For Online Publication: Supplement to the paper

Identification of and correction for publication bias

Isaiah Andrews Maximilian Kasy

May 30, 2019

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 Camerer et al. (2018) and Croke et al. (2016) 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

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 Proposition 1: We divide the proof of this proposition into two lemmas which might be of independent interest. In the first lemma we do not impose the assumptions of Proposition 1, so this result holds outside of the normal case. The second lemma then shows that the conditions of Proposition 1 imply the conditions of the first lemma. We omit conditioning on Σ throughout, since it makes no difference as shown by Equation (1).

Lemma A.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, $\hat{\omega}_\alpha(z) \leq \omega$ 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 A.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 A.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 Proposition 2: Denote

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

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 that as in Equation (3) of the paper

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

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 in 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 Proposition 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

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

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 (1) for the truncated density at $\pi = 1$ yields

$$g(z) = \partial_\pi \log f_{Z|\Sigma}(z|1/\pi)|_{\pi=1} = C_1 + \partial_\pi \log f_{Z^*|\Sigma^*}(z|1/\pi)|_{\pi=1} \tag{2}$$

for a constant C_1. Note how, as we differentiate $\log f_{Z|\Sigma}(z|1/\pi)$ with respect to π at a given value z, the term $p(z)$ drops out of the resulting equation (this plays the same role as the ratio (4) discussed in the paper). 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

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

from which we conclude (recall that $h(z) = f_{Z^*|\Sigma^*}(z|1)$)

$$h''(z) = (C_1 - 1 - g(z)) \cdot h(z) - z \cdot h'(z). \tag{3}$$

Equation (3) 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 (3); 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. The remainder of the proof shows that these parameters are identified as well.

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 $g(\cdot)$. We show below that calculations similar to those for the first derivative with respect to π yield the fourth order differential equation

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

To complete this proof, after deriving (4) we 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 (4).

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 (3) and (4). Differentiating equation (3) three times yields

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

and differentiating equation (4) yields

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

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.[1]

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

Remarks:

- The proof of Proposition 3 shows that our model is overidentified. If we consider higher order derivatives of equations (3) and (4), 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 $h(z)=const.\cdot\exp\Big(-\frac{1}{2(\tau^2+1)}(z-\bar{\mu})^2\Big)$. Denoting $e(z)=\partial_z \log h(z)$, we can rewrite equation (3) 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

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

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

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

- The proof of Proposition 3 could be equivalently stated in terms of linear operators rather than differential equations. In particular, the ordinary differential equations (3) and (4) 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 I.B 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 $\hat{\omega}_\alpha(Z;\beta)$ be the α-quantile unbiased estimator under β. For many specifications of $p(\cdot)$, and in particular for those used in our applications, $\hat{\omega}_\alpha(z;\beta)$ is continuously differentiable in β for all z. If we have a consistent and asymptotically normal estimator $\hat{\beta}$ for β, for $0<\delta<\alpha$, consider the interval

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

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 $\hat{\sigma}_U(z)$ are delta-method standard errors for $\hat{\omega}_{\frac{\alpha-\delta}{2}}\left(z;\hat{\beta}\right)$ and $\hat{\omega}_{1-\frac{\alpha-\delta}{2}}\left(z;\hat{\beta}\right)$, 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.[2]

[2]Even in cases where we do not have an asymptotically normal estimator for β, for example because

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 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 Proposition 1 conditional on C to obtain corrected estimates. Likewise, if we have replication estimates we can apply Proposition 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 $\mu_{\Theta|C}$. Finally, if Θ^* and Σ^* are independent conditional on covariates in the population of latent studies, we can apply Proposition 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 III.A and III.B, 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 B.1 *1. Consider the setup of Proposition 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 Proposition 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.

---

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 $\left[\inf_{\beta\in CS_\beta}\hat{\omega}_{\frac{\alpha-\delta}{2}}(Z;\beta),\sup_{\beta\in CS_\beta}\hat{\omega}_{1-\frac{\alpha-\delta}{2}}(Z;\beta)\right]$.

