

Empirical Bayes Meta-Analysis Through G-Modeling

Johannes Vilsmeier

The usual approach to empirical Bayes modeling, referred to as *f*-modeling, builds on the assumption of a prior distribution with known shape up to a parameter γ . Inference is carried out by setting γ equal to an estimate obtained from the marginal distribution of observations y_1, \dots, y_n . A less restrictive modeling approach, referred to as *g*-modeling, relaxes this assumption by requiring solely that the prior distribution belongs to the exponential family of distributions indexed by a parameter α . So far, applications of *g*-modeling comprise parallel estimation problems in relatively large data sets. Meta-Analyses also give rise to parallel estimation problems. Moreover, *f*-modeling approaches to meta-analysis are well established. Here, a *g*-modeling based approach for conducting a meta-analysis is developed and its performance is compared to both maximum- and restricted maximum-likelihood based random-effects meta-analysis in simulations. Results indicate that the *g*-modeling meta-analysis compares reasonably well. Zero-variance estimates of between-study effects are completely avoided by construction, and coverage probabilities of Wald-type confidence intervals for the average meta-analytic effect were consistently better than those obtained in the maximum- and restricted maximum-likelihood meta-analytic frameworks.

1 Introduction

Consider the following *Bayes model* (Lehmann and Casella 1998, 240) for $n \in \mathbb{N}$ random variables, $y_1, \dots, y_n \in \mathbb{R}$:

$$\begin{aligned} \text{I. } & y_i | \theta_i \stackrel{\text{ind.}}{\sim} p_i(y_i | \theta_i) \\ \text{II. } & \theta_i | \gamma \stackrel{\text{i.i.d.}}{\sim} g(\theta_i | \gamma), \quad i = 1, \dots, n. \end{aligned} \tag{1}$$

Each of the y_i 's is regarded as an independent draw from a known family of probability distributions with conditional density $p_i(y_i | \theta_i)$ indexed by a parameter $\theta_i \in \Theta \subseteq \mathbb{R}$. The

parameter space Θ is itself equipped with a probability distribution (i.e., the *prior distribution*) with density $g(\theta|\gamma)$, indexed by a parameter γ . The θ_i 's ($i = 1, \dots, n$) are independently and identically distributed according to $g(\theta|\gamma)$. Collecting the y_i 's in a vector $y = (y_1, \dots, y_n)$ and the θ_i 's in a vector $\theta = (\theta_1, \dots, \theta_n)$, the joint conditional density of y and the joint density of θ will be denoted by $p(y|\theta) = \prod_{i=1}^n p_i(y_i|\theta_i)$ and $g(\theta|\gamma) = \prod_{i=1}^n g(\theta_i|\gamma)$, respectively.

The primary goal is to infer about θ - the parameter (a vector) of interest. This is done through the posterior distribution,

$$g(\theta|y, \gamma) = \frac{p(y|\theta)g(\theta|\gamma)}{\int_{\Theta^n} p(y|\theta)g(\theta|\gamma)d\theta} = \frac{p(y|\theta)g(\theta|\gamma)}{f(y|\gamma)}. \quad (2)$$

It is straightforward to apply (2) once the Bayes model (1) has been fully specified, including the parameter γ of the prior distribution (or more generally, the prior distribution g). There are several strategies for specifying the prior distribution through a parameter γ in (1). Two of these are referred to as *single-prior Bayes* and *empirical Bayes* in Lehmann and Casella (1998, Ch. 5).

In *single-prior Bayes*, γ (and hence, the entire prior distribution g) is chosen by the analyst. For instance, if $y_1, \dots, y_n | \mu \stackrel{\text{i.i.d.}}{\sim} \text{Normal}(\mu, \sigma^2)$, $\sigma^2 \in (0, \infty)$ known, one might choose a conjugate $\text{Normal}(\mu_0, \sigma_0^2)$ prior for μ . Here, γ is a vector (μ_0, σ_0^2) and it is set to a particular value in $\mathbb{R} \times (0, \infty)$ defined by the analyst *a priori*.

