Austrian Journal of Statistics
MMMMMM YYYY, Volume VV, 1–xx.
http://www.ajs.or.at/

Infinite Diameter Confidence Sets in Hedges' Publication Bias Model

Jonas Moss
University of Oslo

Abstract

Meta-analysis, the statistical analysis of results from separate studies, is a fundamental building block of science. But the assumptions of classical meta-analysis models are not satisfied whenever publication bias is present, which causes inconsistent parameter estimates. Hedges' selection function model takes publication bias into account, but estimating and inferring with this model is tough. Using a generalized Gleser–Hwang theorem, we show there is no confidence set of guaranteed finite diameter for the parameters of Hedges' selection model. This result provides a partial explanation for why inference with Hedges' selection model is fraught with difficulties.

Keywords: meta-analysis, confidence intervals, file-drawer problem, publication bias, selection models, weight function models.

1. Introduction

A meta-analysis is a statistical analysis that quantitatively combines results from separate scientific studies. When the studies measure the same phenomenon, pooling of information allows us to predict the common effect size with larger precision than we could have done with one study alone. Meta-analyses are ubiquitous in the empirical sciences and forms a key component of most systematic reviews such as Cochrane's reviews (Higgins, Thomas, Chandler, Cumpston, Li, Page, and Welch 2019).

Most meta-analytic techniques assume honest and unbiased reporting of results. But there is ample evidence that the scientific literature is not unbiased, as studies with significant p-values tend to be published with a greater probability than other studies (Easterbrook, Berlin, Gopalan, and Matthews 1991), a phenomenon called publication bias by Sterling (1959) and the file-drawer problem by Rosenthal (1979). When publication bias is present, there is no reason to trust the results of meta-analytic methods that do not account for it, as the parameter estimates will be inconsistent (Carter, Schönbrodt, Gervais, and Hilgard 2019).

Hedges' (1992) publication bias model takes publication bias explicitly into account using a selection model, and is arguably the most appropriate model for publication bias (Carter *et al.* 2019). Despite there being an `R` (R Core Team 2020) package for maximum likelihood estimation of this model, called `weightr` (Coburn, Vevea, and Coburn 2019), the model has not yet taken off. Its maximum likelihood estimation methods are numerically unstable and its estimates can be off even when they converge (Coburn *et al.* 2019; Stanley 2005). The estimate of the mean effect size may be negative and of unrealistically large magnitude, and the estimated heterogeneity parameter might be improbably large. It turns out there are ridges in the likelihood function that can be linked to this behavior (McShane, Böckenholt, and Hansen 2016), but it has not been stated in clear terms exactly what the consequences are for inferential procedures. The purpose of this note is to show there is no confidence set for the mean effect size that has infinite diameter with probability 0. In the terminology of Berger, Liseo, and Wolpert (1999), Hedges' publication bias model belongs to the *Gleser–Hwang* class.

2. Hedges' publication bias model

The most popular and well-known meta-analysis method is the random effects model with normal likelihoods (Hedges and Vevea 1998). Written in hierarchical notation, it equals

$$\begin{aligned} \theta_i &\sim N(\theta_0, \tau), \\ x_i \mid \theta_i, \sigma &\sim N(\theta_i, \sigma_i). \end{aligned}$$

Here x_i is the effect size and σ_i is the standard deviation of the ith study, and $i = 1, \ldots, N$. Following the convention in meta-analysis, we assume all σ_is to be known. The mean parameter θ is the population effect size, $N(\theta_0, \tau)$ is the effect size distribution, and τ is the heterogeneity parameter. The purpose of the effect size distribution is to model the fact that most effect size estimates plugged into a meta-analysis do not appear to measure the same phenomenon. By integrating out θ_i, we find the density of x_i,

$$f(x_i; \theta_0, \tau, \sigma_i) = \phi(x_i; \theta_0, (\tau^2 + \sigma_i^2)^{1/2}),$$

where ϕ is the density of a normal random variable.

We will assume that the random effects meta-analysis model is true in the absence of publication bias. The mechanisms that cause publication bias modify the density in a suitable way. Consider the case when only significant studies at some specified level α are published. Assuming one-sided tests, the p-values are $u_i = \Phi(-x_i/\sigma_i)$, or normal one-sided p-values. We will only deal with with one-sided p-values in this paper, as there is usually, but not always, just one direction that is interesting to researchers, reviewers, and editors. A one-sided p-value can also be used if the researchers reported a two-sided value, since $p = 0.05$ for a two-sided hypothesis corresponds to $p = 0.025$ for a one-sided hypothesis, et cetera.

Define $c_\alpha = \Phi^{-1}(1-\alpha)$, the cutoff for significance at level α. The *basic publication bias model* is a truncated normal model with density

$$f(x_i; \theta_0, \sigma_i) = \Phi\left(\frac{\theta_0 - c_\alpha}{(\sigma_i^2 + \tau^2)^{1/2}}\right)^{-1} \phi(x; \theta_0, (\sigma_i^2 + \tau^2)^{1/2}) 1[x_i/\sigma_i > c_\alpha], \tag{2.1}$$

This model for publication bias was introduced by Hedges (1984) in the context of F-distributions.

The basic publication bias model is unrealistic. It requires that no non-significant studies are published. But even in the fields most severely affected by publication bias, such as psychology, a non-negligible number of non-significant studies are published (Motyl, Demos, Carsel, Hanson, Melton, Mueller, Prims, Sun, Washburn, Wong, Yantis, and Skitka 2017). Moreover, the basic publication bias model does not allow for different cutoffs for significance. For instance, it is likely that some editors will accept studies reaching a significance at $\alpha = 0.025$, corresponding to $x_i/\sigma_i > 1.96$ but not at $\alpha = 0.05$, corresponding to $x_i/\sigma_i > 1.64$.