Proof of Corollary B.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 Proposition 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

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

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 $|\Omega^*|$. □

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 Proposition 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 $|\Theta^*|$. □

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

$$f_{Z,Z^r,\Delta}(z,z^r,\delta)=\frac{p(z)\int\varphi(z-\omega)\cdot\frac{1}{\delta}\varphi\big(\frac{z^r-\omega}{\delta}\big)d\mu_\Omega(\omega)}{\iint p(z')\cdot\varphi(z'-\omega)dz'd\mu_\Omega(\omega)}f_{\Delta^*|Z^*}(\delta|z). \tag{5}$$

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 $\mathcal{L}(p,\mu_\Omega)=\prod_{j=1}^J f_{Z,Z^r,\Delta}(z_j,z_j^r,\delta_j)$. To fit a given model, we maximize this likelihood with respect to $p(\cdot)$ and μ_Ω. Since $f_{\Delta^*|Z^*}$ enters multiplicatively,

it plays no role in maximum likelihood estimation of $p(\cdot)$ and μ_Ω. 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 B.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 B.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

$$f_{X,\Sigma}(x,\sigma) = \frac{p(\frac{x}{\sigma}) \cdot \int \varphi\left(\frac{x-\theta}{\sigma}\right) d\mu_\Theta(\theta)}{\int p(\frac{x'}{\sigma}) \cdot \varphi\left(\frac{x'-\theta}{\sigma}\right) dx' d\mu_\Theta(\theta)} f^*_\Sigma(\sigma). \tag{6}$$

Again denoting the total number of observations by J, this yields joint likelihood $\mathcal{L}(p,\mu_\Theta) = \prod_{j=1}^{J} f_{X,\Sigma}(x_j,\sigma_j)$, which we again use to estimate $p(\cdot)$ and μ_Θ. 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.[3] 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.[4] 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 effect 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

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

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

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

Results are reported in Table 1. 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.

Replication				Meta-study			
κ	λ	$\beta_{p,1}$	$\beta_{p,2}$	$\tilde{\kappa}$	$\tilde{\lambda}$	$\beta_{p,1}$	$\beta_{p,2}$
0.337	2.411	0.016	0.218	1.847	0.131	0.021	0.786
(0.233)	(1.057)	(0.021)	(0.336)	(1.582)	(0.065)	(0.030)	(1.497)

Table 1: 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.

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 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 2 we report parameter estimates using the replication and meta-study specifications discussed in Section III.B, 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.[5] Our results are broadly similar for this restricted

[5]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.

sample and for the full data.

Replication				Meta-study			
κ	λ	$\beta_{p,1}$	$\beta_{p,2}$	$\tilde{\kappa}$	$\tilde{\lambda}$	$\beta_{p,1}$	$\beta_{p,2}$
0.192	1.523	0.007	0.143	0.869	0.138	0.018	0.247
(0.012)	(0.367)	(0.005)	(0.078)	(0.648)	(0.059)	(0.012)	(0.142)

Table 2: 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.

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 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 III.B on the 51 endorsed replications, we obtain the estimates reported in Table 3. 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 1 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	$\beta_{p,1}$	$\beta_{p,2}$	$\tilde{\kappa}$	$\tilde{\lambda}$	$\beta_{p,1}$	$\beta_{p,2}$
0.433	1.231	0.016	0.358	0.634	0.198	0.022	0.440
(0.182)	(0.350)	(0.010)	(0.160)	(0.503)	(0.079)	(0.014)	(0.217)

Table 3: 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), Journal of Personality and Social Psychology (JPSP), and Journal of Learning Memory and Cognition (JLMC). In this section we estimate a model where we allow the publication rule to vary by journal,