In *empirical Bayes*, γ is replaced by an estimate $\hat{\gamma}$ obtained from the marginal distribution $f(y|\gamma)$. The maximum likelihood (ML) estimate of γ

$$\hat{\gamma}_{\text{ML}} = \arg \max_{\gamma} f(y|\gamma) = \arg \max_{\gamma} \int_{\Theta^n} p(y|\theta)g(\theta|\gamma)d\theta$$

may be used, however $\hat{\gamma}$ does not have to be ML-based.

The posterior of θ is then approximated by a plug-in version of (2):

$$g(\theta|y, \hat{\gamma}) = \frac{p(y|\theta)g(\theta|\hat{\gamma})}{f(y|\hat{\gamma})}.$$

Within empirical Bayes, two different strategies can be distinguished: *f-modeling* and *g-modeling* (Efron 2014). The strategy described in the previous paragraph describes *f-modeling*; the prior density g is assumed to belong to a known parametric family of distributions indexed by a parameter γ , while $\hat{\gamma}$ is estimated from the marginal distribution f (hence the name *f-modeling*).

The second empirical Bayes strategy, *g-modeling*, is concerned with estimating the entire prior distribution through its density g from the data (hence the name *g-modeling*), making no

parametric assumption in stage II. of (1). That is, stage II of the Bayes model is simply $\theta_i \stackrel{\text{i.i.d}}{\sim} g$, where g denotes the density of a prior distribution for the θ_i 's.

Letting \hat{g} denote the estimate of the prior density, the posterior is then given by

$$\hat{g}(\theta|y) = \frac{p(\theta|y)\hat{g}(\theta)}{f(y)} = \frac{p(\theta|y)\hat{g}(\theta)}{\int_{\Theta^n} p(\theta|y)\hat{g}(\theta)d\theta}.$$

Unfortunately, this non-parametric ideal version of g -modeling is practically infeasible as convergence rates of \hat{g} to g are poor (Carroll and Hall 1988; Narasimhan and Efron 2020). The g -modeling approach considered here and in e.g., Chapter 21 of Efron and Hastie (2016) may be described as a compromise between a fully non-parametric strategy and f -modeling: The prior g is assumed to belong to the exponential family of distributions indexed by a parameter α (further details are given in Section 2.1 below). This restriction on g allows for efficient estimation as shown by Efron (2016).

Examples of applications of the exponential-family-based parametric version of g -modeling (from here on simply referred to as g -modeling) to real-world data can be found in Efron (2014), Efron (2016), Efron and Hastie (Efron and Hastie 2016, Ch. 21) and Narasimhan and Efron (2020). They encompass applications of g -modeling to problems such as estimating the number of words in William Shakespeare's vocabulary, the number of yet to be observed species of butterflies, the proportion of cancerous lymph nodes, or the mean differences of thousands of genes between two groups [?]. These examples all share a defining feature: they encompass relatively large data sets with multiple parallel estimation problems posed at once. For instance, estimating the proportion of cancerous nodes for each individual in a sample of size n amounts to n parallel estimation problems.

Parallel estimation problems also occur whenever a researcher conducts a meta-analysis (although meta-analyses are usually not framed in that way). In a meta-analysis, individual estimates of an effect of interest obtained over the course of multiple studies are combined in order to achieve an overall 'better' estimate of said effect. In addition, the combined information in the collection of individual study estimates can also be used to obtain 'better' estimates for each of the effects underlying the individual studies, which describes a parallel estimation problem. Moreover, empirical Bayes meta-analyses derived from the f -modeling strategy have been around for decades (Raudenbush and Bryk 1985). It thus seems natural to extend the list of potential areas of applications of g -modeling and assess its usefulness in the context of meta-analyses.

The sections below are structured as follows: Section 2.1 and Section 2.2 provide condensed accounts of the technical details underlying g -modeling and classical meta-analytic methods employing maximum-likelihood (ML) and restricted-maximum-likelihood (REML) estimation, respectively. In Section 2.4 an approach to meta-analysis based on g -modeling is developed. In Section 3 the g -modeling meta-analysis is compared to ML- and REML-based meta-analysis in a simulation study.