These problems can be rectified by adopting the selection model for publication bias of Iyengar and Greenhouse (1988), which models the following scenario.

> **Publication bias scenario.** Alice the editor receives a study with the p-value u. Her publication decision is a random function of this p-value. That is, she will publish the study with some probability $w(u)$ and reject it with probability $1 - w(u)$. Every study you will ever read in Alice's journal has survived this selection mechanism, the rest are lost forever.

Let $w(u_i)$ be a function of the p-value $u_i = 1 - \Phi(-x_i/\sigma_i)$ taking values in $[0, 1]$. Then $w(u_i)$ is a probability for every u_i, and the selection model

$$f(x_i; \theta_0, (\sigma_i^2 + \tau^2)^{1/2}) \propto \phi(x_i; \theta_0, (\tau^2 + \sigma_i^2)^{1/2})w(u) \tag{2.2}$$

models the publication bias scenario exactly. This model can be viewed as a rejection sampling procedure (Flury 1990; von Neumann 1951), where ϕ serves as proposal distribution for f. Variants of this model, with and without covariates, has been studied by e.g. Dear and Begg (1992); Vevea and Hedges (1995); Vevea and Woods (2005); Citkowicz and Vevea (2017).

Hedges (1992) studies the selection model when w is a step function with fixed steps. Let α be a vector with elements $0 = \alpha_0 < \alpha_1 < \ldots < \alpha_K = 1$ and ρ be a K-ary non-negative, non-increasing vector having its first element equal to $\rho_1 = 1$. Define the step function w based on α and ρ as

$$w(u; \rho, \alpha) = \sum_{k=1}^{K} \rho^k 1_{(\alpha_{k-1}, \alpha_k]}(u). \tag{2.3}$$

We call the selection model with a step function *Hedges' publication bias model.* Its density is

$$f(x_i; \theta_0, (\tau^2 + \sigma_i^2)^{1/2}) \propto \sum_{k=1}^{K} \rho^k \phi(x_i; \theta_0, (\tau^2 + \sigma_i^2)^{1/2}) 1_{(\alpha_{k-1}, \alpha_k]}(u_i). \tag{2.4}$$

Interpreting Hedges' publication bias model is easy. When the editor receives a study with p-value u, she finds the k such that $u \in (\alpha_{k-1}, \alpha_k]$ and accepts with probability ρ^k. Since $\rho^1 = 1$, she always accepts when $u \in [0, \alpha_1]$. The vector ρ is non-increasing since a publication decision based solely on p-values should always act favorably towards lower p-values. The parameters (θ_0, τ, ρ) of the model are identified when α is fixed (Moss and De Bin 2020, Web Appendix A).

Hedges' publication bias model allows both for non-significant studies to be published and allows the editor to act differently towards different cutoffs such as $\alpha = 0.025$ and $\alpha = 0.05$. In addition, the model can approximate any non-increasing selection function w by increasing the number of steps.

We can write Hedges' model as a mixture model on the form

$$f(x_i; \theta_0, (\tau^2 + \sigma_i^2)^{1/2}) = \sum_{k=1}^{K} \pi^k f^k(x_i; \theta_0, (\tau^2 + \sigma_i^2)^{1/2}). \tag{2.5}$$

Here f^K is a normal density, $f^k, k < K$ are normal densities truncated to $[\Phi^{-1}(1-\alpha_{j-1}), \Phi^{-1}(1-\alpha))$, and π^k are mixture probabilities, i.e., $\pi^k > 0$ for each k and $\sum_{k=1}^{K} \pi^k = 1$. The mixture probabilities π^k are functions of $(\theta_0, \tau, \sigma_i, \rho)$, see the appendix (p. 10) for their formula.

The main benefit of Hedges' publication bias model (2.4) is how it models p-values based publication bias directly, there is no approximation involved. If you believe in the random effects meta-analysis model and the p-value based publication bias scenario, Hedges' publication bias model is simply the correct model. Most statistical methods correcting for publication bias in the literature either do not make use of an explicit statistical model or do not estimate the parameters θ_0 and τ. For instance, the funnel plot of Egger, Davey Smith, Schneider, and Minder (1997) is a graphical method, while the trim-and-fill method of Duval and Tweedie (2000) is a non-parametric method based on making the funnel plot symmetric. Stanley