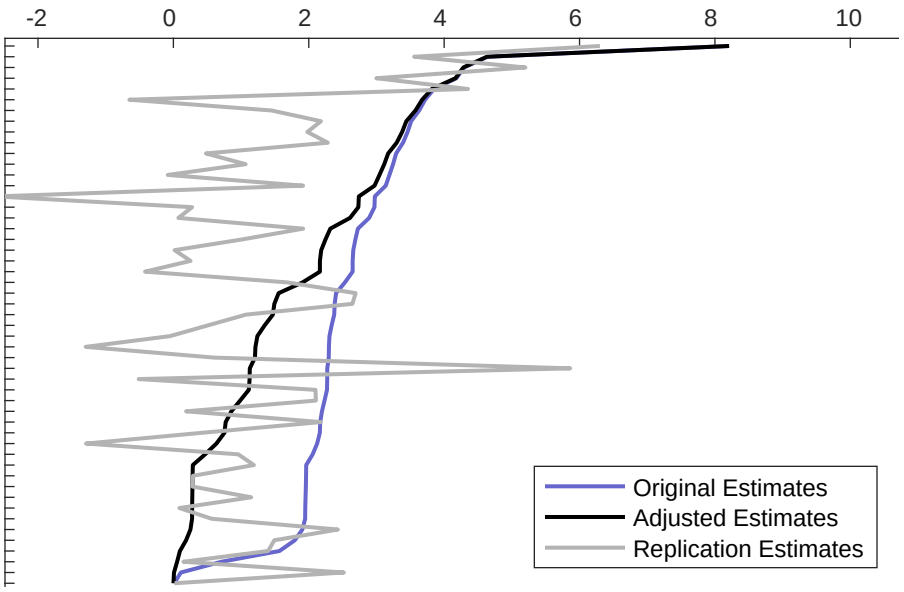

Figure 1: This figure plots the estimates W and W^r from the original and replication studies in Open Science Collaboration (2015), limiting the sample to approved replications, along with the median unbiased estimate $\hat{\theta}_{\frac{1}{2}}$ based on the estimated selection model and the original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

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 4. 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	$\beta_{p,1}$	$\beta_{p,2}$	$\beta_{p,3}$	$\beta_{p,4}$	$\beta_{p,5}$	$\beta_{p,6}$
0.311	1.314	0.008	0.002	-0.001	0.426	-0.287	-0.331
(0.117)	(0.294)	(0.008)	(0.011)	(0.011)	(0.244)	(0.263)	(0.259)

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
(0.565)	(0.054)	(0.014)	(0.019)	(0.026)	(0.320)	(0.350)	(0.364)

Table 4: 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 5 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 papers published by 2013, 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)	(0.378)	(0.332)	(0.147)	(0.140)

Table 5: 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)	(0.299)	(0.750)	(0.175)	(0.603)	(0.118)	(0.486)	(0.113)

Table 6: 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 2. Corrections based on replication estimates from the Open Science Collaboration (2015) data are plotted in Figure 3. Corrections based on estimates using data from Wolfson and Belman (2015) are reported in Figure 4.

C Additional Empirical Results

This section provides additional empirical results to supplement those in the paper. Section C.1 describes method of moments based estimation approaches that allow us to drop

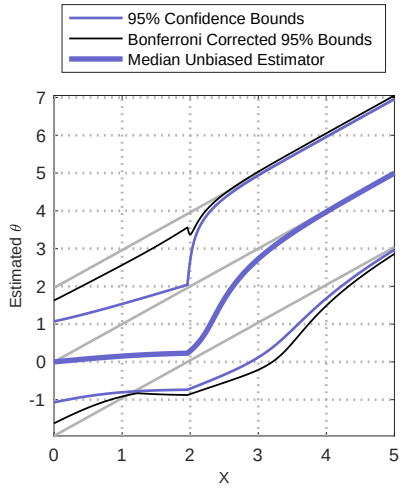

Figure 2: This figure plots 95% confidence bounds and the median unbiased estimator for the selection estimates based on replication data on economics lab experiments. The usual (uncorrected) estimator and confidence bounds are plotted in grey for comparison.

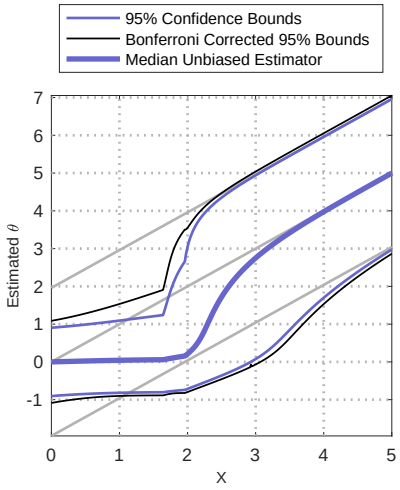

Figure 3: This figure plots 95% confidence bounds and the median unbiased estimator for the selection estimates based on replication data on psychology lab experiments. The usual (uncorrected) estimator and confidence bounds are plotted in grey for comparison.