2 Statistical Framework

2.1 G-Modeling theory

This section briefly summarizes a number of results required to be able to apply g -modeling to meta-analysis below. More comprehensive accounts can be found in Efron Efron (2016), Efron (2014), Efron and Hastie (2016, Ch. 21), and Narasimhan and Efron (2020).

Suppose we have n independent observations y_1, \dots, y_n generated by the Bayes model (1). Marginally, this implies

$$y_i \sim f_i(y_i) = \int_{\Theta} p_i(y_i|\theta)g(\theta)d\theta, \quad i = 1, \dots, n. \quad (3)$$

The researcher observes $y = (y_1, \dots, y_n)$ from the joint marginal density $f = \prod_{i=1}^n f_i$. The task is to estimate the prior density g from the sample y , which may be described as *deconvolving* f , hence, the estimate \hat{g} of g is also referred to as a *deconvolution estimate* (e.g., Efron 2016).

For convenience, we may assume that g sits on a finite subset of \mathbb{R} ; that is, the parameter space is given by $\Theta = \{t_1, \dots, t_m\}$, $t_j \in \mathbb{R}, j = 1, \dots, m$ (see Efron 2014 for a description of the continuous case). Moreover, g is restricted to the exponential family of distributions [? ?]:

$$g(\theta|\alpha) = \exp \left\{ S(\theta)^\top \alpha - \phi(\alpha) \right\}, \quad (4)$$

where $\mathbf{S} : \Theta \rightarrow \mathbb{R}^p$ denotes the sufficient statistic, $\alpha \in \mathbb{R}^p$ the natural parameter and $\phi(\alpha) = \log \left(\sum_{j=1}^m \exp\{S(t_j)^\top \alpha\} \right)$ a constant, ensuring that $g(\theta|\alpha)$ sums to 1.

With Θ being a finite set, we can evaluate S at every point $t_j \in \Theta$ and assemble the resulting m vectors, each containing p elements, in what is called a *structure matrix* $Q \in \mathbb{R}^{m \times p}$,

$$Q = \begin{bmatrix} S(t_1)^\top \\ \vdots \\ S(t_m)^\top \end{bmatrix} = \begin{bmatrix} S(t_1)_1 & \dots & S(t_1)_p \\ \vdots & \ddots & \vdots \\ S(t_m)_1 & \dots & S(t_m)_p \end{bmatrix}.$$

The structure matrix will prove helpful in the derivation of the log-likelihood, the observed Fisher information, the bias and the variance, below. It is convenient to express the prior distribution in (4) for θ evaluated at some point t_j in terms of the structure matrix Q

$$\begin{aligned} g_j(\alpha) = g(t_j|\alpha) &= \exp \left\{ Q_j^\top \alpha - \phi(\alpha) \right\}, \quad j = 1, \dots, m \\ \phi(\alpha) &= \log \left(\sum_{j=1}^m e^{Q_j^\top(\alpha)} \right), \end{aligned} \quad (5)$$

where Q_j^\top denotes the j th row of Q .

In theory, one is free to choose the function S and hence the structure matrix Q as one sees fit. A common choice for S (e.g., Efron 2014) is

$$S(t_j) = (1, \psi_1(t_j), \dots, \dots \psi_{p-1}(t_j))^\top,$$

where $\psi_k, k = 1, \dots, p-1$ denote basis functions for a natural cubic spline, with $p-1$ degrees of freedom returned by the R function `splines::ns(x = Θ , df = p-1)`. In this example, `splines::ns` is supplied with a vector containing the parameter space Θ as its first argument \mathbf{x} and determines a sequence of $p \in \mathbb{N}$ knots (supplied via the argument `df`), $t_1 = \tau_1 < \tau_2 < \dots < \tau_p = t_m$, from the elements in Θ . The interior knots $(\tau_2, \dots, \tau_{p-1})$ are chosen at uniform quantiles of Θ . R maps each point $t_j \in \Theta$ to a $p-1$ row vector of natural cubic spline basis functions $\psi_1(t_j), \dots, \psi_{p-1}(t_j)$ and collects these row vectors in a matrix, $A \in \mathbb{R}^{m \times (p-1)}$. Adding a column of 1's to A yields the structure matrix Q . Consequently, the scalar $Q_j^\top \alpha$ in the exponent of (5) is the value of a natural cubic spline that approximates the unnormalized log-density of the prior, at the point t_j . These approximations are obtained for every $t_j \in \Theta$. The pairs $(t_1, Q_1^\top \alpha), \dots, (t_m, Q_m^\top \alpha)$ represent an approximation of the unnormalized log-prior density. The unknown $\alpha \in \mathbb{R}^p$ is estimated from the data.

