-one mapping between the distribution of p-values P and the distribution of z-statistics Z, since $P = 1 - \Phi(Z)$ for 1-sided tests of the null hypothesis $\theta = 0$ or, equivalently, $\omega = 0$.[14] Under our model, absent selectivity in the publication process the distribution f_Z is equal to f_{Z^*}. For $Z^*|\Omega^* \sim N(\Omega^*,1)$ and $\Omega^* \sim \mu_\Omega$, this implies that

$$f_Z(z) = f_{Z^*}(z) = (\mu_\Omega * \varphi)(z) = \int \varphi(z-\omega) d\mu_\Omega(\omega).$$

This model implies that the density f_{Z^*} is infinitely differentiable. If selectivity is present, by contrast, then $f_Z(z) = \frac{p(z)}{E[p(Z^*)]} \cdot f_{Z^*}(z)$. Any discontinuity of $f_Z(z)$ (for instance at critical values such as $z = 1.96$) thus identifies a corresponding discontinuity of the conditional publication probability $p(z)$:

$$\frac{\lim_{z \downarrow z_0} f_Z(z)}{\lim_{z \uparrow z_0} f_Z(z)} = \frac{\lim_{z \downarrow z_0} p(z)}{\lim_{z \uparrow z_0} p(z)}. \tag{4}$$

If we impose that $p(\cdot)$ is a step function, this identifies $p(\cdot)$ up to scale.

The model without selectivity, $f_Z(z) = f_{Z^*}(z) = (\mu_\Omega * \varphi)(z)$, has testable implications

[12] Stanley (2008) and Doucouliagos and Stanley (2009) note this bias but suggest that one can still use $H_0 : \gamma_1 = 0$ to test the hypothesis of zero true effect if there is no heterogeneity in the true effect Θ^* across latent studies.

[13] Other recent work examining selective publication in economics and finance using non meta-regression approaches includes Chen and Zimmermann (2017) and Hou et al. (2017).

[14] For two-sided tests, the mapping is between p-values and absolute z-statistics $|Z|$.

beyond smoothness. In particular, the density f_{Z^*} precludes excessive bunching, since for all $k \geq 0$ and all z, $\partial_z^k f_{Z^*}(z) \leq \sup_z \partial_z^k \varphi(z)$ and $\partial_z^k f_{Z^*}(z) \geq \inf_z \partial_z^k \varphi(z)$ so for example $f_{Z^*}(z) \leq \varphi(0)$ and $f''_{Z^*}(z) \geq \varphi''(0) = -\varphi(0)$ for all z. Spikes in the distribution of Z thus likewise indicate the presence of selectivity or inflation.

5.4 Observability

The setup of Definition 1 assumes that we only observe the draws (X^*,Σ^*) for which $D=1$. In some cases, however, additional information may be available. First, we might know of the existence of unpublished studies, for example from experimental preregistrations, without observing their results X^*. In this case, called censoring, we observe i.i.d. draws of (Y,D), where $Y = D \cdot Z^*$.[15] The corresponding censored likelihood is

$$f_{Y,D|\Omega^*}(y,d|\omega^*) = d \cdot p(y) \cdot \varphi(y-\omega) + (1-d) \cdot (1 - E[p(Z^*)|\Omega^* = \omega^*]).$$

Second, we might additionally observe the results Z^* from unpublished working papers as in Franco et al. (2014). The likelihood in this case is

$$f_{Z^*,D|\Omega^*}(z,d|\omega) = p(z)^d (1-p(z))^{1-d} \cdot \varphi(z-\omega).$$

Even under these alternative observability assumptions, the truncated likelihood (1) arises as a limited information likelihood that conditions on publication decisions and/or unpublished results. Our identification and inference results therefore continue to apply.

That said, additional information allows identification of $p(\cdot)$ under weaker assumptions. With full observability of unpublished results Z^*, for example, $p(\cdot)$ is identified by simply regressing D on Z^*, cf. Franco et al. (2014).

5.5 Bias and Pseudo-True Values

Bruns and Ioannidis (2016) and Bruns (2017) discuss an additional way in which selectivity may increase bias in observational studies. To cast their concern into our framework, recall that we assume throughout that the distribution of X^* in latent studies is normal and centered on Θ^*, and that the publication decision depends only on Z^* (and possibly the standard error Σ^*). There are different ways this model can be interpreted.

A first interpretation, in line with our exposition in Section 2, is that Θ_i^* is the "true" parameter of interest in study i. This would for example be the case for randomized

[15] We could also observe the standard error Σ for published studies, but suppress this for simplicity.

experiments where we have no reason to doubt the internal validity of each study. In this case any variation of Θ_i^* across studies i considering the same question is due to issues of external validity, for instance to different populations of experimental subjects, or to effects changing over time. In this setting our corrections yield valid estimates and confidence sets for the parameters of interest.

A second interpretation of our model is that researchers consider different estimates X^* of the same parameter. These estimates might for instance be based on different controls, different outcome variables, different estimation methods, and so on. These estimates have expectations Θ^* that vary across specifications, so not all Θ^* correspond to the "true" effect of interest. Put differently, variation of Θ^* across studies might be due to violations of internal validity, in addition to issues of external validity.[16] Under this second interpretation, we have additional sources of bias. First, $E[\Theta] \neq E[\Theta^*]$ in general, so selection can lead to different average biases among published and latent studies. This effect can persist even as sampling noise goes to zero.[17] Second, even if we avoid this bias by using our approach to identify μ_Θ and therefore $E[\Theta^*]$, there is no guarantee that $E[\Theta^*]$ corresponds to the parameter of interest. Hence, while our corrections can undo selection bias and allow inference on either the parameter Θ in a given study or the distribution μ_Θ of Θ^* in the population of latent studies, we cannot correct deficiencies in the underlying studies.

5.6 Manipulation and P-hacking

Some authors consider the possibility that researchers manipulate their results (Brodeur et al., 2016; Furukawa, 2017), while others consider the selection of results within papers, which Simonsohn et al. (2014) term "p-hacking." Our primary focus in this paper is on researchers decisions whether or not to submit findings, and journal decisions whether or not to publish submissions, rather than on manipulation or p-hacking. Nonetheless, depending on the form manipulation or p-hacking takes, it may still be consistent with our baseline model.

To illustrate, consider an experimental setting where researchers run two independent versions of an experiment, or estimate two regression specifications for the same estimand. Suppose first that they decide whether to report an estimate for each experiment or specification separately. In this case our baseline model applies, save that Θ_i^* is no longer i.i.d. Suppose now alternatively that the researcher decides to always report only

[16] If some studies are viewed as more credible than others, this highlights the value of conducting inference on Θ for individual studies, rather than merely on the distribution μ_Θ.

[17] Consider for instance the case where $E[\Theta^*]=0$ and positive results are more likely to be published.

the larger of the two estimates. In this case, the probability of publication of the first estimate depends on the underlying parameter via the second estimate, i.e., publication probabilities are of the form $p(Z_i^*,\Theta_i^*)$.

To accommodate such violations of our baseline model, we discuss the extension of our approach to settings where the selection probability may depend on both Z and Θ in Section D.3 of the supplement. Given normal replication estimates X^r, we show that in this setting we can still identify enough features of the model to apply selection-corrections. We also develop specification tests for our baseline model against this more general alternative, however, and in no case do we reject our baseline model where $p(\cdot)$ does not depend on Θ given Z.

6 Conclusion

This paper makes three contributions relative to the existing literature. First, we provide methods to calculate bias-corrected estimators and confidence sets when the form of selectivity is known. Second, we provide nonparametric identification results for selectivity based on replications and meta-studies. Third, we apply the proposed methods to several literatures, documenting the varying scale and kind of selectivity. In cases where both our replication and meta-study approaches apply, they yield similar conclusions.

Implications for empirical research What can researchers and readers of empirical research take away from this paper? First, when conducting a meta-analysis of the findings of some literature, researchers may wish to apply our methods to assess the degree of selectivity in this literature, and to apply appropriate corrections to individual estimates, tests, and confidence sets. We provide code on our webpages which implements the proposed methods for a flexible family of selection models.[18]

Second, our results provide guidance for how to interpret published empirical results. In particular, if a reader has a view about how the selection process operates in a given literature, they can adjust published estimates and confidence sets as discussed in Section 4. Even if one is concerned that the selection model does not capture all sources of bias, these corrections aid interpretation by showing how much selection, considered in isolation, changes the interpretation of published results. A positive message from our results is that published estimates remain informative even when publication is quite selective.

It should be emphasized that we do not advocate adjusting publication standards to

[18]We have also implemented out meta-study approach in a web app: https://maxkasy.github.io/home/metastudy/

reflect our corrected critical values. If these cutoffs were to be systematically used in the publication process, this would simply entail an "arms race" of selectivity, rendering the more stringent critical values invalid again.

Optimal publication rules One might take the findings in this paper, and the debate surrounding publication bias more generally, to indicate that the publication process should be non-selective with respect to findings. This might for instance be achieved by instituting some form of result-blind review (see for example American Society of Health Economists (2015)) The hope would be that non-selectivity of the publication process might restore the validity (unbiasedness, size control) of standard inferential methods.

Note, however, that optimal publication rules may depend on results. This is for instance the case in models where policy decisions are made based on published findings. Section D.6 in the supplement provides a stylized example of such a setting. Alternatively, given evidence that experts can forecast experimental results quite well (cf. DellaVigna and Pope, 2018), excessively surprising findings might be interpreted as evidence of implementation problems and so weigh against publication. A broader study of the question of optimal publication from a journal's perspective can be found in Frankel and Kasy (2018).

Supplement The supplement contains a wide variety of results to complement those discussed in the main text. Section A provides proofs, while Section B gives additional details for our empirical applications and considers a range of robustness checks, including allowing publication probabilities to depend on covariates such as the journal or the year in which a paper was initially circulated. Section C derives novel GMM estimation approaches that leave the distribution of true effects unrestricted, and reports results for our applications. Section C also reports ML estimates for the Croke et al. (2016) and Camerer et al. (2018) applications. Finally, Section D reports additional theoretical results, including extensions of our identification results to allow publication probabilities to depend on Σ (e.g. to reflect a preference for precise estimates) and on Ω (to nest violations of our baseline model). This section also extends our inference results to cases where selection is driven by multiple variables, and discusses the effect of selection on Bayesian inference.

References

American Society of Health Economists (2015). Editorial statement on negative findings. https://www.ashecon.org/american-journal-of-health-economics/.

Anderson, C. J., Bahník, Š., Barnett-Cowan, M., Bosco, F. A., Chandler, J., Chartier, C. R., ..., and Zuni, K. (2016). Response to comment on estimating the reproducibility of psychological science. *Science*, 351(6277):1037–1037.

Andrews, D. W. (1993). Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica*, 61(1):139–165.

Brodeur, A., Cook, N., and Heyes, A. (2018). Methods matter: P-hacking and causal inference in economics. Working Paper.

Brodeur, A., Lé, M., Sangnier, M., and Zylberberg, Y. (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics*, 8(1):1–32.

Bruns, S. B. (2017). Meta-regression models and observational research. *Oxford Bulletin of Economics and Statistics*, 79(5):637–653.

Bruns, S. B. and Ioannidis, J. P. (2016). P-curve and p-hacking in observational research. *PLoS One*, 11(2):e0149144.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., and Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280):1433–1436.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., Isaksson, S., Manfredi, D., Rose, J., Wagenmakers, E.-J., and Wu, H. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behaviour*, 2(9):637–644.

Card, D. and Krueger, A. B. (1995). Time-series minimum-wage studies: A meta-analysis. *American Economic Review*, 85(2):238–243.

Carter, E., Schönbrodt, F., Gervais, W. M., and Hilgard, J. (2017). Correcting for bias in psychology: A comparison of meta-analytic methods. Unpublished Manuscript.

Chang, A. C. and Li, P. (2018). Is economics research replicable? sixty published papers from thirteen journals say" usually not". *Critical Finance Review*, 7.

Chen, A. Y. and Zimmermann, T. (2017). Selection bias and the cross-section of expected returns. Unpublished Manuscript.

Christensen, G. S. and Miguel, E. (2016). Transparency, reproducibility, and the credibility of economics research. NBER Working Paper No. 22989.

Clemens, M. A. (2017). The meaning of failed replications: a review and proposal. *Journal of Economic Surveys*, 31(1):326–342.

Croke, K., Hicks, J. H., Hsu, E., Kremer, M., and Miguel, E. (2016). Does mass deworming affect child nutrition? Meta-analysis, cost-effectiveness, and statistical power. Technical Report 22382, National Bureau of Economic Research.

De Long, J. B. and Lang, K. (1992). Are all economic hypotheses false? *Journal of Political Economy*, 100(6):1257–1272.

DellaVigna, S. and Pope, D. (2018). Predicting experimental results: Who knows what? *Journal of Political Economy*. Forthcoming.

Doucouliagos, H. and Stanley, T. (2009). Publication selection bias in minimum-wage research? A meta-regression analysis. *British Journal of Industrial Relations*, 47(2):406–428.

Duval, S. and Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95(449):89–98.

Egger, M., Smith, G. D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, 315(7109):629–634.

Franco, A., Malhotra, N., and Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203):1502–1505.

Frankel, A. and Kasy, M. (2018). Working paper. Which Findings Should Be Published?

Furukawa, C. (2017). Unbiased publication bias: Theory and evidence. Unpublished Manuscript.

Gelman, A. (2018). The failure of null hypothesis significance testing when studying incremental changes, and what to do about it. *Personality and Social Psychology Bulletin*, 44(1):16–23. Unpublished Manuscript.

Gertler, P., Galiani, S., and Romero, M. (2018). How to make replication the norm. *Nature*, 554:417–419.

Gilbert, D. T., King, G., Pettigrew, S., and Wilson, T. D. (2016). Comment on "Estimating the reproducibility of psychological science". *Science*, 351(6277):1037.

Havránek, T. (2015). Measuring intertemporal substitution: The importance of method choices and selective reporting. *Journal of the European Economic Association*, 13(6):1180–1204.

Hedges, L. V. (1992). Modeling publication selection effects in meta-analysis. *Statistical Science*, pages 246–255.

Hou, K., Xue, C., and Zhang, L. (2017). Replicating anomalies.

Ioannidis, J., Stanley, T. D., and Doucouliagos, H. (2017). The power of bias in economics research. *The Economic Journal*, 127(605).

Ioannidis, J. P. (2008). Why most discovered true associations are inflated. *Epidemiology*, 19(5):640–648.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Med*, 2(8).

Johnson, V. E., Payne, R. D., Wang, T., Asher, A., and Mandal, S. (2017). On the reproducibility of psychological science. *Journal of the American Statistical Association*, 112(517):1–10.

McCrary, J., Christensen, G., and Fanelli, D. (2016). Conservative tests under satisficing models of publication bias. *PloS one*, 11(2):e0149590.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251):aac4716.

Patil, P. and Peng, R. D. (2016). What should researchers expect when they replicate studies? A statistical view of replicability in psychological science. *Perspectives on Psychological Science*, 11(4):539–44.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3):638–641.

Rothstein, H. R., Sutton, A. J., and Borenstein, M. (2006). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. John Wiley & Sons.

Schuemie, M. J., Ryan, P. B., DuMouchel, W., Suchard, M. A., and Madigan, D. (2014). Interpreting observational studies: why empirical calibration is needed to correct p-values. *Statistics in medicine*, 33(2):209–218.

Simonsohn, U. (2015). Small telescopes detectability and the evaluation of replication results. *Psychological Science*, 26(5):559–569.

Simonsohn, U., Nelson, L. D., and Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2):534–547.

Stanley, T., Doucouliagos, H., and Ioannidis, J. (2017). Finding the power to reduce publication bias. *Statistics in medicine*, 36(10):1580–1598.

Stanley, T. D. (2008). Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and Statistics*, 70(103-127).

Stanley, T. D. and Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5(1):60–78.

Stock, J. and Watson, M. (1998). Median unbiased estimation of coefficient variance in a time-varying parameter model. *Journal of the American Statistical Association*, 93(441):349–358.

Wolfson, P. J. and Belman, D. (2015). 15 years of research on us employment and the minimum wage. *Available at SSRN 2705499*.

For Online Publication: Supplement to the paper

Identification of and correction for publication bias

Isaiah Andrews Maximilian Kasy

October 4, 2018

This appendix contains proofs and supplementary results for the paper "Identification of and correction for publication bias." Section A collects proofs for the results stated in the main text. Section B discusses several details and extensions of the applications, including accounting for estimation error in $p(\cdot)$, identification with sign-normalized data, conditioning on covariates, and a range of robustness checks. Section C contains further empirical results, including moments and results for our GMM approaches which leave the distribution of true effects unrestricted, and empirical results for the Croke et al. (2016) and Camerer et al. (2018) applications. Finally, Section D gathers additional theoretical results on topics including the extension of our identification results to cases where publication probabilities depend on Σ and Ω, the interpretation of meta-regression coefficients in the presence of selectivity, the extension of our inference results to multivariate settings, the effect of selection on Bayesian inference, and optimal selection in a stylized model.