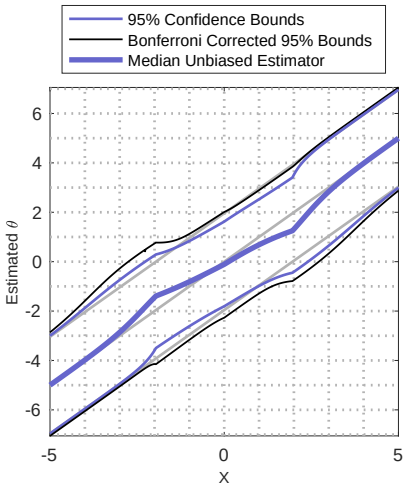

Figure 4: This figure plots 95% confidence bounds and the median unbiased estimator for the selection estimates based on metastudy data on the minimum wage. The usual (uncorrected) estimator and confidence bounds are plotted in grey for comparison.

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 II 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 II.A 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.

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.[6] 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

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

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

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

$$E\left[\frac{1}{p(Z)}(\{|Z|>c_1,|Z^r|\leq c_2\}-1\{|Z^r|>c_1,|Z|\leq c_2\})\right]=0 \tag{7}$$

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 (7) 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,

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

To eliminate the added noise $(\varepsilon,\varepsilon^r)$ in these moments, we can take the conditional expectation of each component given the data and define

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

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

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

$$E\left[\frac{1}{p(Z)}(h(Z,1,Z^r,\Delta)-h(Z^r,\Delta,Z,1))\right]=0 \tag{8}$$

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

$$E\left[\frac{1}{p(Z)}\left(\left(Z^2-1\right)-\left(Z^{r2}-\Delta^{r2}\right)\right)\right]=0 \tag{9}$$

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 II.B 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 $\varepsilon \sim N(0,1)$

$$E\left[1\{X_i^* < c\Sigma_i^*\} - 1\left\{X_{i'}^* + \sqrt{\Sigma_i^{*2} - \Sigma_{i'}^{*2}}\varepsilon < c\Sigma_i^*\right\} \middle| \Sigma_i^* = \sigma_1, \Sigma_{i'}^* = \sigma_2\right] = 0.$$

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

$$E\left[\frac{1}{p(X_i/\Sigma_i)}\frac{1}{p(X_{i'}/\Sigma_{i'})}(1\{X_i < c\Sigma_i\} - h(X_{i'}, \Sigma_{i'}, \Sigma_i)) \middle| \Sigma_i' > \Sigma_i\right] = 0 \tag{10}$$

which depend only on $p(\cdot)$ and observables and so can be used for estimation.[7]

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 (8), 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 very small 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,

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

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 sizes and small values of β_p. To address this issue, in Table 7 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	0.048	0.000	∞

Table 7: Identification-robust 95% confidence sets for β_p for lab experiments in economics. The left panel reports results based on our baseline moments (8) for replication models, while the right panel reports results based on the alternative moments (9). 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 (8) 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 (9), 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 (10). The results from this approach are reported in Table 8. 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 8: Moment-based results for lab experiments in economics. The left panel reports an estimate and standard error based on our moments (10) 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