The assumption of a discrete and finite sample space Θ allows for the representation of the prior density g as a vector $g = (g_1(\alpha), \dots, g_m(\alpha))^\top$. The marginal density f_i for the i th observation is thus (by analogy to (3))

$$y_i \sim f_i = f_i(y_i) = \sum_{j=1}^m p_i(y_i | \theta_i = t_j) g_j(\alpha). \quad (6)$$

To further simplify the notation, define

$$p_{ij} := p_i(y_i | \theta_i = t_j) \quad (7)$$

and let $P = (p_{ij})_{1 \leq i \leq n, 1 \leq j \leq m}$ be a $n \times m$ matrix, where the i th row is given by (p_{i1}, \dots, p_{im}) . The vector of marginal densities $f = (f_1, \dots, f_n)^\top$ can thus be represented concisely using matrix multiplication

$$f = Pg. \quad (8)$$

Since the n observations are assumed to be independent, the (joint) log-likelihood is

$$l(\alpha) = \sum_{i=1}^n \log f_i = \sum_{i=1}^n \log P_i^t g(\alpha)$$

where P_i^t denotes the i th row of P .

Let $W_i(\alpha) = (g_1(p_{i1}/f_i - 1), \dots, g_m(p_{im}/f_i - 1))^\top$. The score function and observed Fisher information matrix are given by

$$\dot{l}(\alpha) = Q^\top \sum_{i=1}^n W_i(\alpha) =: Q^\top W_+(\alpha) \quad (9)$$

and

$$\begin{aligned} I(\alpha) &= -\frac{\partial \dot{l}(\alpha)}{\partial \alpha^\top} \\ &= Q^\top \left(\sum_{i=1}^n W_i(\alpha) W_i(\hat{\alpha})^\top + W_+(\alpha) g(\alpha)^\top + g(\alpha) W_+(\alpha)^\top \right. \\ &\quad \left. - \text{diag}\{W_+(\alpha)\} \right) Q, \end{aligned} \quad (10)$$

respectively (see Efron (2016) for a proof).

In estimating α , regularization is needed (Efron 2016). To this end, the *penalized log-likelihood* with penalty term $s(\alpha)$

$$l_{\text{pen.}}(\alpha) = l(\alpha) - s(\alpha)$$

is maximized rather than the log-likelihood function, $l(\alpha)$. Efron (2016) and Efron and Hastie (2016) use the following penalty:

$$s(\alpha) = c_0 \|\alpha\| = c_0 \left(\sum_{l=1}^p \alpha_l^2 \right)^{1/2}, \quad c_0 > 0, \quad (11)$$

where c_0 is a user-defined penalty factor.

The penalty introduces bias (but decreases variance) by drawing $\hat{\alpha}$ towards the origin and thus $\hat{g} = g(\hat{\alpha})$ towards a flat prior as c_0 increases.

The gradient of $s(\alpha)$ is given by

$$\dot{s}(\alpha) = \frac{\partial s(\alpha)}{\partial \alpha} = c_0 \frac{1}{2} \frac{1}{\|\alpha\|} 2\alpha = \frac{c_0}{\|\alpha\|} \alpha. \quad (12)$$

The Hessian matrix is

$$\ddot{s}(\alpha) = \left(\frac{\partial^2 s}{\partial \alpha_i \partial \alpha_j} \right) = \frac{c_0}{\|\alpha\|} \left(I - \frac{\alpha \alpha^t}{\|\alpha\|^2} \right). \quad (13)$$

The penalized log-likelihood $l_{\text{pen.}}$ and its gradient $\dot{l}_{\text{pen.}}(\alpha) = \dot{l}(\alpha) - \dot{s}(\alpha)$ are needed to find the maximizer $\hat{\alpha}$ of $l_{\text{pen.}}$ numerically in the implementation of g -modeling in the simulation study below Section 3.