Appendix Contents

A Proofs

Proof of Equation 1: By construction, and Bayes' rule

$$\begin{aligned} f_{Z|\Omega,\Sigma}(z|\omega,\sigma) &= f_{Z^*|\Omega^*,\Sigma^*,D}(z|\omega,\sigma,1) \\ &= \frac{P(D=1|Z^*=z,\Omega^*=\omega,\Sigma^*=\sigma)}{P(D=1|\Omega^*=\omega,\Sigma^*=\sigma)} \cdot f_{Z^*|\Omega^*}(z|\omega) \\ &= \frac{p(z)}{E[p(Z^*)|\Omega^*=\omega]}\varphi(z-\omega). \end{aligned}$$

□

Proof of Theorem 1: We divide the proof of this theorem into two Lemmas, which might be of independent interest. We omit conditioning on Σ throughout, since it makes no difference, as shown by Equation (1).

Lemma 1

If for all z, $F_{Z|\Omega}(z|\omega)$ is continuous and strictly decreasing in ω, tends to one as $\omega \to -\infty$, and tends to zero as $\omega \to \infty$, then $\hat{\omega}_\alpha(z)$ as defined in (2) exists, is unique, and is continuous and strictly increasing for all z. If, further, $F_{Z|\Omega}(z|\omega)$ is continuous in z for all ω then $\hat{\omega}_\alpha(Z)$ is α-quantile unbiased for ω under the truncated sampling setup of Definition 1,

$$P(\hat{\omega}_\alpha(Z) \leq \omega | \Omega = \omega) = \alpha \text{ for all } \omega.$$

Proof: For the first claim, note that since $F_{Z|\Omega}(z|\omega)$ tends to zero as $\omega \to -\infty$ and tends to one as $\omega \to \infty$, for any z and any $\alpha \in (0,1)$ there exist $\omega_l(z)$ and $\omega_u(z)$ such that

$$F_{Z|\Omega}(z|\omega_u(z)) < \alpha < F_{Z|\Omega}(z|\omega_l(z)),$$

where since $F_{Z|\Omega}(z|\omega)$ is decreasing in ω we know that $\omega_l(z) < \omega_u(z)$. Thus, since $F_{Z|\Omega}(z|\omega)$ is continuous in ω, the intermediate value theorem implies that there exists $\hat{\omega}_\alpha(z) \in (\omega_l(z), \omega_u(z))$ such that $F_{Z|\Omega}(z|\hat{\omega}_\alpha(z)) = \alpha$. Since $F_{Z|\Omega}(z|\omega)$ is strictly decreasing we know this $\hat{\omega}_\alpha(z)$ is unique, while its strict monotonicity and continuity likewise follow from strict monotonicity and continuity of $F_{Z|\Omega}$ in both arguments.

For the second claim, note that since $F_{Z|\Omega}(z|\omega)$ is strictly decreasing in ω, we have $\hat{\omega}_\alpha(z) \leq \omega$ if and only if $F_{Z|\Omega}(z|\omega) \leq \alpha$. Continuity of $F_{Z|\Omega}(z|\omega)$ in z, however, means that Z is continuously distributed conditional on $\Omega = \omega$ for all ω, and thus that $F_{Z|\Omega}(Z|\omega)$ is uniformly distributed conditional on $\Omega = \omega$. Thus,

$$P\big(F_{Z|\Omega}(z|\omega) \leq \alpha | \Omega = \omega\big) = \alpha,$$

so

$$P(\hat{\omega}_\alpha(Z) \leq \omega | \Omega = \omega) = \alpha \text{ for all } \omega,$$

as we aimed to show. □

Lemma 2

If the distribution of latent draws Z^* conditional on (Ω^*, Σ^*) is $N(\Omega^*, 1)$, $p(z) > 0$ for all z, and $p(\cdot)$ is almost everywhere continuous, then the assumptions of Lemma 1 are satisfied.

Proof: Under the stated assumptions, Z is continuously distributed under all $\omega \in \mathbb{R}$, with density given by Equation (1). To prove the strict monotonicity of $F_{Z|\Omega}(z|\omega)$ in ω, we adapt the proof of Lemma A.1 in Lee et al. (2016).

In particular, note that for $z_1 > z_0$ and $\omega_1 > \omega_0$,

$$\frac{f_{Z|\Omega}(z_1|\omega_1)}{f_{Z|\Omega}(z_0|\omega_1)} > \frac{f_{Z|\Omega}(z_1|\omega_0)}{f_{Z|\Omega}(z_0|\omega_0)},$$

as can be verified from multiplying out these expressions. This means, however, that

$$f_{Z|\Omega}(z_1|\omega_1) f_{Z|\Omega}(z_0|\omega_0) > f_{Z|\Omega}(z_1|\omega_0) f_{Z|\Omega}(z_0|\omega_1).$$

Integrating both sides with respect to z_0 from $-\infty$ to $z<z_1$, and with respect to z_1 from z to ∞, we obtain that

$$(1-F_{Z|\Omega}(z|\omega_1))F_{Z|\Omega}(z|\omega_0)>(1-F_{Z|\Omega}(z|\omega_0))F_{Z|\Omega}(z|\omega_1),$$

and thus that $F_{Z|\Omega}(z|\omega_0)>F_{Z|\Omega}(z|\omega_1)$. Since this argument applies for all z and all ω_0, ω_1, we have shown that $F_{Z|\Omega}(z|\omega)$ is strictly decreasing in ω for all z.

To prove that $F_{Z|\Omega}(z|\omega)\to 0$ as $\omega\to\infty$, note that by our assumption that $p(z)$ is almost everywhere continuous, for any z_0 there exists a point $z_1>z_0$, and an open neighborhood $(z_1-\varepsilon,z_1+\varepsilon)$ of z_1 such that $p(\cdot)$ is continuous on the closure of this neighborhood, and $z_0<z_1-2\varepsilon$. Note, however, that for $\omega>z_1+\varepsilon$, $f_{Z|\Omega}(z|\omega)$ for $z\leq z_0$ is bounded above by $\varphi((z-\omega)/\sigma)/(\sigma\cdot E[p(Z)|\Omega^*=\omega])$. On the other hand, the infimum of $f_{Z|\Omega}(z|\omega)$ over $(z_1-\varepsilon,z_1+\varepsilon)$ is bounded below by $p_l\cdot\varphi((z_1-\varepsilon-\omega)/\sigma)/(\sigma\cdot E[p(Z)|\Omega^*=\omega])$ for

$$p_l=\inf_{z\in[z_1-\varepsilon,z_1+\varepsilon]}p(z)>0.$$

Integrating and taking the ratio, we see that

$$\frac{P(z\leq z_0|\Omega=\omega)}{P(z\in(z_1-\varepsilon,z_1+\varepsilon)|\Omega=\omega)}\leq\frac{\Phi((z_0-\omega)/\sigma)}{2\varepsilon p_l\cdot\varphi((z_1-\varepsilon-\omega)/\sigma)/\sigma}.$$

This expression can in turn be bounded above by

$$\frac{\Phi((z_0-\omega)/\sigma)}{2\varepsilon p_l\cdot\varphi((z_0-\omega)/\sigma)/\sigma},$$

which is proportional to Mill's ratio and tends to zero and $\omega\to\infty$ (see, for example, Baricz (2008)). This immediately implies that $F_{Z|\Omega}(z_0|\omega)\to 0$, as we aimed to show. The claim that $F_{Z|\Omega}(z|\omega)\to 1$ as $\omega\to-\infty$ can be proved analogously. □

Proof of Equation (3)**:** By construction, when $\Sigma\equiv\Sigma^r$,

$$\begin{aligned}f_{Z,Z^r}(z,z^r)&=f_{Z^*,Z^{*r}|D}(z,z^r|d=1)\\&=\frac{P(D=1|Z^*=z,Z^{*r}=z^r)}{P(D=1)}\cdot f_{Z^*,Z^{*r}}(z,z^r)\\&=\frac{p(z)}{E[p(Z^*)]}f_{Z^*,Z^{*r}}(z,z^r),\end{aligned}$$

and, since $Z^* \perp Z^{*r}|\Omega^*$,

$$f_{Z^*,Z^{*r}}(z,z^r) = \int \varphi(z-\omega)\varphi(z^r-\omega)d\mu_\Omega(\omega).$$

□

Proof of Theorem 2: Denote

$$\Delta^* = \frac{\Sigma^{*r}}{\Sigma^*} = \sqrt{\text{Var}(Z^r|\Omega)},$$

and similarly for Δ. Consistent with our convention of using lower case variables for realizations, let us denote realizations of Δ by δ.

We first show identification of $p(\cdot)$ conditional on Δ. We begin by considering the symmetric case, where $\sigma = \sigma^r$ and thus $\delta = 1$. We then allow $\sigma \neq \sigma^r$, recovering the symmetric case by (de-)convolution of Z^r with normal noise. Finally, we show that the distribution μ_Θ of Θ^* is identified.

The symmetric case: For the case $\delta = 1$, we have

$$f_{Z,Z^r|\Delta}(z,z^r|1) = \frac{p(z)}{E[p(Z^*)|\Delta^* = 1]} f_{Z^*,Z^{*r}|\Delta^*}(z,z^r|1).$$

It immediately follows that

$$f_{Z,Z^r|\Delta}(a,b|1)\cdot p(b) = f_{Z,Z^r|\Delta}(b,a|1)\cdot p(a)$$

for all a,b. Note, next, that Z^r has full support given Z,Δ, and thus that if $f_{Z,Z^r|\Delta}(a,b|1) > 0$, for some (a,b), $f_{Z,Z^r|\Delta}(a,c|1) > 0$, for all $c \in \mathbb{R}$. This in turn implies that

$$p(c) = p(a)\cdot \frac{f_{Z,Z^r|\rho}(c,a|1)}{f_{Z,Z^r|\rho}(a,c|1)}$$

for all $c \in \mathbb{R}$, where $p(a)$ is the only unknown on the right hand side. We thus find that $p(\cdot)$ is identified up to scale. Note that we have not used normality in this argument, so the result continues to hold cases where Z^*, Z^{*r} are non-normal but have the same distribution conditional on Ω^*.

The case $\delta^* \neq 1$: We already proved identification of $p(\cdot)$ for the case $\delta = 1$. We will next show that we can reduce the case where $\delta \not\equiv 1$ to this special case. Let $\tilde{Z}^r$ be such that

$$\tilde{Z}^{*r} | Z^*, D, \Omega^*, \Delta^* \sim N(\Omega^*, 1).$$

If $f_{\tilde{Z}^r|Z}$ is identified, we are done. Note that

$$f_{\tilde{Z}^r|Z} = f_{\Omega|Z} * \varphi,$$

for φ the standard normal density and

$$f_{Z^r|Z,\Delta} = f_{\Omega|Z,\Delta} * \varphi_\Delta$$

for φ_Δ the $N(0, \Delta^2)$ density. Based on the last equation, $f_{\Omega|Z,\Delta}$ is identified using deconvolution (this is a standard result; see for instance Wasserman (2006), Chapter 10.1, equation 10.18. An extensive discussion of deconvolution can be found in Meister (2009)). We then recover

$$f_{\Omega|Z}(\omega|z) = \int f_{\Omega|Z,\Delta}(\omega|z,\delta) f_{\Delta|Z}(\delta|z) d\delta,$$

and identification of $p(\cdot)$ follows.

Identification of μ_Θ Knowledge of $p(\cdot)$ up to scale allows us to recover the joint density f_{X^*,Σ^*} via

$$f_{X^*,\Sigma^*}(x,\sigma) = \frac{E[p(Z^*)]}{p(x/\sigma)} f_{X,\Sigma}(x,\sigma).$$

Deconvolution then identifies $\mu_{\Theta^*|\Sigma^*}$, since $f_{Z^*|\Sigma^*} = \mu_{\Theta^*|\Sigma^*} * \varphi_{\Sigma^*}$. Integrating over the marginal distribution of Σ^* yields μ_Θ. □

Proof of Theorem 3: Assume without loss of generality that $\sigma = 1$ lies in the interior of the support of Σ, and let

$$h(z) = f_{Z^*|\Sigma^*}(z|1).$$

If $h(\cdot)$ is identified, then so are $p(\cdot)$ and μ_Θ. We will show that $h(\cdot)$ is identified, which immedaitely identifies μ_Θ by deconvolution, since $h = \mu_\Theta * \varphi$. We can then identify

$p(z)$ as before, since the truncated conditional density of Z is given by

$$f_{Z|\Sigma}(z|\sigma) = \frac{p(z)}{E[p(Z^*)|\Sigma^* = \sigma]} f_{Z^*|\Sigma^*}(z|\sigma), \tag{5}$$

and thus

$$p(z) = const. \cdot \frac{f_{Z|\Sigma}(z|1)}{h(z)}.$$

A second order ODE for $h(\cdot)$. Let $\pi = 1/\sigma$ be the precision of an estimate. Differentiating the log of expression (5) for the truncated density at $\pi = 1$ yields

$$g(z) = \partial_\pi \log f_{Z|\Sigma}(z|1) = C_1 + \partial_\pi \log f_{Z^*|\Sigma^*}(z|1) \tag{6}$$

for a constant C_1. Note how, as we differentiate $\log f_{Z|\Sigma}(z|1)$ with respect to π at a given value z, the term $p(z)$ drops out of the resulting equation. The function g is identified under our assumptions.

Recall now that the definition of the standard normal density implies $\varphi'(z) = -z\varphi(z)$. The density $f_{Z^*|\Sigma^*}$ is given by $\mu_\Theta * \varphi_\Sigma$, and thus $f_{Z^*|\Sigma^*}(z|1/\pi) = \int \varphi(z-\theta\pi) d\mu_\Theta(\theta)$, which implies

$$\begin{aligned}
\partial_z f_{Z^*|\Sigma^*}(z|1) &= -\int (z-\theta)\varphi(z-\theta) d\mu_\Theta(\theta) \\
\partial_z^2 f_{Z^*|\Sigma^*}(z|1) &= -f_{Z^*|\Sigma^*}(z|1) + \int (z-\theta)^2 \varphi(z-\theta) d\mu_\Theta(\theta) \\
\partial_\pi f_{Z^*|\Sigma^*}(z|1) &= \int \theta(z-\theta)\varphi(z-\theta) d\mu_\Theta(\theta) \\
&= -\left[f_{Z^*|\Sigma^*}(z|1) + z \cdot \partial_z f_{Z^*|\Sigma^*}(z|1) + \partial_z^2 f_{Z^*|\Sigma^*}(z|1) \right],
\end{aligned}$$

from which we conclude

$$h''(z) = (C_1 - 1 - g(z)) \cdot h(z) - z \cdot h'(z). \tag{7}$$

Equation (7) is a second order linear homogeneous ordinary differential equation.

Two free parameters Given the initial conditions $h(0) = h_0$ and $h'(0) = h_0'$, and given C_1, the solution to this equation exists and is unique, because all coefficients are continuous in z; cf. Murphy (2011). Furthermore, the general solution to this differential equation can be written in the form $h(z, C_1, h_0, h_0') = h_0 \cdot h_1(z, C_1) + h_0' \cdot h_2(z, C_1)$, where the functions

$h_1(\cdot)$ and $h_2(\cdot)$ are determined by equation (7); cf. Murphy (2011), chapter B. This leaves three free parameters to be determined, C_1,h_0 and h_0'. The constraint $\int h(z)dz=1$ pins down h_0 or h_0' given the other two parameters, so that there remain two free parameters.

A fourth order ODE for $h(\cdot)$. We next turn to the second derivative $k(\cdot)$ defined by

$$k(z)=\partial_\pi^2 \log f_{Z|\Sigma}(z|1)=C_2+\partial_\pi^2 \log f_{Z^*|\Sigma^*}(z|1),$$

which is identified under our assumptions, just like $g(\cdot)$. Calculations similar to those for the first derivative with respect to π yield the fourth order differential equation

$$h^{(4)}(z)=\big(k(z)-C_2+(g(z)-C_1)^2-2\big)h(z)-4zh'(z)-(z^2+5)h''(z)-2zh^{(3)}(z). \quad (8)$$

To complete this proof, we now (i) derive the fourth order differential equation (8) and (ii) show that it allows us to pin down the remaining free parameters. We provide further discussion immediately following the proof.

Derivation of the fourth order ODE for $h(\cdot)$ Differentiating $\log f_{Z^*|\Sigma^*}$ twice yields

$$\partial_\pi^2 \log f_{Z^*|\Sigma^*}(z|1)=\frac{\partial_\pi^2 f_{Z^*|\Sigma^*}(z|1)}{h(z)}-(g(z)-C_1)^2,$$

so that

$$\partial_\pi^2 f_{Z^*|\Sigma^*}(z|1)=h(z)\cdot\big(k(z)-C_2+(g(z)-C_1)^2\big).$$

From $f_{Z^*|\Sigma^*}(z|1/\pi)=\int\varphi(z-\theta\pi)d\mu_\Theta(\theta)$ we note that

$$\partial_\pi^2 f_{Z^*|\Sigma^*}(z|1)=\int\big(-\theta^2+\theta^2(z-\theta)^2\big)\varphi(z-\theta)d\mu_\Theta(\theta).$$

We furthermore have

$$\begin{aligned}
h^{(3)}&=-3h'(z)-\int(z-\theta)^3\varphi(z-\theta)d\mu_\Theta(\theta)\\
h^{(4)}&=-3h''(z)-3\int(z-\theta)^2\varphi(z-\theta)d\mu_\Theta(\theta)+\int(z-\theta)^4\varphi(z-\theta)d\mu_\Theta(\theta)\\
&=-6h''(z)-3h(z)+\int(z-\theta)^4\varphi(z-\theta)d\mu_\Theta(\theta).
\end{aligned}$$

Comparing coefficients on θ between $\partial_\pi^2 f_{Z^*|\Sigma^*}$ and the derivatives of $h(\cdot)$, we get the fourth order differential equation (8).