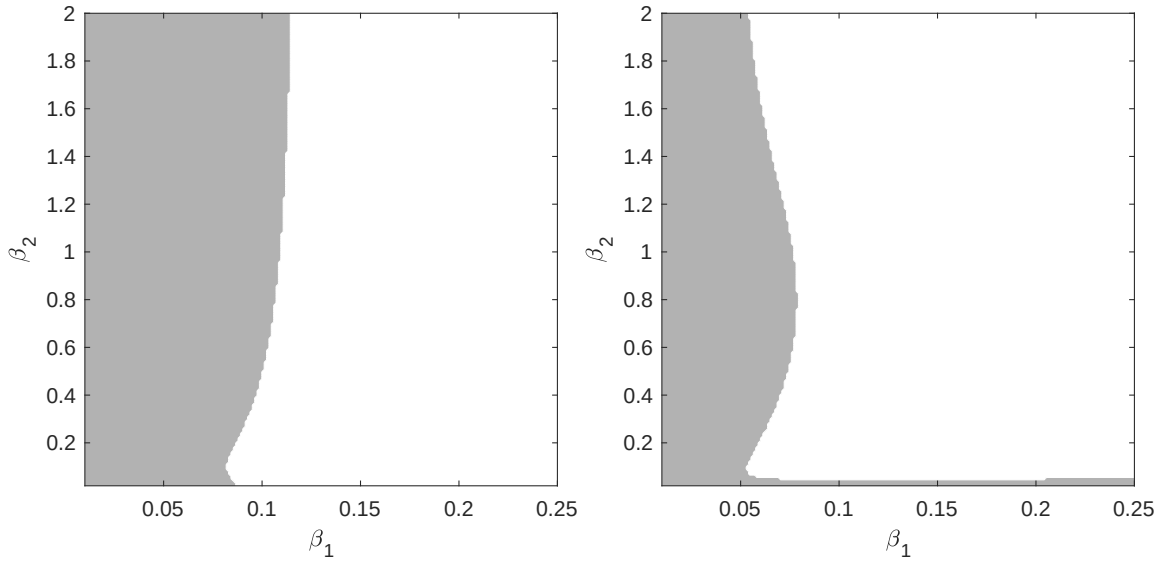

Figure 5: This figure plots 95% identification-robust joint confidence sets for $\beta_{p,1}$ and $\beta_{p,2}$ using data from lab experiments in psychology. The left panel plots results based on the baseline replication moments (8), while the right panel plots results based on the metastudy moments (10).

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 5. 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 (8), we can instead consider the moments (9). 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 9. These results highlight that we still obtain informative results in this setting if we restrict attention to selection on significance at the 5% level.

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	0.051	0.000	0.115

Table 9: 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 (9) 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.

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 10 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 10: 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 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,

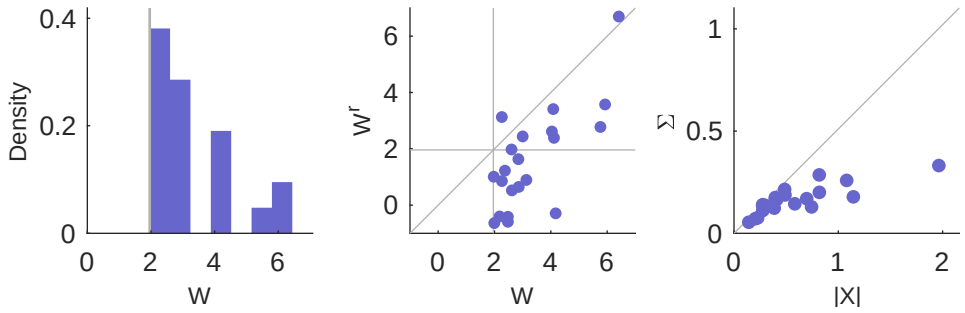

Figure 6: The left panel shows a binned density plot for the normalized z-statistics $W = |X|/\Sigma$ using data from Camerer et al. (2018). The grey line marks $W = 1.96$. The middle panel plots the z-statistics W from the initial study against the estimate W^r from the replication study. The grey lines 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 $|X|/\Sigma = 1.96$.

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 6. 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 6 plots the joint distribution of W, W^r in the replication data of Camerer et al. (2018). We consider 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 11. These estimates therefore imply that results 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 6 shows a meta-study plot for the Camerer et al. (2018) data. As for the replication case we consider the

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
(0.107)	(0.372)	(0.000)	(0.236)	(0.083)	(0.324)	(0.000)	(0.178)

Table 11: 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.