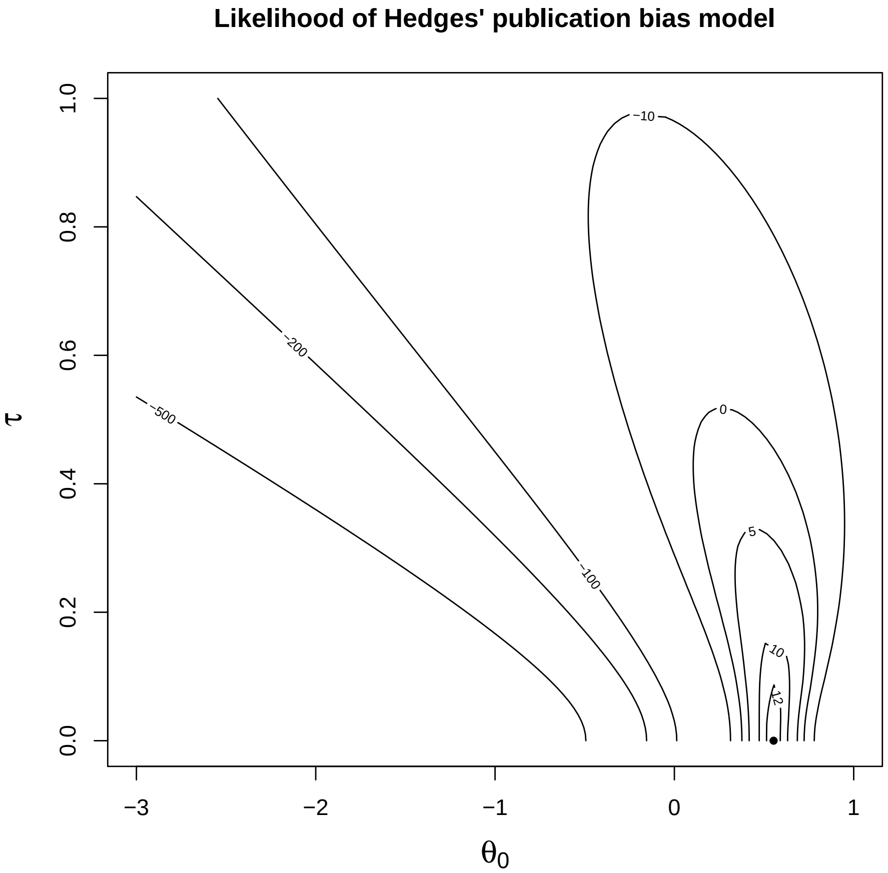

Figure 2.1: Contour lines for the log-likelihood for the simple publication model using power posing data of Cuddy *et al.* (2018).

(2005); Stanley and Doucouliagos (2014) discuss various misspecified regression-based estimators of the corrected effect size θ based on the fixed effect Hedges' publication bias model. The estimating p-curve method of Simonsohn, Nelson, and Simmons (2014) and p-uniform of van Assen, van Aert, and Wicherts (2015); van Aert, Wicherts, and van Assen (2016) are two methods for dealing with publication bias hailing from psychology. Both are based on a variant of the basic publication bias model, but with fixed instead of random effects, and both employ somewhat unusual estimation methods (McShane *et al.* 2016). Since there is ample evidence of heterogeneity in meta-analysis, restricting oneself to the fixed effects meta-analysis is a mistake.

Hedges' model has some downsides. It models only bias due to selection of p-values, not every source of bias, such as language bias (Egger and Smith 1998). Second, it may not be best model for biases with other causes than the publication process, such as p-hacking (Simmons, Nelson, and Simonsohn 2011). Moss and De Bin (2020) propose a related model that may be more successful at correcting for p-hacking.

Estimation of Hedges' model is difficult. Stanley (2005, Section 6.3) discusses three problematic cases in economics when maximum likelihood was used to estimate Hedges' publication bias model. McShane *et al.* (2016, Appendix A) notes that while estimation of the basic publication bias model is hard, introducing the heterogeneity parameter exacerbates the problem. The likelihood function has contours following approximately $\tau \propto (|\theta|)^{1/2}$. Figure 2.1 shows the contour lines for the meta-analysis of Cuddy, Schultz, and Fosse (2018) where the selection probabilities of the step function (2.3) are fixed at $\rho = (1, 0.6, 0.1)$. Only just around the maximum at $\hat{\theta}_0 = 0.55$ and $\hat{\sigma} = 5 \cdot 10^{-7}$ can the likelihood be approximated with a quadratic function.

3. Confidence sets of infinite diameter

Fix some measurable space $(\Omega, \mathcal{F})$, let $\mathcal{P}$ be a family of dominated probability measures defined on this measurable space, and Π a partition of $\mathcal{P}$. Recall that a partition of $\mathcal{P}$ is a collection of disjoint non-empty subsets π of $\mathcal{P}$ such that $\bigcup_{\pi\in\Pi} \pi = \mathcal{P}$. When p is a density associated with a $P \in \mathcal{P}$, we will use the standard notation $[p]$ to denote the unique π containing P.

Definition 1. A *confidence set* of level α is a family of rejection sets $\{R(\pi)\}, \pi \in \Pi$ such that

$$\sup_{\pi\in\Pi} \sup_{P\in\pi} P(R(\pi)) \leq \alpha.$$

If the inequality is an equality, the confidence set has *size* α.

This definition of confidence sets might look slightly unfamiliar, but it is a straight-forward generalization of the definitions in Casella and Berger (2002, Definition 9.1.5) and Lehmann and Romano (2005, Section 3.5). Usually, a confidence is defined as a set C adhering to the relation

$$\pi_0 \in C \iff \omega \notin R(\pi_0). \tag{3.1}$$

That is, $\pi_0 \in C$ if and only if we accept the null-hypothesis $H_0 : \pi = \pi_0$, but will only need the formulation using rejection sets in this paper. When confidence sets are defined in terms of rejection sets, there is sometimes no partition Π to take the supremum over, and the definition reduces to $\sup_{P\in\mathcal{P}} P(R(P)) \leq \alpha$. The term *confidence interval* is far more common than confidence set, but this requires that C of equation 3.1 is an interval, which we will not require here.

The following example should make Definition 1 familiar.

Example 2. Consider the usual t-confidence interval with n observations. Here $\mathcal{P}$ contains all measures $P^n_{\mu,\sigma}$, where $P_{\mu,\sigma}$ is the probability measure of a normal with mean μ and standard deviation σ. Furthermore, when Q is a probability measure, Q^n denotes the n-fold product measure of Q, corresponding to n independent samples from Q. Let $\pi(\mu) = \{P^n_{\mu,\sigma} \mid \sigma > 0\}$ contain all normal probability measures with mean μ and some positive standard deviation. Then $\{\pi(\mu)\}, \mu \in \mathbb{R}$ defines a partition of $\mathcal{P}$. A two-sided t-confidence interval is a confidence set of size α with partition $\{\pi(\mu)\}, \mu \in \mathbb{R}$ according to definition 1.