The fourth order ODE pins down the remaining free parameters Our proof is complete once we have shown that there is at most one set of values C_1,C_2,h_0 and h_0' such that the resulting h satisfies the two differential equations (7) and (8). Differentiating equation (7) three times yields

$$\begin{aligned}
h''(z) = {} & (-1+C_1-g(z))h(z) && -zh'(z) && && \\
h^{(3)}(z) = {} & -g'(z)h(z) && +(-2+C_1-g(z))h'(z) && -zh''(z) & \\
h^{(4)}(z) = {} & -g''(z)h(z) && -2g'(z)h'(z) && +(-3+C_1-g(z))h''(z) & -zh^{(3)}(z) \\
h^{(5)}(z) = {} & -g^{(3)}(z)h(z) && -3g''(z)h'(z) && -3g'(z)h''(z) & \\
& && && +(-4+C_1-g(z))h^{(3)}(z) & -zh^{(4)}(z),
\end{aligned}$$

and differentiating equation (8) yields

$$\begin{aligned}
h^{(4)}(z) = {} & \big(-2-C_2+(-C_1+g(z))^2+k(z)\big)h(z) && -4zh'(z) & \\
& && -\big(5+z^2\big)h''(z) & -2zh^{(3)}(z), \\
h^{(5)}(z) = {} & (2(-C_1+g(z))g'(z)+k'(z))h(z) && +\big(-6-C_2+(C_1-g(z))^2+k(z)\big)h'(z) & \\
& -6zh''(z) && +\big(-7-z^2\big)h^{(3)}(z) & -2zh^{(4)}(z).
\end{aligned}$$

We can iteratively eliminate the derivatives of $h(\cdot)$ from these equations by substitution. After doing so, we divide by $h(z)$, which is possible since $h(z)>0$ for all z by construction. This yields the following equation involving the constants C_1 and C_2, but not involving the function $h(\cdot)$ or any of its derivatives:

$$\begin{aligned}
& C_1^2+C_2^2+g(z)^2+k(z)^2-z^2g'(z)^2+4k(z)g''(z)+3g''(z)^2 \\
& \qquad -2C_2(g(z)+k(z)+2g''(z))+2g(z)\big(k(z)+2\big(g'(z)^2+g''(z)\big)\big) \\
& \qquad\quad +C_1\big(2C_2-2\big(g(z)+k(z)+2\big(g'(z)^2+g''(z)\big)\big)\big)-2g'(z)g^{(3)}(z)=2g'(z)k'(z)
\end{aligned}$$

This equation again has to hold for all z. Differentiating twice with respect to z yields new equations where the constants C_1 and C_2 enter only linearly, and we can explicitly solve for them.[19]

Substituting the solutions C_1 and C_2 back into one of the first order differential equations we obtained by substitution and elimination of higher order derivatives above, we obtain a solution for h_0' given h_0. Given h_0, h_0' and the constants C_1 and C_2, equation (7) yields a unique solution $h(z)$ for all z. Rescaling any solution $h(\cdot)$ by a constant again yields a solution by linearity of the differential equations. h_0 is finally pinned down by

[19]The resulting expressions are unwieldy and so are omitted here, but are available on request.

the constraint $\int h(z)dz=1$. $\square$

Remarks:

- The proof of Theorem 3 shows that our model is overidentified. If we consider higher order derivatives of equations (7) and (8), or alternatively evaluate them at different values z, we obtain infinitely many restrictions on a finite number of free parameters.

- The proof of identification is considerably simplified if we restrict the model to a normal distribution for Θ^*, $\Theta^* \sim N(\bar{\mu},\tau^2)$, which implies $Z^*|\Sigma^*=1 \sim N(\bar{\mu},\tau^2+1)$, and thus $h(z)=const.\cdot\exp\left(-\frac{1}{2(\tau^2+1)}(z-\bar{\mu})^2\right)$. Denoting $e(z)=\partial_z \log h(z)$, we can rewrite equation (7) as

$$e'(z)=C_1-g(z)-1-ze(z)-e^2(z),$$

  while the normality assumption yields $e(z)=-(z-\bar{\mu})/(\tau^2+1)$ and $e'(z)=-\frac{1}{(\tau^2+1)}$. Plugging in yields

$$-\tfrac{1}{(\tau^2+1)}=C_1-g(z)-1+z\tfrac{z-\bar{\mu}}{(\tau^2+1)}-\left(\tfrac{z-\bar{\mu}}{(\tau^2+1)}\right)^2.$$

  Evaluating this equation at different values z pins down τ^2 and $\bar{\mu}$.

- The proof of Theorem 3 could be equivalently stated in terms of linear operators rather than differential equations. In particular, the ordinary differential equations (7) and (8) are equivalent to the following two linear operator equations, indexed by z and linear in μ,

$$\int [\theta(z-\theta)-(g(z)-C_1)]\varphi(z-\theta)d\mu_\Theta(\theta)=0$$
$$\int \left[\left(-\theta^2+\theta^2(z-\theta)^2\right)-\left(k(z)-C_2+(g(z)-C_1)^2\right)\right]\varphi(z-\theta)d\mu_\Theta(\theta)=0$$

  Identification is then equivalent to existence of at most one (μ_Θ,C_1,C_2) triple solving these equations for all z.

B Details and Extensions of Empirical Applications

This section discusses a variety of details for and extensions of the empirical applications reported in the main text. Section B.1 develops an extension of our confidence set construction approach that allows estimation error in $p(\cdot)$. Section B.2 discusses the extension of our identification results to cases where we condition on covariates. Section B.3 develops an extension of our identification results to allow the sign of Z to be normalized as in two of our applications. Sections B.4 and B.5 describe the likelihood used for estimation in our applications and the details of the data and variable construction, respectively. Finally, Section B.6 discusses a variety additional specifications and robustness checks for the results reported in the main text.

B.1 Estimation Error in $p(\cdot)$

The bias corrections discussed in Section 2.2 assume the conditional publication probability is known. If $p(\cdot)$ is instead estimated with error, median unbiased estimation is challenging, but constructing valid confidence sets for ω is straightforward.

Suppose we parameterize the conditional publication probability by β, and let $\hat{\omega}_\alpha(X_i;\beta)$ be the α-quantile unbiased estimator under β. For many specifications of $p(\cdot)$, and in particular for those used in our applications, $\hat{\omega}_\alpha(x;\beta)$ is continuously differentiable in β for all x. If we have a consistent and asymptotically normal estimator $\hat{\beta}$ for β, for $0<\delta<\alpha$, consider the interval

$$\left[\hat{\omega}_{\frac{\alpha-\delta}{2}}\left(X;\hat{\beta}\right)-c_{1-\frac{\delta}{2}}\hat{\sigma}_L(X),\hat{\omega}_{1-\frac{\alpha-\delta}{2}}\left(X;\hat{\beta}\right)+c_{1-\frac{\delta}{2}}\hat{\sigma}_U(X)\right]$$

where $c_{1-\frac{\delta}{2}}$ is the level $1-\frac{\delta}{2}$ quantile of the standard normal distribution while $\hat{\sigma}_L(x)$ and $\hat{\sigma}_U(x)$ are delta-method standard errors for $\hat{\omega}_{\frac{\alpha-\delta}{2}}\left(x;\hat{\beta}\right)$ and $\hat{\omega}_{1-\frac{\alpha-\delta}{2}}\left(x;\hat{\beta}\right)$, respectively. If our model for $p(\cdot)$ is correctly specified, Bonferroni's inequality implies that this interval covers ω with probability at least $1-\alpha$ in large samples.[20]

B.2 Conditioning on covariates

Our baseline results do not consider any study level covariates, such as journal of publication, year of initial circulation of a study, research topic, identification approach, or author seniority. Heterogeneity in degree of publication bias based on study and author

[20]Even in cases where we do not have an asymptotically normal estimator for β, for example because we consider a fully nonparameteric model for $p(\cdot)$, given an initial level $1-\delta$ confidence set CS_β for β we can form a Bonferroni confidence set for ω as $\left[\inf_{\beta\in CS_\beta}\hat{\omega}_{\frac{\alpha-\delta}{2}}\left(X;\hat{\beta}\right),\sup_{\beta\in CS_\beta}\hat{\omega}_{1-\frac{\alpha-\delta}{2}}\left(X;\hat{\beta}\right)\right]$.

characteristics has been explored by many authors, including Open Science Collaboration (2015), Brodeur et al. (2016) and Brodeur et al. (2018). Conditioning our analysis on such covariates might be interesting for two reasons: (i) to make our identification assumptions more credible, and (ii) to explore variation in $p(\cdot)$ and μ_Θ.

Provided our assumptions hold conditional on the covariates, our results extend directly to this setting. Considering covariates C, Equation (1) could for instance be modified to

$$f_{Z|\Omega,\Sigma,C}(z|\omega,\sigma,c)=\frac{p(z,c)}{E[p(Z^*,c)|\Omega^*=\omega]}\varphi(z-\omega).$$

If $p(z,c)$ is known up to scale for each c, we can apply Theorem 1 conditional on C to obtain corrected estimates. Likewise, if we have replication estimates we can apply Theorem 2 conditional on $C=c$ to identify $p(z,c)$ up to scale for each c, along with the conditional distribution of true effects given the covariate $\mu_{\Theta|C}$. Likewise, if Θ^* and Σ^* are independent conditional on covariates in the population of latent studies, we can apply Theorem 3 conditional on $C=c$ to identify $p(z,c)$ up to scale for each c, along with $\mu_{\Theta|C}$. Note, however, that in both cases μ_Θ is not identified without further restrictions.

In Section B.6 below, we re-estimate our applications allowing $p(\cdot)$ to depend on covariates like journal of publication and year of initial circulation of a study. However, in no case do we reject our baseline specifications at conventional significance levels.

B.3 Sign-normalized data

In the applications of Section 4.1 and 4.2, the sign of the estimates Z in our replication datasets is normalized to be positive, with the sign of Z^r adjusted accordingly. The following corollary shows that under this sign normalization identification of $p(\cdot)$ still holds, using either replication studies or meta-studies, so long as $p(\cdot)$ is symmetric in its argument.

Corollary 1 *1. Consider the setup of Theorem 2. Assume additionally that $p(\cdot)$ is symmetric, $p(z)=p(-z)$, and that $f_{\Sigma|Z^*}(\sigma|z)=f_{\Sigma|Z^*}(\sigma|-z)$ for all z. Suppose that we observe i.i.d. draws of $(W,W^r)=\text{sign}(Z)\cdot(Z,Z^r)$. In this setup $p(\cdot)$ is nonparametrically identified on $\mathbb{R}$ up to scale, and the distribution of $|\Theta^*|$ is identified as well.*

2. Consider the setup of Theorem 3. Assume additionally that $p(\cdot)$ is symmetric, i.e., $p(z)=p(-z)$. Suppose that we observe i.i.d. draws of $(|X|,\Sigma)$. In this setup $p(\cdot)$ is nonparametrically identified on $\mathbb{R}$ up to scale, and the distribution of $|\Theta^*|$ is identified as well.

Proof of Corollary 1:

Replication studies: Let $S^* = \pm 1$ with probability 0.5, independently of $(Z^*, Z^{*r}, \Sigma^*, \Sigma^{r*}, \Theta^*, D)$, and let $S = S^*$ denote S^* for published studies. Define

$$(V, V^r) = S \cdot (W, W^r).$$

We show that (V, V^r) satisfies the assumptions of Theorem 2, from which the claim then follows. Define $\tilde{S}^* = S^* \cdot \text{sign}(Z^*)$, so that $(V, V^r) = \tilde{S} \cdot (Z, Z^r)$, and define $\tilde{\Omega}^* = \tilde{S}^* \cdot \Omega^*$. Since $\tilde{S}$ is independent of (Z, Z^r, Δ, Ω),

$$\tilde{\Omega}^* \sim \tilde{\mu} = \tfrac{1}{2}(\mu_{\Omega^*} + \mu_{-\Omega^*})$$

and

$$f_{V,V^r,\Delta}(v, v^r, \delta) = p(v) \cdot f_{\Delta|Z^*}(\delta|v) \cdot \frac{\int \varphi(v-\omega) \cdot \frac{1}{\delta}\varphi\left(\frac{v^r-\omega}{\delta}\right) d\tilde{\mu}_\Omega(\omega)}{\iint p(v') \cdot \varphi(v'-\omega) dv' d\tilde{\mu}_\Omega(\omega)}.$$

This has the exact same form as the density of (Z, Z^r, Δ) under the symmetric measure $\tilde{\mu}_\Omega$. The claim follows, since identification of $\tilde{\mu}_\Omega$ implies identification of the distribution of $|\Omega^*|$. $\square$

Meta-studies: The proof for meta-studies proceeds similarly to the replication studies case. Let $S^* = \pm 1$ with probability 0.5, independently of $(Z^*, \Sigma^* \Theta^*, D)$, and let $S = S^*$ denote S^* for published studies. Define $V = S \cdot |X|$. We show that (V, Σ) satisfies the assumptions of Theorem 3, from which the claim then follows.

Define $\tilde{S}^* = S^* \cdot \text{sign}(X^*)$, so that $V = \tilde{S} \cdot X$, and define $\tilde{\Theta}^* = \tilde{S}^* \cdot \Theta^*$. Since $\tilde{S}$ is independent of (Z, Σ, Θ), we get $\tilde{\Theta}^* \sim \tilde{\mu} = \frac{1}{2}(\mu_{\Theta^*} + \mu_{-\Theta^*})$ and

$$f_{V/\Sigma|\Sigma}(z|\sigma) = \frac{p(z) \cdot \int \varphi(z - \theta/\sigma) d\tilde{\mu}(\theta)}{\iint p(z') \varphi(z' - \theta/\sigma) dz' d\tilde{\mu}(\theta)}.$$

This has the exact same form as the density of Z given Σ under the symmetric measure $\tilde{\mu}$. The claim follows, where we again use the fact that identification of $\tilde{\mu}$ implies identification of the distribution of $|\Theta^*|$. $\square$

B.4 Likelihood and parametric specifications

B.4.1 Systematic replications

Under the replication setting the marginal density of Z,Z^r,Δ (where again $\Delta=\Sigma^r/\Sigma$) is

$$f_{Z,Z^r,\Delta}(z,z^r,\delta)=\frac{p(z)\int\varphi(z-\omega)\cdot\frac{1}{\delta}\varphi\big(\frac{z^r-\omega}{\delta}\big)d\mu_\Omega(\omega)}{\iint p(z')\cdot\varphi(z'-\omega)dz'd\mu(\theta)}f_{\Delta^*|Z^*}(\delta|z). \tag{9}$$

Denoting the total number of observations by J, the joint likelihood of the observed sample $((z_1,z_1^r,\delta_1),...,(z_J,z_J^r,\delta_J))$ is $\mathcal{L}(p,\mu)=\prod_{j=1}^J f_{Z,Z^r,\Delta}(z_j,z_j^r,\delta_j)$. To fit a given model, we maximize this likelihood with respect to $p(\cdot)$ and μ. Since $f_{\Delta^*|Z^*}$ enters multiplicatively, it plays no role in maximum likelihood estimation of $p(\cdot)$ and μ. Hence, we drop this term from the likelihood used in estimation.

To model $p(\cdot)$, similar to Hedges (1992) we consider step functions

$$p(z)\propto\sum_{k=1}^K\beta_{p,k}\cdot\mathbf{1}(\zeta_{k-1}\leq z<\zeta_k),$$

where $-\infty=\zeta_0<\zeta_1<...<\zeta_K=\infty$ are fixed cutoffs. Since $p(\cdot)$ is only identified up to scale, we normalize $\beta_{p,K}=1$ and estimate $\beta_{p,1},...,\beta_{p,K-1}$. Thus $\beta_{p,k}$ can be interpreted as the publication probability for a latent study with Z^* between ζ_{k-1} and ζ_k, relative to a latent study with $Z^*\geq\zeta_{K-1}$.

Sign normalization The sign of the initial estimate is normalized to be positive in both of our replication datasets. In these applications, we thus follow the approach of Corollary 1 and assume that $p(\cdot)$ is symmetric around zero. We conduct estimation based on the normalized z-statistics $(W,W^r)=\text{sign}(Z)\cdot(Z,Z^r)$ using the marginal likelihood

$$f_{W,W^r,\Delta}(w,w^r,\delta)=f_{Z,Z^r,\Delta}(w,w^r,\delta)+f_{Z,Z^r,\Delta}(-w,-w^r,\delta).$$

In this setting, Corollary 1 implies that $\beta_{p,1},...\beta_{p,K-1}$ and the distribution of $|\Theta^*|$ are identified.