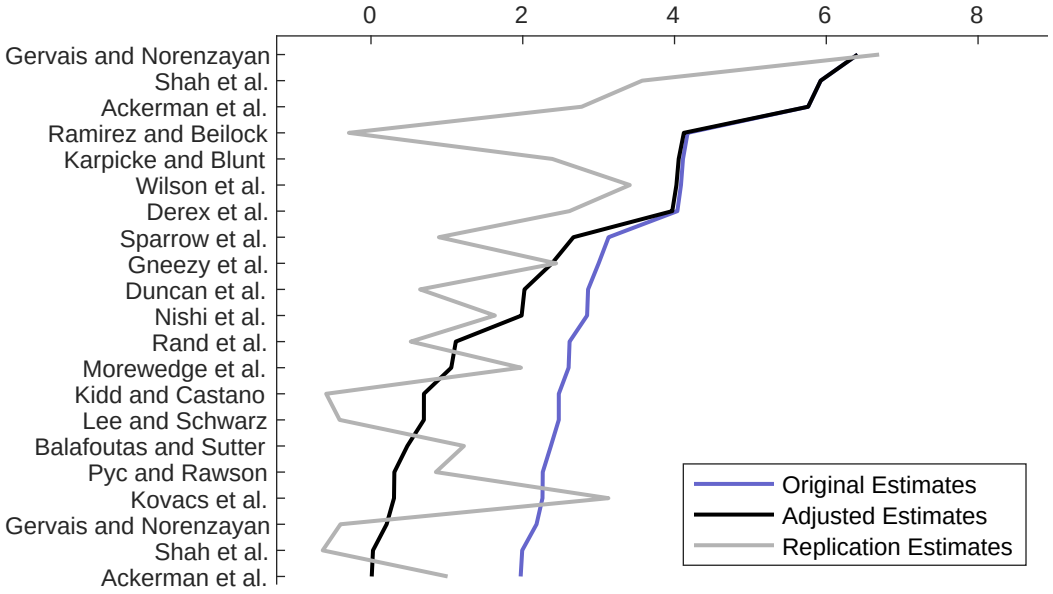

Figure 7: This figure plots the estimates W and W^r from the original and replication studies in Camerer et al. (2018), along with the median unbiased estimate $\hat{\theta}_{\frac{1}{2}}$ based on the estimated selection model and the original estimate.

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 11. Comparing these estimates to those in the left panel, we see that the estimates from the two approaches are again similar.

Bias corrections To interpret our results, we plot our median-unbiased estimates based on the Camerer 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 8. 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.

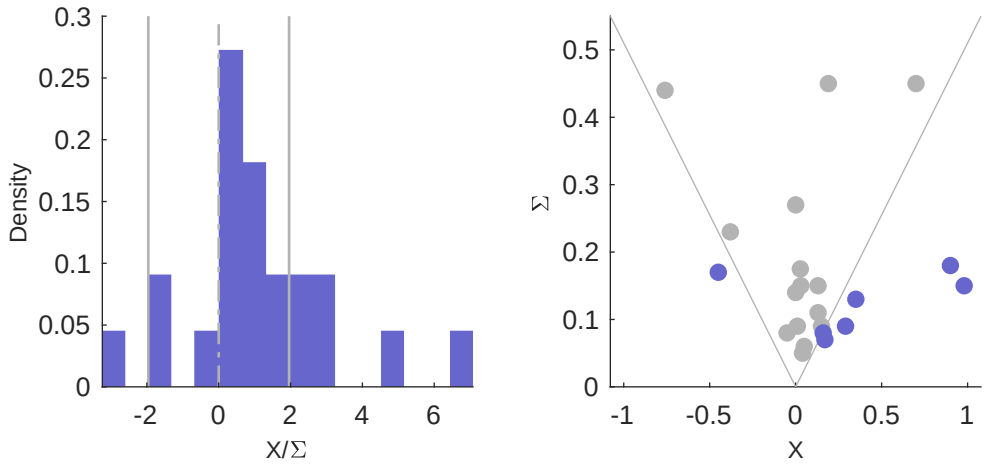

Figure 8: The left panel shows a binned density plot for the z-statistics X/Σ in the deworming metastudy data. The solid grey lines mark $|X|/\Sigma = 1.96$, while the dash-dotted grey line marks $X/\Sigma = 0$. The right panel plots the estimate X against its standard error Σ. The grey lines mark $|X|/\Sigma = 1.96$.

Results from meta-study specifications The right panel of Figure 8 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. Fitting this model yields the estimates reported in Table 12.

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

Table 12: 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.

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) 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 13. 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 8 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.