Now we must find out how to measure the size of confidence sets. Let a *size function* be any positive function $\|\cdot\| : \Pi \to [0, \infty)$. In the t-confidence interval example above, $\|\cdot\|$ can be taken to be $\|\pi\| = |\mu|$ for the unique μ associated with each π.

The diameter of a confidence set is the random variable

$$D(\omega) = \sup_{\pi\in\Pi}\{\|\pi\| \mid \omega \notin R(\pi)\}. \tag{3.2}$$

That is, the diameter tells you the size of the largest accepted π. We will assume that D is Borel measurable.

Definition 3. A confidence set has infinite diameter with P-positive probability if $P(D = \infty) > 0$. It has infinite diameter with positive probability if $P(D = \infty) > 0$ for all $P \in \mathcal{P}$.

This is a modified Gleser–Hwang theorem (Gleser and Hwang 1987, Theorem 1) formulated in terms of partitions. Its proof is in the appendix (p. 8).

Theorem 4 (Gleser–Hwang theorem)**.** *Suppose there is a sequence $\{p_n\}$ of densities from $\mathcal{P}$ satisfying the following:*

(i) There is a density $p^\star$ such that p_n converges to $p^\star$ pointwise,

(ii) $\operatorname{supp} p \supseteq \operatorname{supp} p^\star$ for all densities p derived from $\mathcal{P}$,

Table 1: The three partitions Π for the selection function model

	Symbol	Size $\|\cdot\|$	Confidence set
Mean effect size	θ_0	$\|\pi\| = \|\theta_0\|$	Confidence set for θ_0
Heterogeneity parameter	τ	$\|\pi\| = \tau$	Confidence set for τ
Both parameters	(θ_0, τ)	$\|\pi\| = (\theta_0^2 + \tau^2)^{1/2}$	Joint confidence set for (θ, τ)

(iii) the size of the equivalence class $[p_n]$ *goes to infinity as* n *increases,* $\|[p_n]\| \to \infty$.

Then every confidence set with level $\alpha > 0$ *has infinite diameter with positive probability.*

Following the terminology of Berger *et al.* (1999), we will say that families $\mathcal{P}$ of probabilities satisfying the conclusion of Theorem 4 for a suitable partition Π belong to the Gleser–Hwang class. To make the Gleser–Hwang class more familiar, we will present two examples. More examples can be found in the papers of Gleser and Hwang (1987) and Berger *et al.* (1999).

Fieller's problem is the best known case of a badly behaved confidence set. Let X, Y be an observation from a bivariate normal $N([\mu_1, \mu_2], I\sigma^2)$, where σ^2 is known. We want to form a confidence set for the ratio μ_2/μ_1. The most famous confidence set is due to Fieller (1940). His confidence set can be finite, the whole real line, or the union of two disjoint semi-infinite intervals (Koschat 1987), all with positive probability.

But it is not only Fieller's confidence set that might be infinitely long. The Gleser–Hwang theorem can be used to show that every confidence set for μ_2/μ_1 must be infinitely long with positive probability. This result is almost independent of the distribution of X and Y. To state this result in our notation, let $\mathcal{P}$ be a family of bivariate distributions over (X, Y). All of these distributions have the same support, and all of them have finite means $E_p(X)$ and $E_p(Y)$. Moreover, assume $E_p(X) = 0$ is attainable for some $p \in \mathcal{P}$. Define the partition Π by $p, q \in \pi$ if and only if $E_p(X)/E_p(Y) = E_q(X)/E_q(Y)$, and let $\|[p]\| = |E_q(X)/E_q(Y)|$, i.e., the ratio of means. Choose a sequence $p_n(x, y) = p(x, y)$, where $p(x, y)$ is density with means $E_pX > 0$ and $E_pY = 0$. Then $\|[p]\| = \infty$, the conditions of Theorem 4 are satisfied, and every confidence set with level $\alpha > 0$ has infinite diameter with positive probability.

Bahadur and Savage (1956) studies non-parametric testing of the mean, and concludes the mean cannot be meaningfully tested. They are working with a family $\mathcal{P}$ of densities over $\mathbb{R}$ that covers all finite means, has finite variances, and is closed under convex combinations. Similar problems were considered by Romano (2004) and Donoho (1988).

Using the Gleswer–Hwang theorem, it is easy to verify every confidence set has infinite diameter with positive probability. Define the partition Π by $p, q \in \pi$ if and only if $E_p(X) = E_q(X)$, and let $\|[p]\| = |E_p(X)|$. Let

$$p_n(x) = \left(1 - \frac{1}{n}\right) q_0(x) + \frac{1}{n} q_{n^2}(x),$$

where p_0 has mean 0 and q_{n^2} has mean n^2. Then $\|[p_n]\| = n$, the conditions of Theorem 4 are satisfied, and every confidence set with level $\alpha > 0$ has infinite diameter with positive probability.

There are several natural candidates for Π when working with Hedges' selection function model (2.4). We will work with three of them. First, consider the partition where $p, q \in \pi$ if and only if p, q have the same mean effect size parameter θ_0. We will equip this partition with the size function $\|\pi\| = |\theta_0|$, and it corresponds to a confidence set for θ_0. Second, consider the partition where all $p, q \in \pi$ have the same heterogeneity parameter τ, equipped with $\|\pi\| = \tau$. Finally, we will work with the partition where all $p, q \in \pi$ have the same heterogeneity parameter τ and population effect size θ_0, and equip it with $\|\pi\| = (\theta_0^2 + \tau^2)^{1/2}$. This information is summarized in Table 1 for convenience.