B.4.2 Meta-studies

In the meta-study context, the marginal likelihood of (X,Σ) is

$$f_{X,\Sigma}(x,\sigma)=\frac{p(\frac{x}{\sigma})\cdot\int\varphi\big(\frac{x-\theta}{\sigma}\big)d\mu(\theta)}{\int p(\frac{x'}{\sigma})\cdot\varphi\big(\frac{x'-\theta}{\sigma}\big)dx'd\mu(\theta)}f_\Sigma^*(\sigma). \tag{10}$$

Again denoting the total number of observations by J, this yields joint likelihood $\mathcal{L}(p,\mu)=\prod_{j=1}^{J} f_{X,\Sigma}(x_j,\sigma_j)$, which we again use to estimate $p(\cdot)$ and μ. As before, f_Σ enters multiplicatively and need not be specified. Also as before, we consider step function specifications for $p(\cdot)$.

Sign normalization In contexts where the sign of the initial estimate has been normalized to be positive, we follow the analog of the approach described above, restricting $p(\cdot)$ to be symmetric and conducting estimation based on $|X|=W\cdot\Sigma$ and Σ.

B.5 Details on data and variable construction

In this section, we discuss how we cast the data of Camerer et al. (2016) and Open Science Collaboration (2015) into our framework. The data in Wolfson and Belman (2015) is already in the desired format.

B.5.1 Details for economics laboratory experiments

To apply our approach, we need z-statistics and standard errors for both the original and replication studies. For the application to data from Camerer et al. (2016), we first collect p-values and standardized effect sizes from table S1 in the supplement. Some of the p-values are censored below at .001, so for these studies we also collect the original estimates and standard errors from the replication reports posted online by Camerer et al.[21] and recompute the censored p-values. We then construct z-statistics by inverting the p-value transformation, $|z|=\Phi^{-1}(1-p/2)$. To obtain effect size estimates, we apply the Fisher transformation to standardized effect sizes reported by Camerer et al. Dividing these estimates by the z-statistics finally recovers the standard error.

B.5.2 Details for psychology laboratory experiments

To apply our approach to the data from Open Science Collaboration (2015), we again need z-statistics and standard errors for both the original and replication studies. We draw the inputs for all of these calculations from the RPPdataConverted spreadsheet posted online by the Open Science Collaboration.[22] Since Open Science Collaboration (2015) report p-values for both the original and replication studies, we invert the p-value transform to obtain z statistics. We use the p-values reported in their columns T.pval.USE.O and T.pval.USE.R for the original and replication studies, respectively. Since some of the p-values in this application are based on one-sided tests, we account for this in the inversion step. To compute ef-

[21] Available at `https:/c/experimentaleconreplications.com/replicationreports.html`, accessed September 3, 2016.

[22] Available at `https://osf.io/ytpuq/files/`, accessed January 19, 2017.

fect size estimates, we again apply the Fisher transformation to the standardized effect sizes (columns T.r.O and T.r.R of RPPdataConverted for the original and replication studies, respectively), and then divide these estimates by the z-statistics to construct standard errors.

B.6 Additional maximum likelihood results

This section discusses results from additional specifications estimated by maximum likeihood, intended to complement the results discussed in the main text.

B.6.1 Additional results for economics laboratory experiments

Here we report results based on an alternative specification for the economics replication data from Camerer et al. (2016). We consider specifications which allow the probability of publication to vary depending on whether a latent study is sent to the American Economic Review (AER) or Quarterly Journal of Economics (QJE). The publication probability is identified up to scale separately for each journal. We index the journal by c, and set $p(z,c)$ proportional to one for both journals when the result is significantly different from zero at the 5% level. This ensures that the β parameters can be interpreted as publication probabilities for insignificant results relative to significant results at the same journal. Our ultimate specification is

$$p(Z,S) \propto \begin{cases} \beta_{p,1} & |Z| < 1.96, C = AER \\ \beta_{p,1} + \beta_{p,2} & |Z| < 1.96, C = QJE \\ 1 & |Z| \geq 1.96. \end{cases}$$

Results are reported in Table 4. In both the replication and metastudy specifications we estimate that the QJE is more likely to publish insignificant results. This makes sense given that the sample contains one significant result and one insignificant result published in the QJE, while it contains fifteen significant results and one insignificant result published in the AER. The estimated publication probabilities for the QJE are quite noisy, however, and we cannot reject the hypothesis that $\beta_{p,2} = 0$, so the same publication rule is used at both journals.

B.6.2 Additional results for psychology laboratory experiments

We next report results based on three alternative specifications for the psychology replication data from Open Science Collaboration (2015). We first limit attention to studies with a large number of denominator degrees of freedom. Second, we limit attention to

Replication				Meta-study			
κ	λ	$\beta_{p,1}$	$\beta_{p,2}$	$\tilde{\kappa}$	$\tilde{\lambda}$	$\beta_{p,1}$	$\beta_{p,2}$
0.373	2.153	0.015	0.216	1.847	0.131	0.021	0.786
(0.267)	(1.029)	(0.021)	(0.333)	(1.582)	(0.065)	(0.030)	(1.496)

Table 4: Selection estimates from lab experiments in economics, allowing publication probability to vary by journal. The left panel reports estimates from replication specifications, while the right panel reports results from meta-study specifications. Publication probability β_p is measured relative to omitted category of studies significant at 5% level.

studies where the replication protocols were approved by the original authors. Third, we allow the publication rule to vary by journal.

Denominator degrees of freedom As noted in the main text, our baseline analysis of the Open Science Collaboration (2015) data focuses on studies that use z- or t-statistics (or the squares of these statistics). Our analysis then treats these statistics as approximately normal. A potential problem here is that t-distributions with a small number of degrees of freedom behave differently from normal distributions, and in particular have heavier tails. While the smallest degrees of freedom in the Open Science Collaboration (2015) data is seven, this concern may still lead us to worry about the validity of our approach in this setting. To address this concern, in Table 5 we report parameter estimates using the replication and meta-study specifications discussed in Section 4.2, where

$$p(Z) \propto \begin{cases} \beta_{p,1} & |Z| < 1.64 \\ \beta_{p,2} & 1.64 \leq |Z| < 1.96 \\ 1 & |Z| \geq 1.96, \end{cases}$$

except that we now limit attention to the 52 observations with denominator degrees of freedom at least 30 in the original study.[23] Our results are broadly similar for this restricted sample and for the full data.

Approved replications As discussed in the main text, Gilbert et al. (2016) argue that some of the replications in Open Science Collaboration (2015) deviated substantially from the protocol of the original studies, which might lead to a violation of our assumption that the replication and original results are generated by the same underlying parameter Θ. Before conducting their replications, however, Open Science Collaboration (2015) asked

[23]We screen only on the degrees of freedom in the original study since sample sizes, and thus degrees of freedom, in the replication studies depend on the results in the initial study. Hence, screening on replication degrees of freedom has the potential to introduce additional selection on the results of the original study.

Replication				Meta-study			
κ	λ	$\beta_{p,1}$	$\beta_{p,2}$	$\tilde{\kappa}$	$\tilde{\lambda}$	$\beta_{p,1}$	$\beta_{p,2}$
0.174	1.602	0.007	0.142	0.869	0.138	0.018	0.247
(0.121)	(0.677)	(0.005)	(0.079)	(0.657)	(0.059)	(0.012)	(0.142)

Table 5: Selection estimates from lab experiments in psychology, restricted to observations with denominator degrees of freedom at least 30, with standard errors in parentheses. The left panel reports estimates from replication specifications, while the right panel reports results from meta-study specifications. Publication probability β_p is measured relative to omitted category of studies significant at 5% level.

the authors of each original study to review the proposed replication protocol, and recorded whether the original authors endorsed the replication protocol. We can thus partly address this critique by limiting attention to the subset of studies where the replication was endorsed by the authors of the original study. Re-estimating the specifications of Section 4.2 on the 51 endorsed replications, we obtain the estimates reported in Table 6. These estimates suggest a somewhat smaller degree of selection than our baseline estimates, consistent with a higher rate of replication for approved replications, but are broadly similar to our other estimates. Figure 10 plots the original and replication estimates along with our adjusted estimates, showing somewhat better fit than in Figure 8 in the main text.

Replication				Meta-study			
κ	λ	$\beta_{p,1}$	$\beta_{p,2}$	$\tilde{\kappa}$	$\tilde{\lambda}$	$\beta_{p,1}$	$\beta_{p,2}$
0.490	1.159	0.017	0.365	0.634	0.198	0.022	0.440
(0.267)	(0.400)	(0.011)	(0.165)	(0.503)	(0.079)	(0.014)	(0.217)

Table 6: Selection estimates from lab experiments in psychology, approved replications, with standard errors in parentheses. The left panel reports estimates from replication specifications, while the right panel reports results from meta-study specifications. Publication probability β_p is measured relative to omitted category of studies significant at the 5% level.

Publication rule varies by journal The published studies replicated in Open Science Collaboration (2015) are drawn from Psychological Science (PS), the Journal of Personality and Social Psychology (JPSP), and the Journal of Learning Memory and Cognition (JLMC). In this section we estimate a model where we allow the publication rule to vary

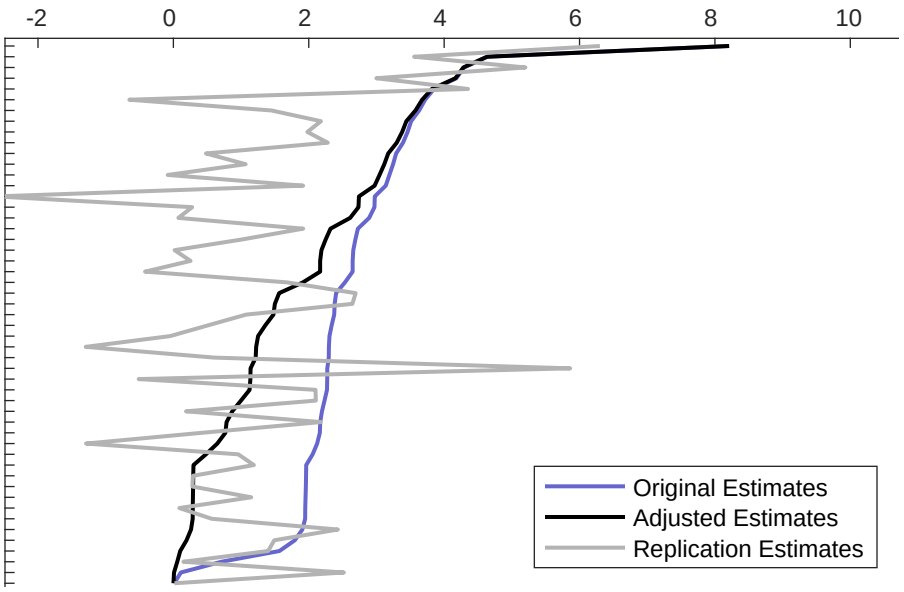

Figure 10: This figure plots the estimates W and W^r from the original and replication studies in Open Science Collaboration (2015), limiting the sample to approved replications, along with the median unbiased estimate $\hat{\theta}_{\frac{1}{2}}$ based on the estimated selection model and the original estimate.

by journal, which we index by C. In particular, we consider the publication rule:

$$p(Z,C) \propto \begin{cases} \beta_{p,1} & |Z|<1.64, C=JLMC \\ \beta_{p,1}+\beta_{p,2} & |Z|<1.64, C=PS \\ \beta_{p,1}+\beta_{p,3} & |Z|<1.64, C=JPSP \\ \beta_{p,4} & 1.64\leq|Z|<1.96, C=JLMC \\ \beta_{p,4}+\beta_{p,5} & 1.64\leq|Z|<1.96, C=PS \\ \beta_{p4}+\beta_{p,6} & 1.64\leq|Z|<1.96, C=JPSP \\ 1 & |Z|\geq 1.96, \end{cases}$$

As discussed in the economics application above, we normalize the publication probability for studies significant at the 5% level to be proportional to one, which allows us to interpret the β coefficients in terms of the publication probability for insignificant studies relative to that for significant studies at the same journal. Such a normalization is necessary since publication probabilities are only identified up to a journal-specific scaling factor.

Results from estimating this model are reported in Table 7. These are noisier than our baseline estimates, as is intuitive given the larger number of parameters, but the JLMC coefficients show roughly the same pattern as our baseline specifications. None of the differences between journal publication probabilities are significant, and a joint test yields a p-value of .78 in the replication specification and .84 in the metastudy specification, so in neither case do we reject the null hypothesis that all the journals use the same publication rule.

				Replication			
κ	λ	$\beta_{p,1}$	$\beta_{p,2}$	$\beta_{p,3}$	$\beta_{p,4}$	$\beta_{p,5}$	$\beta_{p,6}$
0.315	1.308	0.008	0.002	-0.001	0.428	-0.288	-0.332
(0.140)	(0.330)	(0.008)	(0.011)	(0.011)	(0.245)	(0.264)	(0.260)
				Meta-study			
$\tilde{\kappa}$	$\tilde{\lambda}$	$\beta_{p,1}$	$\beta_{p,2}$	$\beta_{p,3}$	$\beta_{p,4}$	$\beta_{p,5}$	$\beta_{p,6}$
0.966	0.154	0.013	0.005	0.008	0.555	-0.360	-0.368
(0.561)	(0.054)	(0.014)	(0.019)	(0.026)	(0.320)	(0.350)	(0.364)

Table 7: Selection estimates from lab experiments in psychology, allowing publication probability to vary by journal. The top panel reports estimates from replication specifications, while the bottom panel reports results from meta-study specifications. Publication probability β_p is measured relative to omitted category of studies significant at 5% level.

B.6.3 Additional results for minimum wage meta-study

This section reports results based on two alternative specifications for the data from Wolfson and Belman (2015). Since Wolfson and Belman (2015) include estimates from both published and working papers, we first reanalyze the data limiting attention to published studies. We then examine whether the publication rules appear to vary with time.

Published Studies Table 8 reports estimates based on the model

$$\Theta^* \sim \bar{\theta} + t(\nu) \cdot \tilde{\tau}, \quad p(Z) \propto \begin{cases} \beta_{p,1} & Z < -1.96 \\ \beta_{p,2} & -1.96 \leq Z < 0 \\ \beta_{p,3} & 0 \leq Z < 1.96 \\ 1 & Z \geq 1.96 \end{cases}$$

based on the subset of published papers, consisting of 705 estimates drawn from 31 studies. As in the main text we cluster our standard errors at the study level. The resulting estimates are broadly similar to those obtained on the full sample.

$\bar{\theta}$	$\tilde{\tau}$	$\tilde{\nu}$	$\beta_{p,1}$	$\beta_{p,2}$	$\beta_{p,3}$
0.022	0.044	1.697	0.838	0.365	0.387
(0.012)	(0.025)	(0.380)	(0.331)	(0.146)	(0.140)

Table 8: Meta-study selection estimates from minimum wage data, published studies, with standard errors in parentheses. Publication probability β_p is measured relative to omitted category of studies estimating a positive effect significant at the 5% level.

Time Trends We next examine whether publication rules appear to vary over time. In particular, letting T_i denote the year in which study i was initially circulated, for $\varsigma(x)=\exp(x)/(1+\exp(x))$ the logistic function we consider the model

$$\Theta^* \sim \bar{\theta}+t(\nu)\cdot\tilde{\tau}, \quad p(Z,T)\propto \begin{cases} \varsigma(\beta_{p,1}+\beta_{p,2}(T-2013)) & Z<-1.96 \\ \varsigma(\beta_{p,3}+\beta_{p,4}(T-2013)) & -1.96\leq Z<0 \\ \varsigma(\beta_{p,5}+\beta_{p,6}(T-2013)) & 0\leq Z<1.96 \\ \varsigma(1) & Z\geq 1.96 \end{cases}$$

where we measure time in years relative to 2013, which is the median year observed in the data, and T varies between 2000 and 2015. We use the logistic function here to ensure that publication probabilities lie between zero and one, and without the time trend this would simply be a reparameterization of our baseline model. Publication probabilities are only identified up to a year-specific scaling, so by normalizing the publication coefficient for studies finding a negative and significant effect of the minimum wage on employment to be proportional to one, we again ensure that the β_p coefficients can be interpreted as measuring publication probabilities relative to the publication probability for studies finding a negative and significant effect within the same year.

$\bar{\theta}$	$\tilde{\tau}$	$\tilde{\nu}$	$\beta_{p,1}$	$\beta_{p,2}$	$\beta_{p,3}$	$\beta_{p,4}$	$\beta_{p,5}$	$\beta_{p,6}$
0.019	0.021	1.359	0.284	0.176	-1.231	0.074	-1.089	0.025
(0.009)	(0.013)	(0.300)	(0.845)	(0.178)	(0.602)	(0.117)	(0.478)	(0.113)

Table 9: Meta-study selection estimates from minimum wage data, published studies, with standard errors in parentheses. Publication probability β_p is measured relative to omitted category of studies estimating a positive effect significant at the 5% level.