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,

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

$\bar{\theta}$	$\tilde{\tau}$	$\beta_{p,1}$	$\beta_{p,2}$	$\beta_{p,3}$
-0.714	-0.521	0.008	0.151	1.299
(0.625)	(0.206)	(0.025)	(0.207)	(1.112)

Table 13: 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.

Fitting this model yields the estimates reported in Table 14. 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 8, our choice of this specification was driven by our results in Table 13. 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 14: 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 (Z^*,Ω^*,Σ^*) and shows that we can still identify $f_{Z|\Omega}$ in this setting, which is enough to let us 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 (for example, selection on tests for pre-trends in difference in differences specifications). 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 IV.B 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 or μ_Σ.

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 $p(Z^*, \Sigma^*) = p_Z(Z^*) p_\Sigma(\Sigma^*)$, however, then we can still show identification of $p_Z(\cdot)$.

In particular, if the assumptions of Proposition 3 hold, save that the publication probability is of the form $p(Z^*, \Sigma^*) = p_Z(Z^*) p_\Sigma(\Sigma^*)$, then

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

just as in the case with selection on Z alone. The proof of Proposition 3 applied to this new density thus shows that we can recover $p_Z(\cdot)$ 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. Publication probabilities conditional on Z^*,Σ^* and Ω^* then implicitly average over these variables, resulting in additional dependence on Ω^*. Hence, 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 I.B.

Proposition D.1
Consider the setup of Proposition 2, but allow

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

In this setup $f_{Z|\Omega}$ is nonparametrically identified.

Proof of Proposition D.1: 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 settings 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 calculate a Lagrange multiplier test of our baseline model against a parametric model of this form. If our baseline model is correct this specification test will have correct size (in large samples), while the parametric model we adopt for selection on (Z^*,Ω^*) controls where the specification test will have power if our baseline model is incorrect.

Assume that publication decisions are based on

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

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

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

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 $p(z,\omega)$:

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

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

$$p(z,\omega) = \sum_{k=1}^{K} (\beta_{p,k} + \gamma_{p,k} \cdot \Psi(\zeta_V,\omega)) \cdot 1\{\zeta_{k-1} \leq z < \zeta_k\}. \tag{11}$$

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

Specification test results Since the specification (11) 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 0.48 and 0.38 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 inference results developed in the main text to cases where publication decisions are based not just on a scalar, but instead on a normally distributed 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.

Assumption D.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.[8] 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

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

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

If we are interested in the effect of a policy enacted in state 1 in period 3 and nowhere else in the sample, for example, we would take

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

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

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

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 $\Theta^*=(\Gamma^*,\Lambda^*)$,

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

where the covariance matrix is known.

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

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

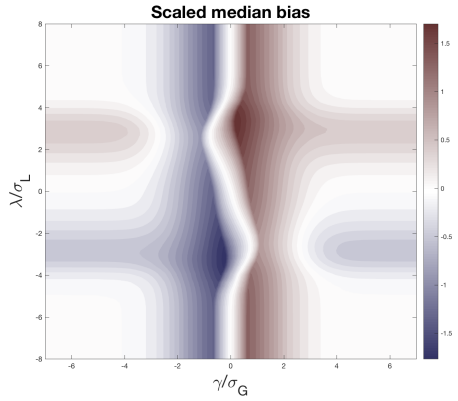

Figure 9: This figure plots the median bias of $(G)\ /\sigma_G$ in the difference in differences example.

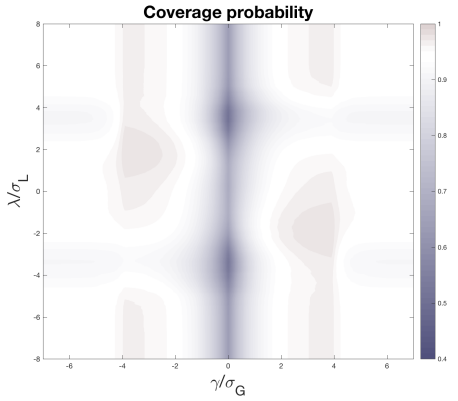

Figure 10: This figure plots the coverage of conventional 95% confidence sets in the difference in differences example.