The bias vector of the maximizer $\hat{\alpha}$ of $l_{\text{pen.}}$ can be estimated using (Efron 2016)

$$\text{Bias}(\hat{\alpha}) = -\{I(\alpha) + \ddot{s}(\hat{\alpha})\}^{-1} \dot{s}(\hat{\alpha}). \quad (14)$$

Plugging $\hat{\alpha}$ into (5) yields the estimate $\hat{g} = (g_1(\hat{\alpha}), \dots, g_m(\hat{\alpha})) = (\hat{g}_1, \dots, \hat{g}_m)$ for the prior g . An estimate of the bias of \hat{g} is given by (see Efron 2016; Efron and Hastie 2016, 443)

$$\begin{aligned} \text{Bias}(\hat{g}) &= D(\hat{\alpha}) Q \text{Bias}(\hat{\alpha}), \\ \text{where } D(\hat{\alpha}) &= \text{diag}\{g(\hat{\alpha})\} - g(\hat{\alpha})g(\hat{\alpha})^\top. \end{aligned} \quad (15)$$

Equations (14) and (15) provide the reason for presenting the observed Fisher information matrix (10), the gradient and the Hessian of the penalty term $s(\alpha)$ in this section: these quantities can be used to estimate, and thus mitigate, the bias incurred by penalizing the log-likelihood.

2.2 Meta-Analysis

A meta-analysis comprises a number $k \in \mathbb{N}$ of different studies, each on approximately the same substantive research question of interest. For instance, there may be k studies on the efficacy of a new drug. Let θ_i denote the true mean difference in the outcome variable (e.g., blood pressure) between a control and a placebo group in study $i = 1, \dots, k$. Each study i gives rise to an estimate $\hat{\theta}_i$ of θ_i , which is assumed normal with mean θ_i and known variance σ_i^2 . If - as is usually the case - the true σ_i^2 's are not known, the sample estimates, $\hat{\sigma}_i^2$'s, are used instead. This setting describes the so-called *normal random-effects model* of meta-analysis (e.g., Higgins, Thompson, and Spiegelhalter 2009), often expressed in the following hierarchical form:

$$\begin{aligned} \text{I. } \hat{\theta}_i | \theta_i &\stackrel{\text{ind}}{\sim} \text{Normal}(\theta_i, \sigma_i^2) \\ \text{II. } \theta_i | \mu, \tau^2 &\stackrel{\text{i.i.d.}}{\sim} \text{Normal}(\mu, \tau^2), \quad i = 1, \dots, k. \end{aligned} \quad (16)$$

Clearly, (16) is a *Bayes model* (1) specifying a normal prior on the true study effects.

In many applications, the primary goal of a meta-analysis is to estimate μ , the so-called *average (treatment) effect*, with τ^2 and the θ_i 's being treated as nuisance parameters. The variability in the true study effects θ_i implied by the variance parameter τ^2 in stage I. of (16) is also referred to as *effect-size heterogeneity* or simply *heterogeneity*.

To derive estimators for μ, τ^2 , and $\theta_1, \dots, \theta_k$ from the collection of k studies, note that the normal random effects model (16) is just a special case of a linear-mixed model:

$$y = X\beta + Z\eta + \varepsilon, \quad (17)$$

where y is a k -vector of random response variables, X a $k \times p$ matrix of known regressors, β a p -vector of fixed effects, Z a $k \times q$ design matrix for the random effects contained in the q -vector η and ε a k -vector of random error terms. Furthermore, it is assumed that $\mathbb{E}[\eta] = \mathbf{0}_q$, $\mathbb{E}[\varepsilon] = \mathbf{0}_k$, and that the variance-covariance matrix is given by

$$\text{Cov}[(\eta, \varepsilon)^\top] = \begin{bmatrix} G & \mathbf{0}_{q \times k} \\ \mathbf{0}_{k \times q} & R \end{bmatrix},$$