These estimates are consistent with our baseline model assuming that publication rules are constant over time, with a p-value of 0.7 for the test of the joint hypothesis that

$\beta_{p,2} = \beta_{p,4} = \beta_{p,6} = 0$.

B.7 Bias corrections based on applications

In this section, we plot our median unbiased estimators and corrected confidence sets, analogous to Figure 2 of the paper, based on the selection estimates from our applications. Corrections based on replication estimates from the Camerer et al. (2016) data are plotted in Figure 11. Corrections based on replication estimates from the Open Science Collaboration (2015) data are plotted in Figure 12. Corrections based on estimates using data from Wolfson and Belman (2015) are reported in Figure 13.

C Additional Empirical Results

This section provides additional empirical results to supplement those in the paper. Section C.1 describes a method of moments based estimation approaches that allow us to drop our parametric assumptions on the distribution of true effects, and reports results from these approaches in our applications. Section C.2 describes results from an additional replication application using data from Camerer et al. (2018), while Section C.3 describes results from an additional metastudy application using data from Croke et al. (2016).

C.1 Moment-based estimation results

In the main text we report estimates based on parametric specifications for the distribution of true effects in latent studies. To confirm that our results are robust to the choice of parametric specification, in this section we report estimates from moment-based approaches that require only that we specify a functional form for the publication probability p, and leave the distribution of true effects fully nonparametric. The moments used to obtain these estimators are motivated by the identification arguments in Section 3 of the paper.

We begin by introducing the moments we consider in the replication and meta-study settings, respectively, and then discuss results in our applications. Overall, we find that while moment-based approaches often yield less precise conclusions, our main findings are robust to dropping our parametric specifications for the distribution of true effects.

C.1.1 Estimation Moments

Replication Moments In our discussion of identification for settings with replication data in Section 3.1 of the main text, we noted that if the original and replication estimates have the same distribution in the population of latent studies, then absent selective publication the joint distribution of published and replication estimates will likewise be symmetric. This observation implies a moment restriction that can be used for estimation.

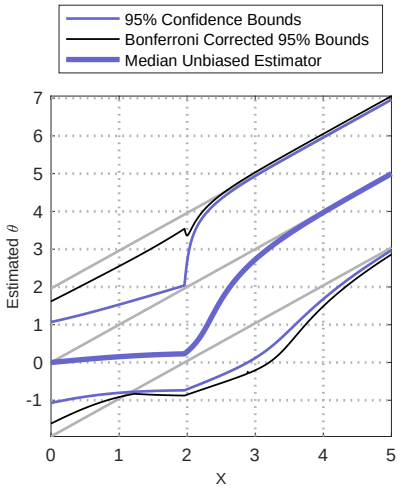

Figure 11: This figure plots 95% confidence bounds and the median unbiased estimator for the selection estimates based on replication data on economics lab experiments. The usual (uncorrected) estimator and confidence bounds are plotted in grey for comparison.

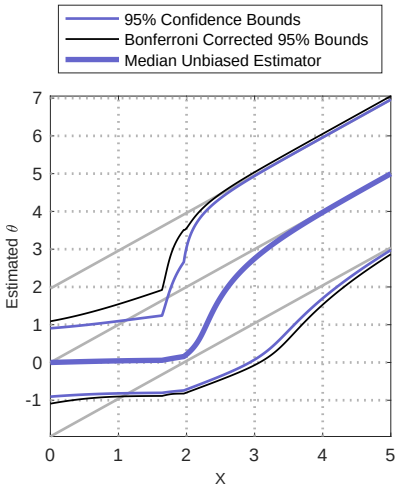

Figure 12: This figure plots 95% confidence bounds and the median unbiased estimator for the selection estimates based on replication data on psychology lab experiments. The usual (uncorrected) estimator and confidence bounds are plotted in grey for comparison.

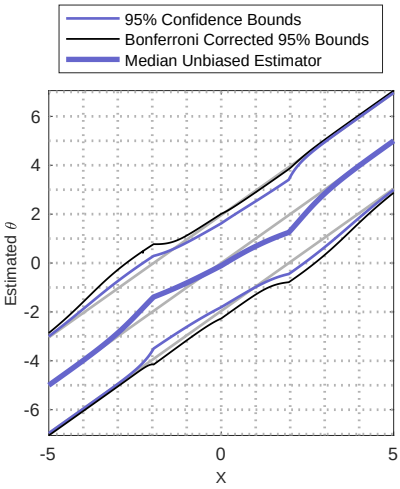

Figure 13: This figure plots 95% confidence bounds and the median unbiased estimator for the selection estimates based on metastudy data on the minimum wage. The usual (uncorrected) estimator and confidence bounds are plotted in grey for comparison.

To derive our moments, we first consider the case where $\Delta^* = \Sigma^{*r}/\Sigma^* = 1$ for all latent studies, so the original and replication studies have the same standard error. For any constants c_1, c_2

$$E[1\{|Z^*| > c_1, |Z^{r*}| \leq c_2\} - 1\{|Z^{r*}| > c_1, |Z^*| \leq c_2\}] = 0,$$

in the population of latent studies no matter the distribution μ of true effects.[24] In particular, this reflects our observation in the main text that, absent selection, we should observe an equal number of cases where the original results are significant and the replications are insignificant, and where the replication results are significant and the original results are insignificant, where we can consider results significant and insignificant at different levels.

We can recover the distribution of latent studies from the distribution of published studies by weighting by the inverse of the publication probability, $E[p(Z^*)]/p(Z)$. This implies the moment restriction

$$E\left[\frac{E[p(Z^*)]}{p(Z)}(1\{|Z| > c_1, |Z^r| \leq c_2\} - 1\{|Z^r| > c_1, |Z| \leq c_2\})\right] = 0$$

in the population of published studies. Since $E[p(Z^*)]$ does not vary across observations, the moment restriction continues to hold if we drop this term, yielding moments

$$E\left[\frac{1}{p(Z)}(\{|Z| > c_1, |Z^r| \leq c_2\} - 1\{|Z^r| > c_1, |Z| \leq c_2\})\right] = 0 \tag{11}$$

which depend only on observables and $p(\cdot)$ and so can be used to estimate $p(\cdot)$.

Thus far, in deriving moments we have assumed that $\Delta^* = 1$. In our applications, however, we in fact have $\Delta^*|Z^* \sim f_{\Delta^*|Z^*}$. If the distribution of Δ is bounded above by some value $\delta_{\max} \geq 1$, we can adapt the moments (11) to account for unequal variances by noising up both the original and replication estimates to noise level $\delta_{\max}$. In particular, for ε, ε^r i.i.d. $N(0,1)$ random variables,

$$E\left[\frac{1}{p(Z)}\begin{pmatrix} 1\left\{\left|Z + \sqrt{\delta_{\max}^2 - 1}\varepsilon\right| > c_1, \left|Z^r + \sqrt{\delta_{\max}^2 - \Delta}\varepsilon^r\right| \leq c_2\right\} - \\ 1\left\{\left|Z^r + \sqrt{\delta_{\max}^2 - \Delta}\varepsilon^r\right| > c_1, \left|Z + \sqrt{\delta_{\max}^2 - 1}\varepsilon\right| \leq c_2\right\} \end{pmatrix}\right] = 0.$$

To eliminate the added noise $(\varepsilon, \varepsilon^r)$ in these moments, we can take the conditional

[24] Here we focus on the absolute value of the original and replication estimates to avoid complications from the sign normalization in our replication applications.

expectation of each component given the data and define

$$h(z,\delta_1,z^r,\delta_2) = E\left[1\left\{\left|Z^* + \sqrt{\delta^2_{\max}-\delta_1}\varepsilon\right| > c_1, \left|Z^{*r} + \sqrt{\sigma^2_{\max}-\delta_2}\varepsilon^r\right| \leq c_2\right\} | Z^* = z, Z^{r*} = z^r\right]$$

$$= \left(1 - \Phi\left(\frac{c_1 - z}{\sqrt{\delta^2_{\max}-\delta_1}}\right) + \Phi\left(\frac{-c_1 - z}{\sqrt{\delta^2_{\max}-\delta_1}}\right)\right)\left(\Phi\left(\frac{c_2 - z^r}{\sqrt{\delta^2_{\max}-\delta_2}}\right) - \Phi\left(\frac{-c_1 - z^r}{\sqrt{\delta^2_{\max}-\delta_2}}\right)\right).$$

By the law of iterated expectations, we obtain the moment restrictions

$$E\left[\frac{1}{p(Z)}(h(Z,1,Z^r,\Delta) - h(Z^r,\Delta,Z,1))\right] = 0 \tag{12}$$

which depends only on observables and p and so can be used for estimation.

To use these moments in practice we need to choose a value of $\delta_{\max}$ and values for c. In our applications we below we take $\delta_{\max}$ to equal sample maximum of Δ, which is about 2.5 for the economics replications and about 2 for the psychology replications, and consider values c in each specification corresponding to the critical values used in p. Setting $\delta_{\max}$ to the sample maximum is ad-hoc, so as a further check we also report results based on the moments

$$E\left[\frac{1}{p(Z)}\left(\left(Z^2 - 1\right) - \left(Z^{r2} - \Delta^{r2}\right)\right)\right] = 0 \tag{13}$$

which can be shown to hold for any μ_Ω by arguments along the same lines as above and do not require that we select a value $\delta_{\max}$.

Metastudy Moments The moments we consider in our metastudy applications are derived using a similar approach. As noted in our discussion of metastudy identification in Section 3.2 of the text, absent selectivity in the publication process our assumptions imply that the distribution of effects for noisier studies is just a noised-up version of the distribution for less noisy studies. In particular, if we consider a pair of values σ_1, σ_2 with $\sigma_2 > \sigma_1$ and a pair of latent studies (i,i') then for any constant c and $\varepsilon \sim N(0,1)$

$$E\left[1\{X^*_i < c\Sigma^*_i\} - 1\left\{X^*_{i'} + \sqrt{\Sigma^{*2}_i - \Sigma^{*2}_{i'}}\varepsilon_< c\Sigma^*_i\right\} \middle| \Sigma^*_i = \sigma_1, \Sigma^*_{i'} = \sigma_2\right] = 0.$$

As above we can eliminate the noise from the added error ε. If we define

$$h(x,\sigma_2,\sigma_1) = E\left[1\left\{X_i^* + \sqrt{\sigma_2^2 - \sigma_1^2}\varepsilon < c\sigma_2^*\right\} | X_i^* = x\right] = \Phi\left(\frac{c\sigma_2 - x}{\sqrt{\sigma_2^2 - \sigma_1^2}}\right)$$

then the law of iterated expectations implies that

$$E[1\{X_i^* < c\Sigma_i^*\} - h(X_{i'}^*, \Sigma_{i'}^*, \Sigma_i^*) | \Sigma_i^* = \sigma_1, \Sigma_{i'}^* = \sigma_2] = 0.$$

As in the replication setting, to obtain moments which hold in the population of published studies, we can weight inversely by the publication probability (now for the pair X_i, $X_{i'}$), again dropping normalizing constants to obtain the moments

$$E\left[\frac{1}{p(X_i/\Sigma_i)}\frac{1}{p(X_{i'}/\Sigma_{i'})}(1\{X_i < c\Sigma_i\} - h(X_{i'}, \Sigma_{i'}, \Sigma_i))\middle| \Sigma_i' > \Sigma_i\right] = 0 \tag{14}$$

which depend only on $p(\cdot)$ and observables and so can be used for estimation.[25]

For estimation, we again consider values of c corresponding to the thresholds used in $p(\cdot)$. Since our moments hold for each pair (i,i') with $\Sigma_i > \Sigma_{i'}$, we average over all pairs of observations and obtain asymptotic distributions using results for estimators based on U-statistics from Honore and Powell (1994).

C.1.2 Empirical Applications

Economics laboratory experiments In our application to data on economics lab experiments from Camerer et al. (2016), we again model the publication probability as

$$p(Z) \propto \begin{cases} \beta_p & \text{if } |Z| \leq 1.96 \\ 1 & \text{otherwise.} \end{cases}$$

When we attempt to estimate β_p based on moments (12), we find that while the system of moments is just-identified and can be solved exactly, the zero of the sample moments corresponds to a negative value of β_p. This occurs because, unlike in likelihood estimation, the GMM moments do not automatically rule out negative values of β_p, though such values are meaningless under our model. Indeed, we see in simulation that even under correct specification negative point estimates arise with non-negligible probability for small sample

[25] In the sign-normalized case, as above we instead form moments based on the absolute value of X_i.

sizes and small values of β_p. To address this issue, in Table 10 we report 95% confidence sets based on Stock and Wright (2000), which are robust both to weak-identification and to parameter-on-the-boundary issues.

Robust CS, Baseline Moments		Robust CS, Alternative Moments	
β_p Lower Bound	β_p Upper Bound	β_p Lower Bound	β_p Upper Bound
0.000	0.049	0.000	∞

Table 10: Identification-robust 95% confidence sets for β_p for lab experiments in economics. The left panel reports results based on our baseline moments (12) for replication models, while the right panel reports results based on the alternative moments (13). Publication probability β_p is measured relative to omitted category of studies significant at the 5% level.

From these results, we see that when we consider our baseline moments (12) we obtain a robust confidence set roughly consistent with the estimate of β_p reported in the main text, even though we are fully relaxing our assumption on the distribution of latent effects. When we consider the alternative moments (13), by contrast, the moments are less informative, and the robust confidence set covers the full parameter space.

As before, instead of using the replication data we can instead focus just on the initial estimates and standard errors and apply our meta-study approach based on the moments (14). The results from this approach are reported in Table 11. For comparability with the replication results above we include both a conventional point estimate and standard error and an identification-robust confidence based on the generalization of Stock and Wright (2000) to the present U-statistic setting. These results are again broadly consistent with those obtained both from the replication moments above and from our likelihood estimates in the main text, showing strong selection in favor of statistically significant results.

Point Estimate	Robust CS	
β_p	β_p Lower Bound	β_p Upper Bound
0.040	0.000	0.177
(0.042)		

Table 11: Moment-based results for lab experiments in economics. The left panel reports an estimate and standard error based on our moments (14) for metastudy models, while the right panel reports a 95% identification-robust confidence set based on the same moments. Publication probability β_p is measured relative to omitted category of studies significant at the 5% level.

Psychology laboratory experiments Turning next to the data on lab experiments in psychology from Open Science Collaboration (2015), as in the main text we model the

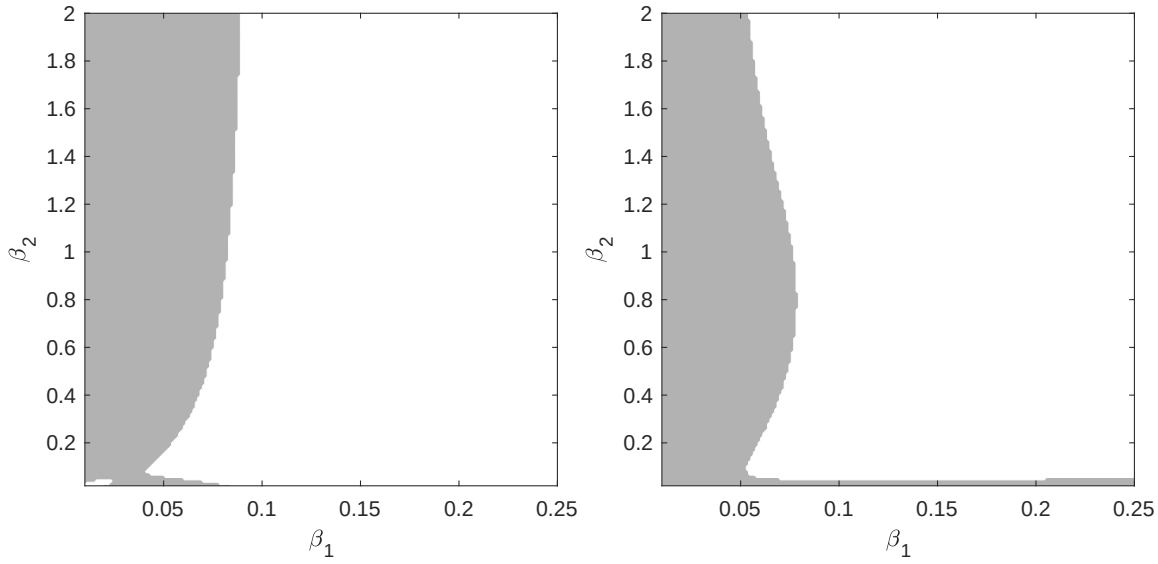

Figure 14: This figure plots 95% identification-robust joint confidence sets for $\beta_{p,1}$ and $\beta_{p,2}$ using data from lab experiments in psychology. The left panel plots results based on the baseline replication moments (12), while the right panel plots results based on the metastudy moments (14).

publication probability as

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

We find that identification of β_{p2} based on both our replication and metastudy moments appears weak in this setting. We report identification-robust joint confidence sets for (β_{p1},β_{p2}) based on Stock and Wright (2000) in Figure 14. While both confidence sets allow a wide range of possible values β_{p2}, only small values of $\beta_{p,1}$ are consistent with the confidence set based on replication data. On the other hand, results based on our meta-study approach allow a wide range of values for either parameter, though they rule out cases where both are large simultaneously. Both sets of results are consistent with our estimates in the main text, and in the case of the replications specification again provide evidence of selection against insignificant results.