Let us take a look at the basic publication bias model (2.1) again. To use Theorem 4 we need a witnessing sequence of functions $p_n \to p$ satisfying the conditions (ii) and (iii). The next lemma shows how to make such a witness for the truncated normal. Its proof is in the appendix (p. 9).

Lemma 5. *Let f_n be a normal density truncated to $[a,b)$, where $b=\infty$ is allowed, with underlying mean $\theta_n = -n$ and standard deviation $\sigma_n^2 = n^2 + c$ for some $c \in \mathbb{R}$. Then f_n converges pointwise to $\exp(-x)/[\exp(-a)-\exp(-b)]$, the distribution of an exponential variable truncated to $[a,b)$.*

Using Lemma 5 it is not hard to show that the basic publication bias model (2.1) is a member of the Gleser–Hwang class.

Theorem 6. *Assume n independent samples from the basic publication bias model (2.1). Then, for any n, any confidence set for θ_0, τ, or (θ_0,τ) with level $\alpha > 0$ will have infinite diameter with positive probability.*

Proof. Let Π be the partition of $\mathcal{P}$ where $p,q \in \pi$ if and only if they share the same θ_0, and let $||[p]|| = |\theta_0|$. We are dealing with products of densities of the form (2.1), that is,

$$p(x) = \prod_{i=1}^{N} \Phi\left(\frac{\theta_0 - c_\alpha}{(\sigma_i + \tau)^{1/2}}\right)^{-1} \phi(x_i; \theta_0, (\sigma_i^2 + \tau^2)^{1/2}),$$

where σ_i are known parameters. From Lemma 5, p_n converges to a product of truncated exponentials when $\theta_n = -n$ and $\tau_n^2 = n$. Since $||[p_n]|| = n$, the three conditions of Theorem 4 are satisfied. The proofs for $||[p]|| = \tau$ and $||[p]|| = (\tau^2 + \theta_0^2)^{1/2}$ are similar and omitted. □

Proving the analogue of Theorem 6 for Hedges' publication is only somewhat more involved. We will use the mixture representation of (2.5) and a lemma generalizing Theorem 4 to a certain kind of mixtures.

Let $f^1, f^2, \ldots, f^K$ be a sequence of densities, $\pi = (\pi^1, \pi^2, \ldots, \pi^K)$ a probability vector, and $p = \sum_{k\leq K} \pi^k f^k$ a mixture distribution. We will assume that the size of $[p]$ equals the size of any of its mixture components $[f^k]$ for some size $\|\cdot\|$, i.e., $\|[p]\| = \left\|[f^k]\right\|$ for all k. Why we do this will be clear in the proof of Theorem 8, but think of it this way: If all of ps mixture components have the same mean, the mean of p equals the mean of any f^k.

Lemma 7. *Let $\mathcal{P}$ be a class of K-ary mixture distributions and $||[p]||$ be the assumption above. Assume there is a sequence $p_n = \sum_{k\leq K} \pi_n^k f_n^k$ and a subset K' such that*

(i) For all $k \in K'$, there is a density $f^{k\star}$ such that f_n^k converges to $f^{k\star}$ pointwise.

(ii) For all mixtures p, $\operatorname{supp} p \supseteq \operatorname{supp} f^{k\star}$ for some $k \in K'$.

(iii) For all $k \in K'$, the size of $[f_n^k]$ goes to infinity, $\left\|[f_n^k]\right\| \to \infty$.

(iv) The density concentrates on the components indexed by K', $\lim_{n\to\infty} \sum_{k\in K'} \pi_n^k = 1$.

Then every confidence set with level $\alpha > 0$ has infinite diameter with positive probability.

Proof. We employ Theorem 4. By (i) and (iv), p_n converges pointwise to the density $\sum_{k\in K'} \pi_n^{k\star} f_n^{k\star} = p^\star$. That $\operatorname{supp} p \supseteq p^\star$ follows from (ii) and (iv). Finally, from the assumption that $\|[p_n]\| = \left\|[f_n^k]\right\|$, we get that $\|[p_n]\| \to \infty$ too. □

Theorem 8. *Assume n independent samples from the publication bias model (2.1), where the selection probability ρ is unknown and α is known. Then any confidence set for θ_0, τ, or (θ_0, τ) will have infinite diameter with positive probability.*

Proof. Let $n = 1$ and consider confidence sets for θ_0. Let Π be the partition of $\mathcal{P}$ where $p, q \in \pi$ if and only if they share the same θ_0, and let $||[p]|| = |\theta_0|$. Then $||f^k|| = |\theta_0|$ from the mixture representation (2.5). Using Lemma 5, we see that $f^k, k > 1$ converges pointwise to a truncated exponential when $\theta_0 = -n$ and $(\tau^2 + \sigma^2)^{1/2} = n^{1/2}$, so that (i), (ii), (iii) of Proposition 7 are satisfied with $K' = K\backslash\{1\}$. The mixture probabilities for $k \neq K$ can be fixed at e.g. $\pi = 1/(K-1)$, and (iv) is satisfied as well. The proofs for $||[p]|| = \tau$ and $||[p]|| = (\tau^2 + \theta_0^2)^{1/2}$ are similar and omitted. When $N > 1$, expand the expression $\prod_{j=1}^{n} \sum_{k \leq K} \pi_i^k(\sigma_j) f_i^k(\sigma_j)$,and use the same reasoning as in the first part of this proof. □

4. Remarks

Well-behaved confidence sets for Hedges publication bias model do not exist, but well-behaved credibility sets do. Bayesian estimation of Hedges' model can be made routine, as it is easy to find uncontroversial priors for θ_0 and τ. In practical meta-analyses we know that θ_0 cannot be large, and is likely to be close to 0. Moreover, since it is common effects in meta-analyses to be interpreted as the aggregation of many small effects, the central limit theorem justifies using a normal prior. As we want to remove prior mass from negative θ_0s of large magnitude, $N(0, 1)$ is a decent standard prior. Similarly, a half-normal is a reasonable prior for the heterogeneity parameter τ. Moss and De Bin (2020) employed these priors on several examples.