where $G = \text{Cov}[\eta]$ and $R = \text{Cov}[\varepsilon]$ denote the variance-covariance matrices of η and ε , respectively. It then follows that $\mathbb{E}[y] = X\beta$ and that $\text{Cov}[y] = ZGZ^\top + R$. Assuming normality for η and ε , we have

$$y \sim \text{Normal}_k(\mathbf{X}\beta, ZGZ^\top + R).$$

Building on these assumption, the linear-mixed model representation of the normal random-effects meta-analytic model (16) is obtained by setting

$$\begin{aligned} y &= \hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k), & \mathbf{X} &= \mathbf{1}_k = (1, \dots, 1)^\top \in \mathbb{R}^k, & \beta &= \mu \in \mathbb{R}, \\ \mathbf{Z} &= \mathbf{I}_k, & \eta &= (\theta_1 - \mu, \dots, \theta_k - \mu)^\top, & \mathbf{G} &= \text{diag}\{\tau^2, \dots, \tau^2\}, \\ R &= \text{diag}\{\sigma_1^2, \dots, \sigma_k^2\}, \end{aligned}$$

which yields

$$\hat{\theta} = \mathbf{1}_k \mu + \mathbf{I}_k \eta + \varepsilon. \quad (18)$$

As an immediate consequence, the marginal distributions of the $\hat{\theta}_i$'s are normal:

$$\hat{\theta}_i \stackrel{\text{ind.}}{\sim} \text{Normal}(\mu, \sigma_i^2 + \tau^2), \quad i = 1, \dots, k.$$

The joint log-likelihood of $\hat{\theta}_1, \dots, \hat{\theta}_n$ is

$$l(\mu, \tau^2) \propto -\frac{1}{2} \sum_{i=1}^k \log(\sigma_i^2 + \tau^2) - \frac{1}{2} \sum_{i=1}^k \frac{(\hat{\theta}_i - \mu)^2}{\sigma_i^2 + \tau^2}. \quad (19)$$

Maximization of (19) with respect to the parameter vector (μ, τ^2) can be achieved by profiling. For τ^2 fixed, the maximizer of μ is obtained by setting the derivative of (19) with respect to μ equal to zero and solving for μ . This gives the estimator

$$\hat{\mu} = \hat{\mu}(\tau^2) = \frac{\sum_{i=1}^k w_i \hat{\theta}_i}{\sum_{i=1}^k w_i}, \quad (20)$$

for $w_i = (\sigma_i^2 + \tau^2)^{-1}$.

Replacing μ in (19) by the function $\hat{\mu}(\tau^2)$ (20) (τ^2 is no longer assumed fixed) yields the profile log-likelihood for τ^2 , denoted by l_p .

The maximum likelihood estimate $\hat{\tau}_{\text{ML}}^2$ for τ^2 satisfies

$$l'_p(\tau_{\text{ML}}^2) = \left. \frac{\partial l_p(\tau^2)}{\partial \tau^2} \right|_{\tau^2 = \hat{\tau}_{\text{ML}}^2} = -\frac{1}{2} \sum_{i=1}^k w_i + \frac{1}{2} \sum_{i=1}^k \frac{2(\hat{\theta}_i - \hat{\mu}(\tau^2))\hat{\mu}'(\tau^2)w_i^{-1} - (\hat{\theta}_i - \hat{\mu})^2}{w_i^{-2}} \stackrel{!}{=} 0, \quad (21)$$

where $\hat{\mu}'(\tau^2) = -\frac{\sum_{i=1}^k w_i \hat{\theta}_i}{\sum_{i=1}^k w_i} + \frac{\sum_{i=1}^k w_i^2 \cdot \sum_{i=1}^k w_i \hat{\theta}_i}{(\sum_{i=1}^k w_i)^2}$

The solution $\hat{\tau}_{\text{ML}}^2$ to (21) is found numerically. The maximum likelihood estimate $\hat{\mu}_{\text{ML}}$ for μ is then given by plugging $\hat{\tau}_{\text{ML}}^2$ into (20),

$$\hat{\mu}_{\text{ML}} = \hat{\mu}_{\text{ML}}(\hat{\tau}_{\text{ML}}^2).$$

It may happen that solutions of τ_{ML}^2 in 21 converge to values outside the parameter space for τ^2 , namely $\tau^2 < 0$ (i.e., we would estimate the variance to be negative; see Figure 1 for an illustration).