To avoid specifying a value $\delta_{\max}$ to use in the moments (12), we can instead consider the moments (13). Since this yeilds only a single moment restriction, we consider selection only on significance at the 5% level, as in our application to economics lab experiments

above. Robust confidence sets from this specification are reported in Table 12. These

Robust CS, Alternative Replication Moments		Robust CS, Metastudy Moments	
β_p Lower Bound	β_p Upper Bound	β_p Lower Bound	β_p Upper Bound
0.000	0.045	0.000	0.115

Table 12: Identification-robust 95% confidence sets for β_p for lab experiments in psychology, assuming only selection on significance at the 5% level. The left panel reports results based on our alternative moments (13) for replication data, while the right panel reports results based on our metastudy moments. Publication probability β_p is measured relative to omitted category of studies significant at the 5% level.

results highlight that we still obtain informative results in this setting if we restrict attention to selection on significance at the 5% level.

Effect of minimum wage on employment For the data from Wolfson and Belman (2015) we consider the specification

$$p(X/\Sigma) \propto \begin{cases} \beta_{p1} & \text{if } Z < -1.96 \\ \beta_{p2} & \text{if } -1.96 \leq Z < 0 \\ \beta_{p3} & \text{if } 0 \leq Z < 1.96 \\ 1 & \text{if } Z \geq 1/96. \end{cases}$$

Table 13 reports estimates and standard errors. We see that the main message of our

$\beta_{p,1}$	$\beta_{p,2}$	$\beta_{p,3}$
1.174	0.231	0.235
(0.417)	(0.100)	(0.080)

Table 13: Meta-study selection estimates from GMM specifications for minimum wage data, with standard errors in parentheses. Publication probability β_p is measured relative to omitted category of studies estimating a positive effect significant at the 5% level.

likelihood results in this setting, that results finding a significant and negative effect of the minimum wage on employment are favored over insignificant results, again comes through clearly. In contrast to our likelihood results the point estimate for β_{p1} also suggests selection in favor of significant results finding a positive effect of the minimum wage on employment, but given the large standard error associated with this coefficient the results

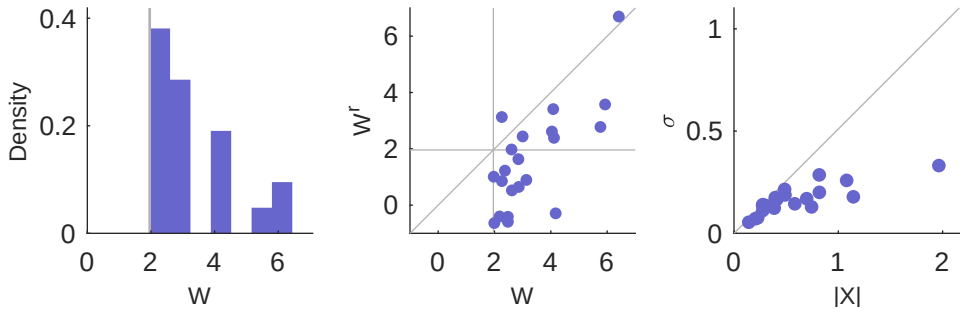

Figure 15: The left panel shows a binned density plot for the normalized z-statistics $W = |X|/\Sigma$ using data from Camerer et al. (2018). The grey line marks $W = 1.96$. The middle panel plots the z-statistics W from the initial study against the estimate W^r from the replication study. The grey lines mark W and $W^r = 1.96$, as well as $W = W^r$. The right panel plots the initial estimate $|X| = W \cdot \Sigma$ against its standard error Σ. The grey line marks $|X|/\Sigma = 1.96$.

are also consistent with selection on statistical significance alone ($\beta_{p1} = 1$, $\beta_{p2} = \beta_{p3}$), with a p-value of .86 for the joint test.

C.2 Results for Nature and Science replication study

This section describes results from applying our method to replication data from Camerer et al. (2018). This study replicated 21 social science studies published in the journals Science and Nature between 2010 and 2015. The authors selected results for replication based on significance at the 5% level, so the function $p(\cdot)$ must be interpreted as the probability that a result was both published and selected for replication. Nonetheless, we can still explore the question of selection in this setting by focusing on selection above the 5% significance threshold.

Histogram Before we discuss our formal estimation results, consider the distribution of originally published estimates $W = |Z|$, shown by the histogram in the left panel of Figure 15. This histogram shows that no results insignificant at the 5% are included in the sample, but it is not obvious from this plot whether there is evidence for selection above the 5% threshold.

Results from replication specifications The middle panel of Figure 15 plots the joint distribution of W, W^r in the replication data of Camerer et al. (2018). We consider

Replication				Meta-study			
κ	λ	$\beta_{p,1}$	$\beta_{p,2}$	$\tilde{\kappa}$	$\tilde{\lambda}$	$\beta_{p,1}$	$\beta_{p,2}$
0.211	1.653	0.000	0.480	0.070	0.663	0.000	0.583
(0.133)	(0.558)	(0.000)	(0.238)	(0.084)	(0.336)	(0.000)	(0.178)

Table 14: Selection estimates from social science experiments published in Nature and Science. The left panel reports estimates from replication specifications, while the right panel reports results from meta-study specifications. The parameters (κ,λ) and ($\tilde{\kappa}$,$\tilde{\lambda}$) are not comparable.

the model

$$|\Omega^*| \sim \Gamma(\kappa,\lambda), \quad p(Z) \propto \begin{cases} \beta_{p,1} & |Z| < 1.96 \\ \beta_{p,2} & 1.96 \leq |Z| < 2.58 \\ 1 & |Z| \geq 2.58. \end{cases}$$

To match the selection of results for replication we set $\beta_{p,1} = 0$ so results insignificant at the 5% level are always excluded from the sample, while we leave $\beta_{p,2}$ free so results significant between 5% and 1% level may be published with different probability than results significant at the 1% level. Fitting this model by maximum likelihood yields the estimates reported in the left panel of Table 14. These estimates therefore imply that results that are significant between the 5% and 1% level are about half as likely to be published as results significant at the 1% level.

Results from meta-study specifications The right panel of Figure 15 shows a meta-study plot for the Camerer et al. (2018) data. As for the replication case we consider the model

$$|\Theta^*| \sim \Gamma(\tilde{\kappa},\tilde{\lambda}), \quad p(Z) \propto \begin{cases} \beta_{p,1} & |Z| < 1.96 \\ \beta_{p,2} & 1.96 \leq |Z| < 2.58 \\ 1 & |Z| \geq 2.58. \end{cases}$$

Fitting this model by maximum likelihood (again with the restriction $\beta_{p,1}$=0) yields the estimates reported in the right panel of Table 14. Comparing these estimates to those in the left panel, we see that the estimates from the two approaches are again similar.

Bias corrections Figure 16 plots our median-unbiased estimates based on Camerer et al. (2018) against the replication estimates.

To interpret our results, we plot our median-unbiased estimates based on the Camerer

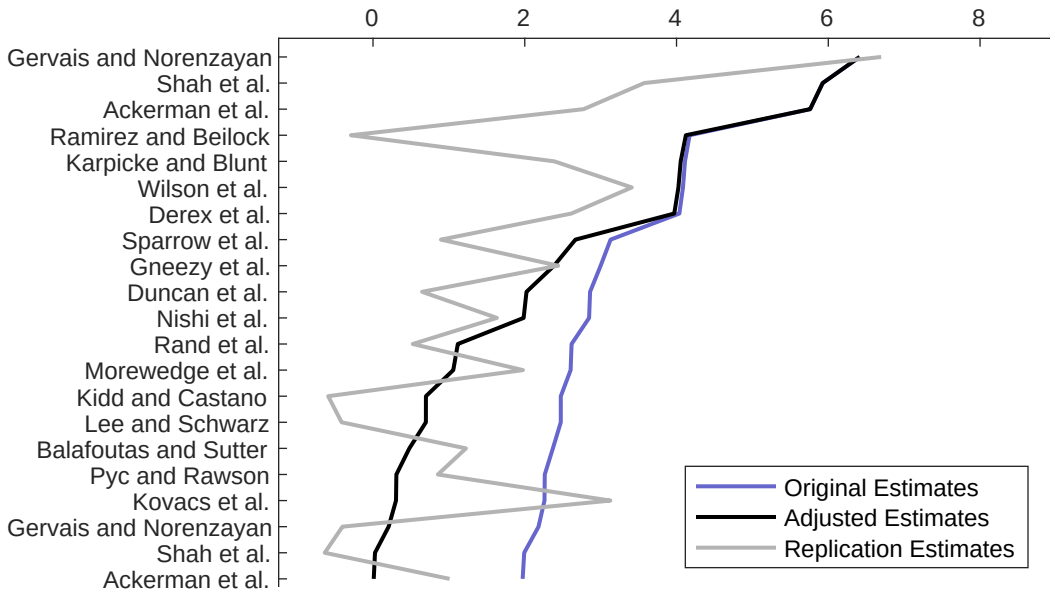

Figure 16: This figure plots the estimates W and W^r from the original and replication studies in Camerer et al. (2018), along with the median unbiased estimate $\hat{\theta}_{\frac{1}{2}}$ based on the estimated selection model and the original estimate.

et al. (2018) data in Figure 8. We see that the adjusted estimates track the replication estimates fairly well for studies with smaller original z-statistics. For studies with larger original z-statistics, our corrected estimates tend to be larger than the replication estimates.

C.3 Results for deworming meta-study

Here we report results based on data from the recent meta-study Croke et al. (2016) on the effect of mass drug administration for deworming on child body weight. They collect results from randomized controlled trials which report child body weight as an outcome, and focus on intent-to-treat estimates from the longest follow-up reported in each study. They include all studies identified by the previous review of Taylor-Robinson et al. (2015), as well as additional trials identified by Welch et al. (2017). They then extract estimates as described in Croke et al. (2016) and obtain a final sample of 22 estimates drawn from 20 studies, which we take as the basis for our analysis. For further discussion of sample construction, see Taylor-Robinson et al. (2015), Croke et al. (2016), and Welch et al. (2017). To account for the presence of multiple estimates in some studies, we again cluster by study.

Histogram Consider first the distribution of the normalized estimates Z, shown by the histogram in the left panel of Figure 17. Given the small sample size of 22 estimates, this histogram should not be interpreted too strongly. That said, the density of Z appears to jump up at 0, which suggests selection toward positive estimates.

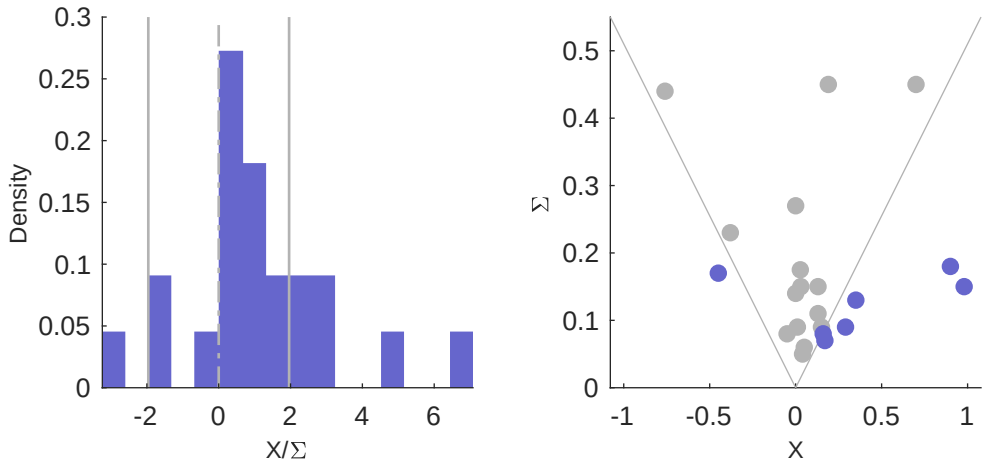

Figure 17: The left panel shows a binned density plot for the z-statistics X/Σ in the deworming metastudy data. The solid grey lines mark $|X|/\Sigma = 1.96$, while the dash-dotted grey line marks $X/\Sigma = 0$. The right panel plots the estimate X against its standard error Σ. The grey lines mark $|X|/\Sigma = 1.96$.

Results from meta-study specifications The right panel of Figure 17 plots the joint distribution of X, the estimated intent to treat effect of mass deworming on child weight, along with the standard error Σ in the Croke et al. (2016) data.

We next consider the model

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

where we constrain the the distribution of Θ^* to be normal and the function $p(\cdot)$ to be symmetric to limit the number of free parameters, which is important since we have only 22 observations. Fitting this model yields the estimates reported in Table 15.

The point estimates here suggest that statistically significant results are less likely to be included in the meta-study of Croke et al. (2016) than are insignificant results. However, the standard errors are quite large, and the difference in publication (inclusion)

$\bar{\theta}$	$\tilde{\tau}$	β_p
0.190	0.343	2.514
(0.120)	(0.128)	(1.869)

Table 15: Meta-study estimates from deworming data, with robust standard errors in parentheses. Publication probabilities β_p measured relative to omitted category of studies significant at 5% level.

probabilities between significant and insignificant results is itself not significant at conventional levels, so there is no basis for drawing a firm conclusion here. Likewise, the estimated $\bar{\theta}$ suggests a positive average effect in the population, but is not significantly different from zero at conventional levels.

We next consider the more flexible specification

$$\Theta^* \sim N(\bar{\theta},\tau^2), \quad p(Z) \propto \begin{cases} \beta_{p,1} & Z < -1.96 \\ \beta_{p,2} & -1.96 \leq Z < 0 \\ \beta_{p,3} & 0 \leq Z < 1.96 \\ 1 & Z \geq 1.96. \end{cases}$$

Results based on this specification are reported in Table 16. These estimates differ substantially from those reported above, and suggest strong selectivity against negative estimates, particularly negative and significant estimates. However, as can be seen from Figure 17 there is only a single negative and statistically significant estimate in the sample, so the reliability of conventional large-sample approximations here is highly suspect.

$\bar{\theta}$	$\tilde{\tau}$	$\beta_{p,1}$	$\beta_{p,2}$	$\beta_{p,3}$
-0.714	0.521	0.008	0.151	1.299
(0.626)	(0.206)	(0.025)	(0.207)	(1.113)

Table 16: Meta-study selection estimates from deworming wage data, flexible specification, with standard errors in parentheses. Publication probability β_p is measured relative to omitted category of studies estimating a positive effect significant at the 5% level.

To reduce the number of free parameters, we estimate a version of the model which does not allow discontinuities in $p(\cdot)$ based on statistical significance, but only based on

the sign of the estimate,

$$\Theta^* \sim N(\bar{\theta},\tau^2), \quad p(Z) \propto \begin{cases} \beta_p & X/\Sigma < 0 \\ 1 & Z \geq 0. \end{cases}$$

Fitting this model yields the estimates reported in Table 17. These estimates suggest strong selectivity on the sign of the estimated effect, where positive effects are estimated to be ten times more likely to be published than negative effects. While this is consistent with the distribution of observations in Figure 17, our choice of this specification was driven by our results in Table 16. Given that this is a form of specification search, it suggests that conventional asymptotic approximations may be unreliable here, and thus that these results should be treated with caution.

$\bar{\theta}$	$\tilde{\tau}$	β_p
-0.217	0.365	0.094
(0.156)	(0.103)	(0.099)

Table 17: Meta-study selection estimates from deworming wage data, restricted asymmetric specification, with standard errors in parentheses. Publication probability β_p is measured relative to omitted category of studies estimating a positive effect significant at the 5% level.

D Additional Theoretical Results

This section provides additional theoretical results to complement those in the main text. Section D.1 provides further discussion of meta-regression coefficients. Section D.2 discusses the extension of our baseline identification results to the case with selection on both Z^* and Σ^*. Section D.3 further generalizes our results to allow selection on both (Z^*,Ω^*), shows that we can still identify enough of the model in this setting to implement selection corrections, and develops specification tests based on this model. Section D.4 extends our corrected inference results to inference on scalar parameters in settings with multidimensional selection. Section D.5 discusses the effect of selection on Bayesian inference, and Section D.6 discusses optimal selection in the context of a simple model.

D.1 Interpretation of meta-regression coefficients

In Section 5.2 of the main text we discussed meta-regressions. We noted that under our assumptions meta-regressions deliver a valid test of the null of no selectivity. We also noted, however, that in the presence of selectivity the function $E[Z|1/\Sigma = \pi]$ is in

general non-linear, and the slope of the best linear predictor cannot be interpreted as a selection-corrected estimate of $E[\Theta^*]$.