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

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 9 plots the median bias of G as an estimator for γ (scaled by the standard error 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 10 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 $M(v)\left(I-\frac{\Xi vv'}{v'\Xi v}\right)$ 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

$$\begin{pmatrix} G^* \\ W^* \end{pmatrix}\sim N\left(\begin{pmatrix} \gamma \\ \omega \end{pmatrix},\begin{pmatrix} \sigma_G^2 & 0 \\ 0 & \Xi_W \end{pmatrix}\right) \tag{12}$$

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 D.3
Under Assumption D.1, the conditional density $G|W,\Gamma$ is given by

$$f_{G|W,\Gamma}(g|w,\gamma) = \frac{p(g,w)}{E[p(G^*,W^*)|W^*=w,\Gamma^*=\gamma]} \frac{1}{\sigma_G}\phi\left(\frac{g-\gamma}{\sigma_G}\right) \tag{13}$$

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 D.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 I.B 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 proposition, 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.

Proposition D.2
Let Assumption D.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 Proposition D.2 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 D.3, this implies that

$$f_{G,W|\Theta}(g,w|\theta)=h(g,w)r(\gamma(\theta),\omega(\theta))\exp(\gamma(\theta)g)\exp\big(\omega(\theta)'w\big) \tag{14}$$

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 (14) 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 proposition then follows immediately Theorem 5.5.9 of Pfanzagl (1994).

Part 2 of the proposition follows by using Theorem 5.5.9 of Pfanzagl (1994) along with (14) 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 11 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 12 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 error Σ is constant and suppress it in our notation. None of the results in this section rely on normality.

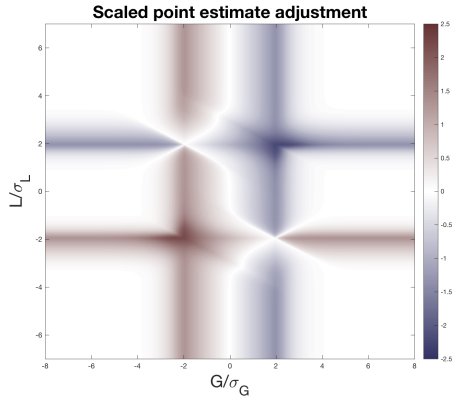

Figure 11: This figure plots the difference between the median-unbiased estiamtor $\hat{\gamma}_{\frac{1}{2}}(X)$ and the conventional estimator G in the difference-in-differences example.

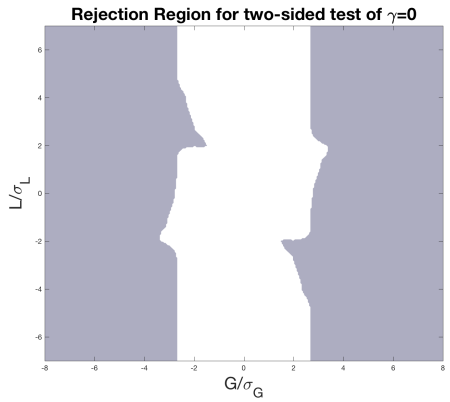

Figure 12: This figure plots the (shaded) rejection region for a 5% test of $H_0 : \gamma = 0$ based on equal-tailed confidence sets for γ in the differences in differences example.

Definition D.1 (Two classes of priors)
Consider the following two classes of prior distributions π_μ for μ_Θ:

1. *Unrelated Parameters: π_μ is a point mass at some μ_Θ, so that μ_Θ 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_Θ, 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 D.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 D.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 unrelated 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 error 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

$$t(x)=\mathbf{1}(E[\Theta|X=x]>0), \tag{15}$$

which results in the expected outcome

$$v(x)=\max(0,E[\Theta|X=x]) \tag{16}$$

where $E[\Theta|X]$ is the policymakers' posterior expectation of Θ after observing X.[9] Suppose 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

$$p(X^*)=\mathbf{1}(E[\Theta^*|X^*]>c). \tag{17}$$

If the conditional expectation is increasing 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 $E[\Theta^*|X^*=x_c]=c$.

[9] Perhaps surprisingly, truncation is irrelevant for this posterior expectation. This stems from the fact that we assume policy makers have unrelated parameters priors as in Definition D.1 above.

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.

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, 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.

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.