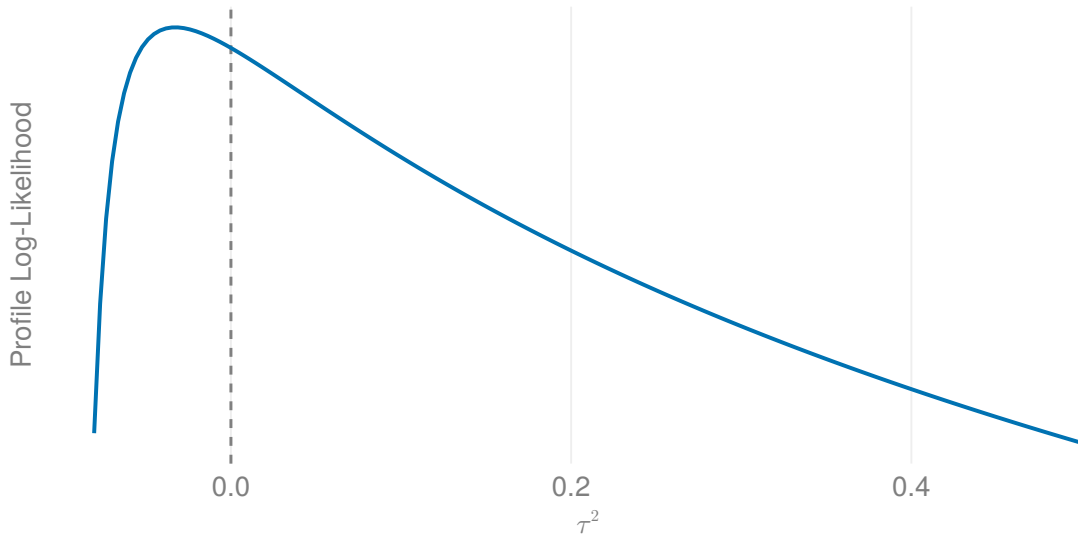


Figure 1: Example of a zero-variance estimate. The profile log-likelihood for the variance parameter τ^2 decreases smoothly and attains its maximum outside of the parameter space at the nonsensical value of approximately -0.02 . In such cases, the maximum-likelihood estimate is usually set to $\hat{\tau}_{\text{ML}}^2 = 0$.

It is common practice to set the maximum likelihood estimate of τ^2 equal to 0 when the maximizer of the profile log-likelihood is located outside of the parameter space. Thus, a truncated version the maximum likelihood estimate for τ^2 given by

$$\hat{\tau}_{\text{ML}}^2 = \max\{\hat{\tau}_{\text{ML}}^2, 0\}. \quad (22)$$

is actually used. These so-called *zero-variance estimates* (i.e., $\hat{\tau}_{\text{ML}}^2 = 0$) occur rather frequently in practice.

It is well known that $\hat{\tau}_{\text{ML}}^2$ has negative bias for τ^2 in finite samples, due to the loss in the degrees of freedom incurred by replacing μ in the profile likelihood (21) with its estimate $\hat{\mu}$ (Viechtbauer 2005; Harville 1977). To correct for this bias, the restricted log-likelihood function $l_R(\mu, \tau^2)$ may be maximized instead of the log-likelihood (19). This is equivalent to maximizing the following function:

$$h(\mu, \tau^2) = l(\mu, \tau^2) - \frac{1}{2} \log \sum_{i=1}^k (\sigma_i^2 + \tau^2)^{-1}.$$

The profile version of $h(\mu, \tau^2)$ is

$$h_p(\tau^2) = l_p(\tau^2) - \frac{1}{2} \log \sum_{i=1}^k (\sigma_i^2 + \tau^2)^{-1},$$

satisfying

$$h'_p(\hat{\tau}_{\text{REML}}^2) = l'_p(\tau_{\text{REML}}^2) - \frac{1}{2} \frac{\partial}{\partial \tau^2} \log \sum_{i=1}^k (\sigma_i^2 + \tau^2)^{-1} = 0, \quad (23)$$

where τ_{REML}^2 denotes the *restricted-maximum-likelihood* (REML) estimate for τ^2 .