To see this, consider the following simple example. Suppose that $\Theta^* \equiv \bar{\theta} > 0$, so there is no parameter heterogeneity across latent studies, and that $p(Z) = \mathbf{1}(Z > z^c)$, so there is strict selection on significant, positive effects. Let $\varepsilon \sim N(0,1)$, and let m be the inverse Mill's ratio, $m(x) = \frac{\varphi(x)}{1-\Phi(x)}$. Then

$$E[Z|1/\Sigma = \pi] = E[\pi\bar{\theta} + \varepsilon | \pi\bar{\theta} + \varepsilon > z^c] = \pi\bar{\theta} + m\big(z^c - \pi\bar{\theta}\big).$$

This is a nonlinear function of π, and the slope and intercept of the best linear predictor approximating this function both depend on the distribution of π (that is, of Σ). If Σ takes on only small values, and thus π only takes on large values, the Mill's ratio term is negligible, and $E^*[Z|1/\Sigma = \pi] \approx \pi\bar{\theta}$. If Σ takes on only large values, a first order approximation around $\pi = 0$ yields

$$E^*[Z|1/\Sigma = \pi] \approx m(z^c) + \bar{\theta}(1 - m'(z^c)) \cdot \pi.$$

This shows in particular that the slope, which in this example equals $\bar{\theta}(1 - m'(z^c))$, is in general different from the average effect $\bar{\theta}$, so that meta-regressions cannot be expected to deliver bias-corrected estimates of $E[\Theta^*]$.

D.2 Selection depending on Σ^* given Z^*

In this section we consider the extension of our identification results to the case where the publication probability takes the form $p(Z^*, \Sigma^*)$, so selection depends both on the z-statistic and the standard error. Such selection could arise, for instance, if journals prefer to publish precise estimates.

In settings with replication data, this extension is straightforward. In particular, we can treat the standard error Σ^* for the original study as a covariate and consider a conditional analysis as discussed in Section B.2. This will allow us to identify $p(z, \sigma)$ up to scale for each σ, as well as the conditional distribution of true effects given the standard error $\mu_{\Theta|\Sigma}$, though we cannot in general identify μ_Θ.

In settings with meta-study data, our identification argument exploits variation in Σ, so conditioning on Σ is not a viable option. If we impose multiplicative separability $p(X^*, \Sigma^*) = p_X(X^*) p_\Sigma(\Sigma^*)$, however, then we can still show identification of $p_X(\cdot)$.

specifically, if the assumptions of Theorem 3 still hold, save that the publication

probability is of the form $p(X^*,\Sigma^*)=p_X(X^*)p_\Sigma(\Sigma^*)$, then the resulting distribution of (X,Σ,Θ) is the same as that arising from the model with $p(X^*)=p_X(X^*)$ a modified distribution of Σ^* (obtained by reweighing the original distribution by $p_\Sigma(\cdot)$). Theorem 3 applied to this modified model shows that we can recover $p_X(\cdot)$ and μ_Θ.

D.3 Selection depending on (Ω^*,Σ^*) given Z^*

Selection of an empirical result for publication might depend not only on the result itself but also on other empirical findings reported in the same manuscript, or on unreported results obtained by the researcher. If that is the case, our assumption that publication decisions are independent of true effects conditional on reported results, $D\perp\Omega^*|Z^*$, may fail. Allowing for a more general selection probability $p(Z^*,\Omega^*,\Sigma^*)$, we can still identify $f_{X|\Omega,\Sigma}$, which is the key object for bias-corrected inference as discussed in Section 2.2.

This generalization allows publication decisions to depend on both the reported estimate and the true effect, and allows a wide range of models for the publication process. In particular, as we show in the next section this accommodates models where publication decisions depend on a variety of additional variables, including alternative specifications and robustness checks not reported in the replication dataset. Publication probabilities conditional on Z^*,Σ^* and Ω^* then implicitly average over these variables, resulting in additional dependence on Ω^*.

Theorem 4

Consider the setup of Theorem 2, but allow for

$$D|Z^*,\Omega^*,\Sigma^*\sim Ber(p(Z^*,\Omega^*,\Sigma^*)).$$

In this setup $f_{X|\Omega,\Sigma}$ is nonparametrically identified.

Proof of Theorem 4: Under the setup considered, and again denoting $\Delta=\Sigma^r/\Sigma$, using the implied conditional independence assumptions we get

$$\begin{aligned}f_{Z^r|Z,\Delta}(z^r,z,\delta)&=\int f_{Z^{*r}|\Delta^*,Z^*,D,\Omega^*}(z^r|\delta,z,1,\omega)f_{\Omega^*|\Delta^*,Z^*,D}(\omega|\delta,z,1)d\theta\\&=\int\varphi_\delta(z^r-\omega)f_{\Omega^*|Z^*,D}(\omega|z,1)d\omega\\&=(f_{\Omega|Z}*\varphi_\delta)(z^r|z).\end{aligned}$$

By deconvolution, this immediately implies that we can identify $f_{\Omega|Z}$. Since f_Z is directly identified, Bayes' rule yields the desired result via

$$f_{Z|\Omega}(z|\omega) = \frac{f_{\Omega|Z}(\omega|z) \cdot f_Z(z)}{\int f_{\Omega|Z}(\omega|z') \cdot f_Z(z') dz'}.$$

□

Note that in this proof we never appealed to normality of Z^* (though we did use normality of Z^{r*}). Hence, the result continues to apply in setting where Z^* is non-normal (for example due to manipulation of results or p-hacking).

D.3.1 Latent selection model

To make the generalized selection model introduced in the last section more concrete, this section shows how selection on (Z^*,Ω^*) can arise from selection on an unobserved variable. We then derive a parametric model of this form which we use as a specification test for our baseline model.

Assume that publication decisions are based on

$$\begin{pmatrix} Z^* \\ V^* \end{pmatrix} | \Theta^* \sim N\left(\begin{pmatrix} \Omega^* \\ \Omega^* \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right),$$

where V^* is a second, independent estimate of the true effect Ω^*, with the same variance as Z^*. Assume further that

$$D|Z^*,V^*,\Omega^* \sim Ber(p(Z^*,V^*)),$$

so publication decisions are based on Z^* and V^*. Since V^* is unobserved, integrating over its distribution gives publication probabilities of the form $p(Z^*,\Omega^*)$.

We want our parametric specification for $p(z,v)$ to nest our baseline specifications,

$$p(z) = \sum_{k=1}^{K} \beta_{p,k} 1\{\zeta_{k-1} \leq z < \zeta_k\}.$$

To ensure this, we consider the generalized specification

$$p(z,v) = \begin{array}{c} \sum_{k=1}^{K} \tilde{\beta}^1_{p,k} 1\{\zeta_{k-1} \leq z < \zeta_k, |v| \geq \zeta_V\} \\ + \sum_{k=1}^{K} \tilde{\beta}^0_{p,k} 1\{\zeta_{k-1} \leq z < \zeta_k, |v| < \zeta_V\}, \end{array}$$

which allows publication probabilities to depend on whether two-sided z-tests based on the latent variable v reject $\Omega^*=0$. Integrating over the distribution of V^* yields the following specification for $p(z,\omega)$:

$$p(z,\omega)=\begin{array}{c}\sum_{k=1}^{K}\tilde{\beta}^1_{p,k}1\{\zeta_{k-1}\leq z<\zeta_k\}\left(1-\tilde{\Psi}(\zeta_V,\omega)\right)\\ +\sum_{k=1}^{K}\tilde{\beta}^0_{p,k}1\{\zeta_{k-1}\leq z<\zeta_k\}\tilde{\Psi}(\zeta_V,\omega),\end{array}$$

where

$$\tilde{\Psi}(\zeta_V,\omega)=Pr\{|V|<\zeta_V|\Omega^*=\omega\}=\Phi(\zeta_V-\omega)-\Phi(-\zeta_V-\omega).$$

One can show that $p(z,\omega)$ is only nonparametrically identified up to a normalization for each value ω. Analogous to our baseline specifications, here we impose the normalization $\tilde{\beta}^1_{p,K}=\tilde{\beta}^0_{p,K}=1$. If we then define

$$\beta_{p,k}=\tilde{\beta}^1_{p,k}+\tilde{\Psi}(\zeta_V,0)\cdot(\tilde{\beta}^0_{p,k}-\tilde{\beta}^1_{p,k}),$$

$$\gamma_{p,k}=\left(\tilde{\beta}^1_{p,k}-\tilde{\beta}^0_{p,k}\right)\cdot\tilde{\Psi}(\zeta_V,0),$$

and

$$\Psi(\zeta_V,\omega)=\frac{\tilde{\Psi}(\zeta_V,\omega)-\tilde{\Psi}(\zeta_V,0)}{-\tilde{\Psi}(\zeta_V,0)},$$

we obtain the specification

$$p(z,\omega)=\sum_{k=1}^{K}(\beta_{p,k}+\gamma_{p,k}\cdot\Psi(\zeta_V,\omega))\cdot1\{\zeta_{k-1}\leq z<\zeta_k\}. \tag{15}$$

Note that our normalization now implies that $\beta_{p,K}=1$ and $\gamma_{p,K}=0$.

Specification test results Since the specification (15) nests our baseline specifications, we can use it to form Lagrange multiplier specification tests in our replication applications (testing the restriction $\gamma_{p,k}=0$ for all k). This specification test yields p-values of 0.53 and 0.42 in our Camerer et al. (2016) and Open Science Collaboration (2015) applications, respectively.

D.4 Inference when selection depends on multiple variables

This section extends the frequentist inference results developed in the main text to cases where publication decisions are based not just on a scalar, but instead on a normally dis-

tributed vector of estimates. Let X_i^* represent the estimates from study i, and assume that

$$X_i^*|\Theta_i^* \sim N(\Theta_i^*, \Xi)$$

for Ξ known. Assume that Ξ is constant across latent studies i; the generalization to the case where latent study i has variance Ξ_i^* is immediate. Since X_i^* is a vector, Ξ is a matrix. We thus obtain the following density for X^* given Θ^*:

Assumption 1
The distribution $f_{X^*|\Theta^*}(x|\theta)$ is multivariate normal with mean θ and variance Ξ:

$$f_{X^*|\Theta^*}(x|\theta) = (2\pi)^{-\frac{k}{2}} |\Xi|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\theta)'\Xi^{-1}(x-\theta) \right).$$

We consider inference on $\Gamma = v'\Theta$ for a known non-zero vector v, treating the other elements of Θ, denoted Ω, as nuisance parameters.[26] To conduct inference on the ith element of Θ we can simply take v to be the ith standard basis vector. To illustrate our results, we consider the example of difference in differences estimation, with selection on both statistical significance and a test for parallel trends.

D.4.1 Illustrative example: difference in differences

Suppose we observe data from two states, $s \in \{1,2\}$ over three time periods $t \in \{1,2,3\}$. Denote the average outcome for residents of state s at time t by Y_{st}, and note that under regularity conditions, Y_{st} will be approximately normally distributed

$$Y_{st} \sim N\left(\mu_{st}, \sigma_{st}^2\right).$$

For simplicity we assume that Y_{st} is independent of $Y_{s't'}$ if $s \neq s'$ or $t \neq t'$.

Suppose we are interested in estimating the effect of a particular state-level policy, and let D_{st} be a dummy for the presence of the policy in state s at time t. The difference in differences model (with no control variables) assumes that

$$\mu_{st} = \alpha_s + \beta_t + D_{st}\gamma.$$

If we are interested in the effect of a policy enacted in state 1 in period 3 and nowhere

[26] In an abuse of notation, this parameter Ω is unrelated to the variance-normalized effect Ω considered in other sections.

else in the sample, for example, we would take

$$D_{st} = \{s = 1, t = 3\}.$$

A key identifying assumption in the difference-in-differences model is that the only source of variation in μ_{st} at the state-by-time level is the policy change of interest. In particular, while we allow state fixed effects α_s and time fixed effects β_t, we rule out state-time-specific effects other than those acting through D_{st}. This is known as the parallel trends assumption.

With only two periods of data this assumption is untestable, since we have four free parameters $(\alpha_1, \alpha_2, \beta_2, \gamma)$ and only four means $(\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22})$. With data from an additional time period, however, we have five free parameters and six means and so can instead consider the model

$$\mu_{st} = \alpha_s + \beta_t + \tilde{D}_{st}\lambda + D_{st}\gamma$$

where

$$\tilde{D}_{st} = \{s = 1, t = 2\}$$

and the parallel trends assumption implies that $\lambda = 0$. Thus, given data from two states in three time periods the parallel trends assumption is testable.

Formal and informal tests of parallel trends are common in applications of difference in differences strategies. To describe a formal test in our setting, note that the natural estimator (G, L) for (γ, λ) has a simple form,

$$(G, L) = ((X_{13} - X_{12}) - (X_{23} - X_{22}), (X_{12} - X_{11}) - (X_{22} - X_{21})).$$

To test the parallel trends assumption in this setting we again want to test that λ, the mean of L, is equal to zero.

Consider a population of latent studies with the structure just described, and let us further simplify the model by setting $\sigma_{st} = 1$ for all t. For latent estimates $X^* = (G^*, L^*)$ and latent true effects $\Theta^* = (\Gamma^*, \Lambda^*)$,

$$\begin{pmatrix} G^* \\ L^* \end{pmatrix} \Bigg| \begin{pmatrix} \Gamma^* \\ \Lambda^* \end{pmatrix} \sim N\left(\begin{pmatrix} \Gamma^* \\ \Lambda^* \end{pmatrix}, \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix} \right)$$

where the covariance matrix is known.

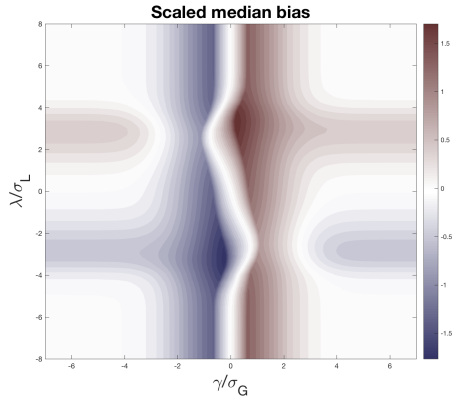

Figure 18: This figure plots the median bias of $(G)\ /\sigma_G$ in the difference in differences example.

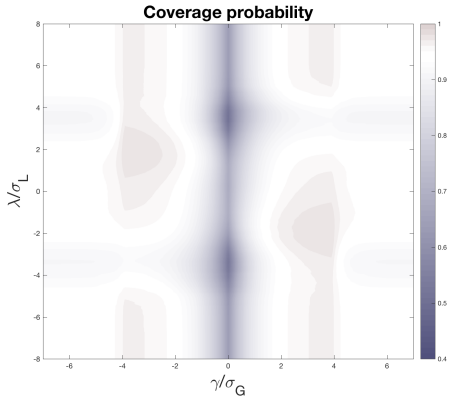

Figure 19: This figure plots the coverage of conventional 95% confidence sets in the difference in differences example.

As in our illustrative example in the main text, assume studies that reject $\gamma=0$ at the 5% level are ten times more likely to be published than studies that do not. In addition, assume studies that reject $\lambda=0$ at the 5% level are ten times *less* likely to be published than studies that do not. This leads to publication probability

$$p(X) \propto 1\left\{\frac{|G^*|}{\sigma_G} > 1.96, \frac{|L^*|}{\sigma_L} \leq 1.96\right\} 1 + 1\left\{\frac{|G^*|}{\sigma_G} > 1.96, \frac{|L^*|}{\sigma_L} \geq 1.96\right\} 0.1$$

$$+ 1\left\{\frac{|G^*|}{\sigma_G} \leq 1.96, \frac{|L^*|}{\sigma_L} \leq 1.96\right\} 0.1 + 1\left\{\frac{|G^*|}{\sigma_G} \leq 1.96, \frac{|L^*|}{\sigma_L} > 1.96\right\} 0.01.$$

This publication rule favors studies that find significant difference in difference estimates, and disfavors studies that reject the parallel trends assumption.

To illustrate the effect of selective publication in this setting, Figure 18 plots the median bias of G as an estimator for γ (scaled by the standard deviation σ_G of G^*). Selective publication results in large bias for the conventional estimator G, which depends on both the parameter of interest γ and the nuisance parameter λ. Analogously, Figure 19 plots the coverage of the usual two-sided confidence set $G^* \pm 1.96\sigma_G$, and shows that selective publication results in substantial coverage distortions.

D.4.2 Sufficient statistic for nuisance parameter

To conduct inference on γ, treating ω as a nuisance parameter, it will be helpful to derive a sufficient statistic for ω. Note that for $M(v)$ a $(\dim(X)-1)\times\dim(X)$ matrix such that

$M(v)\big(I-\frac{\Xi vv'}{v'\Xi v}\big)$ has full row-rank,

$$(G(x),W(x))=\left(v'x,M(v)\left(I-\frac{\Xi vv'}{v'\Xi v}\right)x\right)$$

is a one-to one transformation of x. Thus $(G,W)=(G(X),W(X))$ are jointly sufficient for θ, and rather than basing inference on X we can equally well base inference on (G,W). Note moreover that for $G^*=G(X^*)$ and $W^*=W(X^*)$, $X^*\sim N(\theta,\Xi)$ implies that

$$\begin{pmatrix} G^* \\ W^* \end{pmatrix}\sim N\left(\begin{pmatrix} \gamma \\ \omega \end{pmatrix},\begin{pmatrix} \sigma_G^2 & 0 \\ 0 & \Xi_W \end{pmatrix}\right) \tag{16}$$

for $\omega=M(v)\big(I-\frac{\Xi vv'}{v'\Xi v}\big)\theta$, $\sigma_G^2=v'\Xi v$, and $\Xi_W=M(v)\big(I-\frac{\Xi vv'}{v'\Xi v}\big)\Xi\big(I-\frac{vv'\Xi}{v'\Xi v}\big)M(v)'$. Thus the conditional distribution of G^* given W^* depends only on γ,

$$G^*|W^*\sim N(\gamma,\sigma_G^*),$$

and by conditioning on W^* we can eliminate dependence on the nuisance parameter ω. This property continues to hold for the conditional distribution of published G given W, as the following lemma shows.