Appendix

The following sandwich convergence theorem is used in the proof of the Gleser–Hwang theorem.

Lemma 9 (Billingsley (1995, Exercise 16.4(a))). *Suppose a_n, b_n, f_n converge to a, b, f and $a_n \leq f_n \leq b_n$ for all n. If $\int a_n d\mu \to \int a d\mu$ and $\int b_n d\mu \to \int b d\mu$, then $\int f_n d\mu \to \int f d\mu$ for any measure μ.*

The proof of Theorem 4 closely follows the proof of Gleser and Hwang (1987, Theorem 1).

Proof of Theorem 4. We can assume without loss of generality that $\|[p_n]\| \geq n$, as we can choose a sub-sequence if we have to. By definition of the diameter D (3.2) we see that

$$\{D \geq n\} = \{\omega \in \Omega \mid \text{there is a } \pi \text{ such that } \|\pi\| \geq n \text{ and } \omega \in R^c(\pi)\},$$

and if $||[p_n]|| \geq n$, then $R^c([p_n]) \subseteq \{D \geq n\}$. Since we assume that $\|[p_n]\| \geq n$ and

$$1 - \alpha \leq P_n(R^c([p_n])) = \int_{R^c([p_n])} p_n d\mu$$

by definition of a confidence set, we have that

$$0 < 1 - \alpha \leq \int_{R^c([p_n])} p_n d\mu \leq \int_{D \geq n} p_n d\mu \tag{4.1}$$

for all n. Since p_n and $p^\star$ are densities,

$$\lim_{n\to\infty} \int p_n d\mu = 1 = \int p^\star d\mu = \int \lim_{n\to\infty} p_n d\mu.$$

This allows us to use Lemma 9 with $a_n = 0$, $b_n = p_n$, and $f_n = 1_{D \geq n} p_n$ and conclude that

$$\int_{D \geq n} p_n d\mu \to \int_{D=\infty} p^\star d\mu. \tag{4.2}$$

Combining equations (4.1) and (4.2), we get

$$0 < 1 - \alpha \leq \int_{D=\infty} p^\star d\mu.$$

Let $P \in \mathcal{P}$ be arbitrary, p be its density, and consider

$$P(D=\infty) = \int_{D=\infty} p d\mu \geq \int_{D=\infty \cap \operatorname{supp} p^\star} \left(\frac{p}{p^\star}\right) p^\star d\mu.$$

Since $\int_{D=\infty} p^\star d\mu > 0$ and $p/p^\star > 0$ on $\operatorname{supp} p^\star$ (since $\operatorname{supp} p \supseteq \operatorname{supp} p^\star$ by assumption), we see that $\int_{D=\infty \cap \operatorname{supp} p^\star} (p/p^\star) p^\star d\mu > 0$ too. It follows that $P(D=\infty) > 0$, and, since P is arbitrary, D has infinite diameter with positive probability. □

Now we prove Lemma 5.

Proof of Lemma (5). Let $n^2 > -c$, so that $\sigma_n^2 > 0$. Recall the well-known formula for the normal truncated to $[a,b]$,

$$\begin{aligned} f_n(x) &= \frac{1}{\Phi\left(\frac{b-\theta_n}{\sigma_n}\right) - \Phi\left(\frac{a-\theta_n}{\sigma_n}\right)} \phi(x;\theta_n,\sigma_n) 1[a,b](x), \\ &= \frac{\phi(x;-n,(n+c)^{1/2}) 1[a,b](x)}{\Phi[-(a+n)(n+c)^{-1/2}] - \Phi[-(b+n)(n+c)^{-1/2}]}. \end{aligned} \tag{4.3}$$

The normal density part equals

$$\phi(x;-n,(n+c)^{1/2}) = (2\pi)^{-1/2}(n+c)^{-1/2} \exp\left(-\frac{x^2+2nx+n^2}{2(n+c)}\right).$$

When n is large compared to x, the term $x^2/2(n+c)$ is negligible, hence

$$\begin{aligned} \phi(x;-n,(n+c)^{1/2}) &\approx (2\pi)^{-1/2}(n+c)^{-1/2} \exp(-n^2/2(n+c)) \exp(-x), \\ &= (n+c)^{-1/2} \phi(n/(n+c)^{1/2}) \exp(-x). \end{aligned}$$

From Equation 5 of Borjesson and Sundberg (1979), we know that $\Phi(-x) \approx \phi(x)/x$ as x grows. The part of (4.3) involving cumulative normal distributions become

$$\Phi[-(a+n)(n+c)^{-1/2}] \approx \frac{(n+c)^{1/2}}{n+a} \phi[-(a+n)(n+c)^{-1/2}],$$

and using the same reasoning as above, we find that $\phi[-(a+n)(n+c)^{-1/2}] \approx \phi((n+c)^{1/2}) \exp(-a)$ as n increase. Therefore,

$$\Phi[-(a+n)(n+c)^{-1/2}] \to \frac{(n+c)^{1/2}\phi((n+c)^{1/2})\exp(-a)}{a+n}.$$

