Modified Beta-Function Selection Model

Selection models comprises two components. The first component, herein after termed the *evidence-generating process*, models the distribution of effect sizes before selection, typically using a conventional random-effects model or meta-regression model. The second component, hereinafter termed the *selection process*, identifies how the distribution is changed based on the likelihood of an effect size being reported. The combined model provides parameter estimates that define the selection process, along with meta-analytic estimates that are adjusted for selective reporting.

Following the approach outlined in our previous paper (Pustejovsky, Citkowicz, and Joshi 2025), we model the *marginal* distribution of effect size estimates rather than the joint distribution within studies. To account for dependence among effect sizes, we use cluster-robust variance estimation or clustered bootstrap methods, which accommodate within-study correlation without requiring explicit modeling of the dependence structure. While this strategy limits interpretation to the marginal distribution and does not distinguish between study-level and outcome-level selection, it remains a practical and plausible framework for modeling selective reporting based on the significance of individual estimates.

We use the following notation to describe the model and estimation procedures. Consider a meta-analytic dataset comprising J studies, where study j reports k_j effect size estimates. Let y_{ij} denote the ith effect size estimate from study j, with associated standard error σ_{ij} and one-sided p-value p_{ij}. The one-sided p-value is defined relative to the null hypothesis that the true effect is less than or equal to zero. Let $\mathbf{x}_{ij}$ be a $1 \times x$ row vector of predictors representing characteristics of the effect size, sample, or study procedures. We use $\Phi()$ to denote the standard normal cumulative distribution function and $\phi()$ to denote the standard normal density function.

Evidence-generating process

We assume an evidence-generating process based on a standard random-effects meta-regression model. Let Y^* denote a potentially reported effect size estimate, with standard error σ^*, one-sided p-value p^*, and predictor vector $\mathbf{x}^*$. Then the evidence-generating process is defined as

$$(Y^* | \sigma^*, \mathbf{x}^*) \sim N\left(\mathbf{x}^* \boldsymbol{\beta},\ \tau^2 + \sigma^{*2}\right), \tag{1}$$

where $\boldsymbol{\beta}$ is an $x \times 1$ vector of regression coefficients and τ^2 is the marginal variance of the effect size distribution. This model treats effect sizes as independent and characterizes *total* heterogeneity without decomposing within- and between-study variation.

Selection process

A p-value selection process is defined by a selection function that specifies the probability that an effect size is reported, conditional on its p-value. Let O indicate whether Y^* is observed. The process implies that

$$\Pr\left(O = 1 | p^*\right) \propto w\left(p^*; \boldsymbol{\lambda}\right) \tag{2}$$

where $w\left(.; \boldsymbol{\lambda}\right)$ is a known, strictly positive function on the interval $[0, 1]$ with an unknown $h \times 1$ parameter vector $\boldsymbol{\lambda}$.

Citkowicz and Vevea (2017) defined the selection function using a truncated beta density with two parameters, offering flexibility to capture diverse selection patterns more parsimoniously than the step functions developed by Hedges (1992) and Vevea and Hedges (1995). Since the beta density can be unbounded near 0 and 1, they proposed truncating it to make the model computationally tractable, assuming constant selection probabilities for p-values in the range $[0, \alpha_1]$ and $[\alpha_2, 1]$. Given these pre-specified thresholds α_1 and α_2 and selection parameters $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$, the beta density selection function is expressed by

$$w(p_i^*, \boldsymbol{\lambda}) = \begin{cases} \alpha_1^{\lambda_1 - 1}(1 - \alpha_1)^{\lambda_2 - 1} & \text{if} \quad p_i^* \leq \alpha_1 \\ (p_i^*)^{\lambda_1 - 1}\,(1 - p_i^*)^{\lambda_2 - 1} & \text{if} \quad \alpha_1 < p_i^* < \alpha_2 \\ \alpha_2^{\lambda_1 - 1}(1 - \alpha_2)^{\lambda_2 - 1} & \text{if} \quad \alpha_2 \leq p_i^*. \end{cases} \tag{3}$$

Equation (3) can be written equivalently as

$$w(Y_i^*/\sigma_i^*, \boldsymbol{\lambda}) = \begin{cases} \alpha_1^{\lambda_1 - 1}(1 - \alpha_1)^{\lambda_2 - 1} & \text{if} \quad \sigma_i^*\Phi^{-1}(1 - \alpha_1) \leq Y_i^* \\ \left[\Phi\left(-Y_i^*/\sigma_i^*\right)\right]^{\lambda_1 - 1}\left[\Phi\left(Y_i^*/\sigma_i^*\right)\right]^{\lambda_2 - 1} & \text{if} \quad \sigma_i^*\Phi^{-1}(1 - \alpha_2) < Y_i^* < \sigma_i^*\Phi^{-1}(1 - \alpha_1) \\ \alpha_2^{\lambda_1 - 1}(1 - \alpha_2)^{\lambda_2 - 1} & \text{if} \quad Y_i^* \leq \sigma_i^*\Phi^{-1}(1 - \alpha_2). \end{cases} \tag{4}$$

Citkowicz and Vevea (2017) used extreme truncation points ($\alpha_1 = 10^{-5}$, $\alpha_2 = 1 - 10^{-5}$), but such choices can make the model overly sensitive to rare, extreme p-values, potentially producing implausible estimates (Hedges 2017). Using more moderate, psychologically salient thresholds such as $\alpha_1 = .025$ and $\alpha_2 = .975$ could potentially reduce this sensitivity and yield more plausible selection patterns.

Citkowicz, Martyna, and Jack L Vevea. 2017. "A parsimonious weight function for modeling publication bias." *Psychological Methods* 22 (1): 28–41. https://doi.org/10.1037/met0000119.

Hedges, Larry V. 1992. "Modeling Publication Selection Effects in Meta-Analysis." *Statistical Science* 7 (2): 246–55.

———. 2017. "Plausibility and Influence in Selection Models: A Comment on Citkowicz and Vevea (2017)." *Psychological Methods* 22 (1): 42–46. https://doi.org/10.1037/met0000108.

Pustejovsky, James E., Martyna Citkowicz, and Megha Joshi. 2025. "Estimation and Inference for Step-Function Selection Models in Meta-Analysis with Dependent Effects." *Journal Name*.

Vevea, Jack L, and Larry V Hedges. 1995. "A General Linear Model for Estimating Effect Size in the Presence of Publication Bias." *Psychometrika* 60 (3): 419–35. https://doi.org/10.1007/BF02294384.