Lemma 3

Under Assumption 1, the conditional density $G|W,\Gamma$ is given by

$$f_{G|W,\Gamma}(g|w,\gamma)=\frac{p(g,w)}{E[p(G^*,W^*)|W^*=w,\Gamma^*=\gamma]}\frac{1}{\sigma_G}\phi\left(\frac{g-\gamma}{\sigma_G}\right) \tag{17}$$

for ϕ the standard normal density, where we use the fact that (g,w) is a one-to-one transformation of x to write $p(g,w)=p(x(g,w))$.

Proof of Lemma 3 Note that we can draw from the conditional distribution $G|W=w,\Gamma=\gamma$ by drawing from the conditional distribution $G^*|W^*=w,\Gamma^*=\gamma$ and discarding the draw G^* with probability $1-p(G^*,w)$. The result then follows from Bayes rule. □

Thus, we see that the conditional density of G given W depends only on the parameter of interest γ and not on the nuisance parameter ω. Hence, by conditioning on W we can eliminate the nuisance parameter and conduct inference on γ alone.

D.4.3 Optimal quantile-unbiased estimates

To conduct frequentist inference, we generalize the median-unbiased estimator and equal-tailed confidence set proposed in Section 2.2 of the main text to the present setting. Using a result from Pfanzagl (1994) we show that the resulting quantile-unbiased estimators are optimal in a strong sense.

Formally, define $\hat{\gamma}_\alpha(X)$ by

$$F_{G(X)|W(X),\Gamma}(G|W,\hat{\gamma}_\alpha(X)) = \alpha.$$

This estimator is simply the value γ such that the observed G lies at the α quantile of the corresponding conditional distribution given W. The following theorem, based on the results of Pfanzagl (1994), shows that this estimator is both quantile-unbiased and, in a strong sense, optimal in the class of quantile-unbiased estimators.

Theorem 5
Let Assumption 1 hold, and assume further that the conditional distribution of G given W is absolutely continuous for all γ and almost every W, and that the parameter space for ω given γ contains an open set for all γ. Then

1. *The estimator $\hat{\gamma}_\alpha(X)$ is level-α quantile unbiased:*

$$Pr\{\hat{\gamma}_\alpha(X) \leq \gamma | \Theta = (\gamma,\omega)\} = \alpha \text{for all } \gamma,\omega,$$

2. *This estimator is uniformly most concentrated in the class of level-α quantile-unbiased estimators, in the sense that for any other level-α quantile unbiased estimator $\tilde{\gamma}(X)$ and any loss function $L(d,\gamma)$ that attains its minimum at $d=\gamma$ and is increasing as d moves away from γ,*

$$E[L(\hat{\gamma}_\alpha(X),\gamma)|\Theta=(\gamma,\omega)] \leq E[L(\tilde{\gamma}(X),\gamma)|\Theta=(\gamma,\omega)] \text{for all } \gamma,\omega.$$

Proof of Theorem 5 Since the multivariate normal distribution belongs to the exponential family, we can write

$$f_{G^*,W^*|\Theta^*}(g,w|\theta) = \tilde{h}(g,w)\tilde{r}(\gamma(\theta),\omega(\theta))\exp\big(\gamma(\theta)g + \omega(\theta)'w\big).$$

By the same argument as in the proof of Lemma 3, this implies that

$$f_{G,W|\Theta}(g,w|\theta) = h(g,w)r(\gamma(\theta),\omega(\theta))\exp(\gamma(\theta)g)\exp\big(\omega(\theta)'w\big) \tag{18}$$

for $h(g,w) = p(g,w)\tilde{h}(g,w)$ and

$$r(\gamma,\omega) = \frac{\tilde{r}(\gamma,\omega)}{E[p(X^*)|\Theta^* = \theta(\gamma,\omega)]}.$$

The density (18) has the same structure as (5.5.14) of Pfanzagl (1994), and satisfies properties (5.5.1)-(5.5.3) of Pfanzagl (1994) as well. Part 1 of the theorem then follows immediately Theorem 5.5.9 of Pfanzagl (1994).

Part 2 of the theorem follows by using Theorem 5.5.9 of Pfanzagl (1994) along with (18) to verify the conditions of Theorem 5.5.15 of Pfanzagl (1994). □

Using this result we see that $\hat{\gamma}_{\frac{1}{2}}(X)$ is the optimal median-unbiased estimator for the parameter of interest γ. A natural level-α confidence interval to accompany this estimator is then the equal-tailed confidence interval

$$CS = \big[\hat{\gamma}_{1-\frac{\alpha}{2}}(X), \hat{\gamma}_{\frac{\alpha}{2}}(X)\big].$$

Difference in differences example (continued) To illustrate our corrections in a multivariate setting, Figure 20 plots the difference between our median-unbiased estimator $\hat{\gamma}_{\frac{1}{2}}(X)$ and the conventional estimator $\hat{\gamma} = G$ in the difference-in-differences example. As this plot makes clear, $\hat{\gamma}_{\frac{1}{2}}(X)$ depends on both G and L. Thus, while we are interested only in the difference-in-differences parameter γ, the result for the pretest of parallel trends also plays a role in our estimate. Figure 21 plots the rejection region for a 5% test of $H_0 : \gamma = 0$ based on our equal-tailed confidence interval for γ. As this plot shows, the results of this test likewise depend on both G and L.

D.5 Bayesian inference

In the main text we discuss the effect of selective publication on frequentist inference on θ under known $p(\cdot)$. The effect of selective publication on Bayesian inference is more subtle, and depends on the prior. Here we briefly discuss Bayesian inference on θ under known $p(\cdot)$ for two natural classes of priors. These priors can be thought of as two extreme points of the set of relevant priors. For ease of exposition, we assume the standard deviation Σ is constant and suppress it in our notation. None of the results in this section rely on normality.

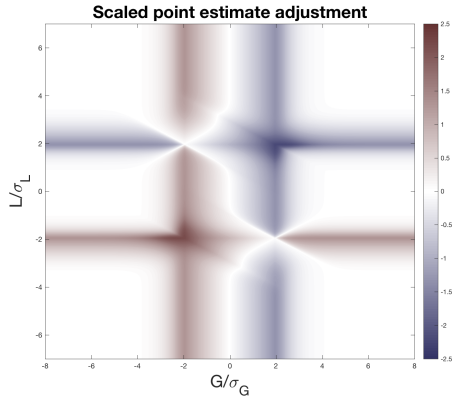

Figure 20: This figure plots the difference between the median-unbiased estiamtor $\hat{\gamma}_{\frac{1}{2}}(X)$ and the conventional estimator G in the difference-in-differences example.

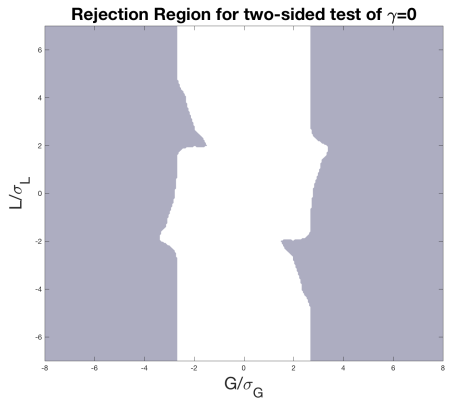

Figure 21: This figure plots the (shaded) rejection region for a 5% test of $H_0 : \gamma = 0$ based on equal-tailed confidence sets for γ in the differences in differences example.

Definition 2 (Two classes of priors)
Consider the following two classes of prior distributions π_μ for μ:

1. *Unrelated Parameters: π_μ is a point mass at some μ, so that μ is known and the prior distribution of Θ_i^* is i.i.d. across i.*

2. *Common Parameters: π_μ assigns positive probability only to point-measures μ, so that Θ_i^* is constant across i (equal to Θ_0^*) with probability* 1.

The unrelated parameters prior corresponds to the case where each latent study considers a different parameter. Thus, under priors in this class, learning the true parameter value Θ_i^* in latent study i conveys no information about the true parameter value $\Theta_{i'}^*$ in latent study i', and Θ_i^* is iid across i. The common parameters prior, by contrast, assumes that all latent studies attempt to estimate the same parameter Θ_0^*. Thus, priors in this class imply that Θ_i^* is perfectly dependent across i.

For both the unrelated and common parameters classes, the marginal prior π_{Θ^*} for Θ^* is unrestricted. For any π_{Θ^*} there is a unique prior in each class consistent with this marginal distribution.

If we observe a single draw X^*, our posterior for Θ^* depends only on the marginal prior π_{Θ^*}, and so is the same whether we consider the unrelated or common parameters priors. By contrast, when we observe a single draw X from the distribution of published papers, which class of priors we use turns out to be important. The following result is closely related to the findings of Yekutieli (2012).

Lemma 4 (Two posterior distributions)
Based on single observation of X, we obtain the following posteriors:

1. *Under unrelated parameters priors:*

$$f_{\Theta|X}(\theta|x) = f_{X^*|\Theta^*}(x|\theta)\cdot\pi_{\Theta^*}(\theta)/\pi_{X^*}(x)$$

2. *Under common parameters priors:*

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

$$\propto f_{X|\Theta}(x|\theta)\cdot\pi_{\Theta^*}(\theta)$$

Proof of Lemma 4:

1. Unrelated parameters: By construction $D \perp \Theta|X^*$, and thus

$$\begin{aligned} f_{\Theta|X}(\theta|x) &= f_{\Theta^*|X^*,D}(\theta|x,d=1) \\ &= f_{\Theta^*|X^*}(\theta|x) \\ &= f_{X^*|\Theta^*}(x|\theta)\cdot\pi_{\Theta^*}(\theta)/f_{X^*}(x). \end{aligned}$$

2. Common parameters: This follows immediately from the truncated likelihood (1).

□

Under the unrelated parameters prior, our posterior $f_{\Theta|X}(\theta|x)$ after observing $X=x$ is the same as our posterior had we observed $X^*=x$. The form of $p(\cdot)$ has no effect on our posterior distribution, and inference proceeds exactly as in the case without selection. Under the common parameters prior, by contrast, our posterior $f_{\Theta|X}(\theta|x)$ corresponds to updating our marginal prior π_{Θ^*} using the truncated likelihood $f_{X|\Theta}(x|\theta)$.

The fact that selection has no effect on our posterior under the common parameters prior may be surprising, but reflects the fact that under this prior, selection changes the marginal prior π_Θ for true effects in published studies. In particular, under this prior we have

$$\pi_\Theta(\theta) = \frac{E[p(X^*)|\Theta^*=\theta]}{E[p(X^*)]}\pi_{\Theta^*}(\theta),$$

which reflects the fact that the distribution of true effects for published studies differs from that for latent studies under this prior. When we update this prior based on observation of

X, however, the adjustment by $E[p(X^*)|\Theta^*=\theta]$ in the prior cancels that in the likelihood, and selection has no net effect on the posterior. Under the common parameters prior, by contrast, $\pi_{\Theta^*} = \pi_\Theta$, so the adjustment term in the prior due to selective inference continues to play a role in the posterior. For related discussion, see Yekutieli (2012).

D.6 Optimal selection for publication in a simple model

In the main text we discuss how to account for selective publication in inference and how to identify selectivity. It is natural to ask, however, whether selective publication is a good idea in the first place or just a misguided application of statistics leading to either publication bias or needlessly complicated inference. The answer to this question depends on the journal's objective function. One possibility is as follows. Suppose that published estimates are inputs into policy decisions, for instance in development economics, education, public finance, or medicine. If there are constraints on how many studies are published and read, then selectivity of the sort we observe might be justified.

We discuss a stylized version of this idea in a development economics context, though our model might also be considered a stylized description of medical publishing and doctors' prescriptions of treatments for patients. As in the last section we suppress the standard deviation Σ, and the results here do not rely on normality.

Suppose that each i corresponds to a different policy intervention. Suppose the distribution μ of true treatment effects Θ^* is known to journal editors and readers, and that the expected effect $E[\Theta^*]$ of a randomly chosen treatment on the likelihood of escaping poverty is non-positive. Suppose further that the journal is read by policy makers who aim to minimize poverty. Assume finally that each treatment is relevant for a population of equal size, normalized to 1. A policy maker wishes to implement a given treatment if the expected impact on the outcome considered is positive, conditional on the observed estimate $X=x$. Thus, their optimal treatment assignment rule is

$$t(x) = \mathbf{1}(E[\Theta|X=x] > 0), \tag{19}$$

which results in the expected outcome

$$v(x) = \max(0, E[\Theta|X=x]) \tag{20}$$

where $E[\Theta|X]$ is the policymakers' posterior expectation of Θ after observing X.[27] Suppose

[27] Perhaps surprisingly, truncation is irrelevant for this posterior expectation. This stems from the

the journal also aims to minimize poverty, but faces a marginal (opportunity) cost of c, in units comparable to treatment outcomes, when publishing a given study. Policymakers update their behavior only for published studies with $E[\Theta|X]>0$. This updated behavior results in an expected poverty reduction of $E[\Theta|X]$ relative to the status quo. It follows that the optimal publication rule for the journal is

$$p(X^*)=\mathbf{1}(E[\Theta^*|X^*]>c). \tag{21}$$

If the conditional expectation is monotonic in X^*, this rule is equivalent to

$$p(X^*)=\mathbf{1}(X^*>x_c),$$

so that results should get published if they are positive and "significant" relative to the critical value x_c, defined via $E[\Theta^*|X^*=x_c]=c$.

This result rationalizes selectivity in the publication process: the optimal rule derived here corresponds to one-sided testing. A more realistic version of this story allows for variation across latent studies in the variance of X^*, the cost of implementing treatment, the size of the populations to be treated, etc. All of these would affect the critical value x_c, which thus should vary across i and need not be equal to conventional critical values of hypothesis tests. What remains true, however, is that publication decisions that are optimal according to the above model are selective in a way which leads to publication bias, and correct inference needs to account for this selectivity.

Supplement References

Baricz, Á. (2008). Mills' ratio: Monotonicity patterns and functional inequalities. *Journal of Mathematical Analysis and Applications*, 340(2):1362–1370.

Brodeur, A., Cook, N., and Heyes, A. (2018). Methods matter: P-hacking and causal inference in economics. Working Paper.

Brodeur, A., Lé, M., Sangnier, M., and Zylberberg, Y. (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics*, 8(1):1–32.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai,

fact that we assume policy makers have unrelated parameters priors as in Definition 2 above.

T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., and Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280):1433–1436.

Croke, K., Hicks, J. H., Hsu, E., Kremer, M., and Miguel, E. (2016). Does mass deworming affect child nutrition? Meta-analysis, cost-effectiveness, and statistical power. Technical Report 22382, National Bureau of Economic Research.

Gilbert, D. T., King, G., Pettigrew, S., and Wilson, T. D. (2016). Comment on "Estimating the reproducibility of psychological science". *Science*, 351(6277):1037.

Hedges, L. V. (1992). Modeling publication selection effects in meta-analysis. *Statistical Science*, pages 246–255.

Honore, B. and Powell, J. (1994). Pairwise difference estimators for censored and truncated regression models. *Journal of Econometrics*, 64:241–278.

Lee, J. D., Sun, D. L., Sun, Y., Taylor, J. E., et al. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, 44(3):907–927.

Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. Springer.

Murphy, G. M. (2011). *Ordinary differential equations and their solutions*. Courier Corporation.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251):aac4716.

Pfanzagl, J. (1994). *Parametric Statistical Theory*. De Gruyter.

Stock, J. H. and Wright, J. H. (2000). Gmm with weak identification. *Econometrica*, 68(5):1055–1096.

Taylor-Robinson, D. C., Maayan, N., Soares-Weiser, K., Donegan, S., and Garner, P. (2015). Cochrane database of systematic reviews.

Wasserman, L. (2006). *All of nonparametric statistics*. Springer Science & Business Media.

Welch, V. A., Ghogomu, E., Hossain, A., Awasthi, S., Bhutta, Z. A., Cumberbatch, C., Fletcher, R., McGowan, J., Krishnaratne, S., Kristjansson, E., Sohani, S., Suresh, S.,

Tugwell, P., White, H., and Wells, G. A. (2017). Mass deworming to improve developmental health and wellbeing of children in low-income and middle-income countries: a systematic review and network meta-analysis. *The Lancet Global Health*, 5(1):e40–e50.

Wolfson, P. J. and Belman, D. (2015). 15 years of research on us employment and the minimum wage. *Available at SSRN 2705499*.

Yekutieli, D. (2012). Adjusted bayesian inference for selected parameters. *Journal of the Royal Statistical Society Series B*, 74(3):515–541.