Since this reasoning applies to b as well, we get that f approaches

$$\begin{aligned} & \frac{(n+c)^{-1/2}\phi(n/(n+c)^{1/2})\exp(-x)}{\Phi[-(a+n)n^{-1/2}] - \Phi[-(b+n)n^{-1/2}]}, \\ \approx\ & \frac{(n+c)^{-1/2}\phi(n/(n+c)^{1/2})\exp(-x)}{(n+c)^{1/2}\phi((n+c)^{1/2})\left[\frac{\exp(-a)}{a+n} - \frac{\exp(-b)}{b+n}\right]}, \\ =\ & \frac{\phi(-n^2/2(n+c))\exp(-x)}{\phi((n+c)^{1/2})\left[\frac{\exp(-a)}{a+n} - \frac{\exp(-b)}{b+n}\right]}, \\ \approx\ & \exp(-x) n^{-1} \left[\frac{\exp(-a)}{a+n} - \frac{\exp(-b)}{b+n}\right]^{-1}, \\ \to\ & \exp(-x)\left[\exp(-a) - \exp(-b)\right]^{-1}. \end{aligned}$$

Here the third line follows from $\phi(n/(n+c)^{1/2})/\phi((n+c)^{1/2}) \to 1$. □

For convenience, we show the formula for the mixture probabilities π_i here. Let $c_j = \Phi^{-1}(1 - \alpha_j)$ and define

$$c = \sum_{k=1}^{K} \rho^k [\Phi(c_{j-1}; \theta_0, (\tau^2 + \sigma_i^2)^{1/2}) - \Phi(c_j; \theta_0, (\tau^2 + \sigma_i^2)^{1/2})].$$

Then $\pi^K = \rho^k c^{-1}$ and

$$\pi^k = c^{-1}(\rho^k - \rho^k)[\Phi(c_{k-1}; \theta_0, (\tau^2 + \sigma_i^2)^{1/2}) - \Phi(c_k; \theta_0, (\tau^2 + \sigma_i^2)^{1/2})]$$

References

Bahadur RR, Savage LJ (1956). "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems." *Annals of Mathematical Statistics*, **27**(4), 1115–1122. ISSN 0003-4851. doi:10.1214/aoms/1177728077.

Berger JO, Liseo B, Wolpert RL (1999). "Integrated Likelihood Methods for Eliminating Nuisance Parameters." *Statistical Science*, **14**(1), 1–28. ISSN 0883-4237, 2168-8745. doi:10.1214/ss/1009211804.

Billingsley P (1995). *Probability and Measure*. John Wiley & Sons.

Borjesson P, Sundberg C (1979). "Simple Approximations of the Error Function Q(x) for Communications Applications." *IEEE Transactions on Communications*, **27**(3), 639–643. ISSN 1558-0857. doi:10.1109/TCOM.1979.1094433.

Carter EC, Schönbrodt FD, Gervais WM, Hilgard J (2019). "Correcting for Bias in Psychology: A Comparison of Meta-Analytic Methods." *Advances in Methods and Practices in Psychological Science*, **2**(2), 115–144. ISSN 2515-2459. doi:10.1177/2515245919847196.

Casella G, Berger RL (2002). *Statistical Inference*. Duxbury Press.

Citkowicz M, Vevea JL (2017). "A Parsimonious Weight Function for Modeling Publication Bias." *Psychological Methods*, **22**(1), 28–41. ISSN 1082-989X, 1939-1463. doi:10.1037/met0000119.

Coburn KM, Vevea JL, Coburn (2019). *R package 'weightr'*. URL https://cran.rstudio.org/web/packages/weightr/weightr.pdf.

Cuddy AJC, Schultz SJ, Fosse NE (2018). "P-Curving a More Comprehensive Body of Research on Postural Feedback Reveals Clear Evidential Value for Power-Posing Effects: Reply to Simmons and Simonsohn (2017)." *Psychological Science*, **29**(4), 656–666. ISSN 0956-7976, 1467-9280. doi:10.1177/0956797617746749.

Dear KBG, Begg CB (1992). "An Approach for Assessing Publication Bias Prior to Performing a Meta-analysis." *Statistical Science*, **7**(2), 237–245. ISSN 0883-4237. doi:/10.1214/ss/1177011363.

Donoho DL (1988). "One-Sided Inference about Functionals of a Density." *Annals of Statistics*, **16**(4), 1390–1420. doi:10.1214/aos/1176351045.

Duval S, Tweedie R (2000). "A Nonparametric "Trim and Fill" Method of Accounting for Publication Bias in Meta-analysis." *Journal of the American Statistical Association*, **95**(449), 89–98. ISSN 0162-1459. doi:10.1080/01621459.2000.10473905.

Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR (1991). "Publication Bias in Clinical Research." *The Lancet*, **337**(8746), 867–872. ISSN 0140-6736. doi:10.1016/0140-6736(91)90201-y.

Egger M, Davey Smith G, Schneider M, Minder C (1997). "Bias in Meta-analysis Detected by a Simple, Graphical Test." *British Medical Journal*, **315**(7109), 629–634. ISSN 0959-8138. doi:10.1136/bmj.315.7109.629.

Egger M, Smith GD (1998). "Bias in Location and Selection of Studies." *British Medical Journal*, **316**(7124), 61–66. ISSN 0959-8138. doi:10.1136/bmj.316.7124.61.

Fieller EC (1940). "The Biological Standardization of Insulin." *Supplement to the Journal of the Royal Statistical Society*, **7**(1), 1–64. ISSN 1466-6162. doi:10.2307/2983630.

Flury BD (1990). "Acceptance–Rejection Sampling Made Easy." *SIAM Review*, **32**(3), 474–476. ISSN 0036-1445. doi:10.1137/1032082.

Gleser LJ, Hwang JT (1987). "The Nonexistence of 100(1 - α)% Confidence Sets of Finite Expected Diameter in Errors-in-Variables and Related Models." *Annals of Statistics*, **15**(4), 1351–1362. ISSN 0090-5364. doi:10.1214/aos/1176350597.

Hedges LV (1984). "Estimation of Effect Size under Nonrandom Sampling: The Effects of Censoring Studies Yielding Statistically Insignificant Mean Differences." *Journal of Educational and Behavioral Statistics*, **9**(1), 61–85. ISSN 1076-9986, 0362-9791. doi:10.3102/10769986009001061.

Hedges LV (1992). "Modeling Publication Selection Effects in Meta-analysis." *Statistical Science*, **7**(2), 246–255. ISSN 0883-4237. doi:10.1214/ss/1177011364.

Hedges LV, Vevea JL (1998). "Fixed- and Random-Effects Models in Meta-analysis." *Psychological Methods*, **3**(4), 486. ISSN 1082-989X. doi:10.1037/1082-989x.3.4.486.

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (2019). *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons.

Iyengar S, Greenhouse JB (1988). "Selection Models and the File Drawer Problem." *Statistical Science*, **3**(1), 109–117. ISSN 0883-4237. doi:10.1214/ss/1177013012.

Koschat MA (1987). "A Characterization of the Fieller Solution." *Annals of Statistics*, **15**(1), 462–468. ISSN 0090-5364, 2168-8966. doi:10.1214/aos/1176350282.

Lehmann EL, Romano JP (2005). *Testing Statistical Hypotheses*. Springer Science & Business Media. doi:10.1007/0-387-27605-x.

McShane BB, Böckenholt U, Hansen KT (2016). "Adjusting for Publication Bias in Meta-analysis: An Evaluation of Selection Methods and Some Cautionary Notes." *Perspectives on Psychological Science*, **11**(5), 730–749. ISSN 1745-6916, 1745-6924. doi:10.1177/1745691616662243.

Moss J, De Bin R (2020). "Modelling Publication Bias and p-hacking." *arXiv preprints*. URL https://arxiv.org/abs/1911.12445.

Motyl M, Demos AP, Carsel TS, Hanson BE, Melton ZJ, Mueller AB, Prims JP, Sun J, Washburn AN, Wong KM, Yantis C, Skitka LJ (2017). "The State of Social and Personality Science: Rotten to the Core, Not So Bad, Getting Better, or Getting Worse?" *Journal of Personality and Social Psychology*, **113**(1), 34–58. ISSN 0022-3514, 1939-1315. doi:10.1037/pspa0000084.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Romano JP (2004). "On Non-parametric Testing, the Uniform Behaviour of the t-test, and Related Problems." *Scandinavian Journal of Statistics*, **31**(4), 567–584. ISSN 0303-6898. doi:10.1111/j.1467-9469.2004.00407.x.

Rosenthal R (1979). "The File Drawer Problem and Tolerance for Null Results." *Psychological Bulletin*, **86**(3), 638–641. ISSN 0033-2909, 1939-1455. doi:10.1037/0033-2909.86.3.638.

Simmons JP, Nelson LD, Simonsohn U (2011). "False-positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science*, **22**(11), 1359–1366. ISSN 0956-7976, 1467-9280. doi:10.1177/0956797611417632.

Simonsohn U, Nelson LD, Simmons JP (2014). "p-Curve and Effect Size: Correcting for Publication Bias Using Only Significant Results." *Perspectives on Psychological Science*, **9**(6), 666–681. ISSN 1745-6916, 1745-6924. doi:10.1177/1745691614553988.

Stanley TD (2005). "Beyond Publication Bias." *Journal of Economic Surveys*, **19**(3), 309–345. ISSN 0950-0804. doi:10.1111/j.0950-0804.2005.00250.x.

Stanley TD, Doucouliagos H (2014). "Meta-Regression Approximations to Reduce Publication Selection Bias." *Research Synthesis Methods*, **5**(1), 60–78. doi:10.1002/jrsm.1095.

Sterling TD (1959). "Publication Decisions and their Possible Effects on Inferences Drawn from Tests of Significance—or Vice Versa." *Journal of the American Statistical Association*, **54**(285), 30–34. ISSN 0162-1459. doi:10.1080/01621459.1959.10501497.

van Aert RCM, Wicherts JM, van Assen MALM (2016). "Conducting Meta-Analyses Based on p Values: Reservations and Recommendations for Applying p-Uniform and p-Curve." *Perspectives on Psychological Science*, **11**(5), 713–729. ISSN 1745-6916, 1745-6924. doi:10.1177/1745691616650874.

van Assen MALM, van Aert RCM, Wicherts JM (2015). "Meta-analysis Using Effect Size Distributions of Only Statistically Significant Studies." *Psychological Methods*, **20**(3), 293–309. ISSN 1082-989X, 1939-1463. doi:10.1037/met0000025.

Vevea JL, Hedges LV (1995). "A general linear model for estimating effect size in the presence of publication bias." *Psychometrika*, **60**(3), 419–435. ISSN 0033-3123. doi:10.1007/bf02294384.

Vevea JL, Woods CM (2005). "Publication Bias in Research Synthesis: Sensitivity Analysis Using A Priori Weight Functions." *Psychological Methods*, **10**(4), 428–443. ISSN 1082-989X. doi:10.1037/1082-989X.10.4.428.

von Neumann J (1951). "Various Techniques Used in Connection with Random Digits." *Applied Math Series*, **12**(36-38), 5.

Affiliation:

Jonas Moss
Department of Mathematics
University of Oslo
PB 1053, Blindern, NO-0316, Oslo, Norway
E-mail: jonasmgj@math.uio.no

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume VV *Submitted:* yyyy-mm-dd
MMMMMM YYYY *Accepted:* yyyy-mm-dd