Again, the solution $\hat{\tau}_{\text{REML}}^2$ to (23) is found numerically with the constraint that $\hat{\tau}_{\text{REML}}^2$ is truncated to 0 if the solutions converge to values of $\tau^2 < 0$. The resulting REML estimator for μ is given by

$$\hat{\mu}_{\text{REML}} := \hat{\mu}(\hat{\tau}_{\text{REML}}^2). \quad (24)$$

For τ^2 known, the variance of $\hat{\mu}(\tau^2)$ in (20) is given by

$$\text{Var}[\hat{\mu}(\tau^2)] = \frac{1}{\sum_{i=1}^n (\sigma_i^2 + \tau^2)^{-1}}. \quad (25)$$

Commonly employed estimators of the variances of $\hat{\mu}_{\text{ML}}$ and $\hat{\mu}_{\text{REML}}$ are given by

$$\hat{\text{Var}}[\hat{\mu}_{\text{ML}}] = \text{Var}[\hat{\mu}(\tau_{\text{ML}}^2)] \quad \text{and} \quad \hat{\text{Var}}[\hat{\mu}_{\text{REML}}] = \text{Var}[\hat{\mu}(\tau_{\text{REML}}^2)]. \quad (26)$$

The estimator $\hat{\mu}$ in (20) and its variance $\text{Var}[\hat{\mu}]$ are used to construct $(1 - \alpha)\%$ Wald-type confidence intervals (CI) for μ based on a normal approximation of the distribution of $\hat{\mu}$. The general form of a Wald-type CI (for τ^2 fixed) is given by

$$\hat{\mu}(\tau^2) \pm z_{1-\alpha/2} \sqrt{\text{Var}[\hat{\mu}(\tau^2)]}, \quad (27)$$

where $z_{1-\alpha/2}$ denotes the $(1 - \alpha)$ -quantile of the standard normal distribution. In applications, confidence intervals are obtained by plugging τ_{ML}^2 or τ_{REML}^2 into (27), depending on the method used for estimating τ^2 .

2.3 F -Modeling posterior inference

If μ and τ^2 are known, stage II. of the Bayes model (16) could be interpreted as a normal single-prior distribution for each of the θ_i . The sampling distribution of the $\hat{\theta}_i$ is also normal. Thus, by conjugacy, the posterior distribution for θ_i is

$$\begin{aligned} \theta_i | \hat{\theta}_i, \mu, \tau^2 &\sim \text{Normal}(\tilde{\theta}_i, V), \\ \text{for } V &= \left\{ \frac{1}{\sigma_i^2} + \frac{1}{\tau^2} \right\}^{-1} \text{ and } \tilde{\theta}_i = V \left\{ \frac{\hat{\theta}_i}{\sigma_i^2} + \frac{\mu}{\tau^2} \right\}. \end{aligned} \quad (28)$$

The posterior in (28) serves as the starting point for any inferences about the individual θ_i 's, e.g., by way of $(1 - \alpha)\%$ ($\alpha \in (0, 1)$) credible intervals constructed from posterior quantiles. Thus meta-analysis naturally gives rise to k parallel estimation problems (one for each $\theta_i, i = 1, \dots, k$).

In the f -modeling strategy to empirical Bayes, posterior distributions for the θ_i 's are obtained by replacing μ and τ^2 in (28) by estimates $\hat{\mu}$ and $\hat{\tau}^2$. Both the ML- and REML-estimates are suitable choices.

2.4 G -Modeling meta-analysis

The g -modeling equivalent of the Bayes model (16) is given by

$$\begin{aligned} \text{I. } \hat{\theta}_i | \theta_i &\stackrel{\text{ind}}{\sim} \text{Normal}(\theta_i, \sigma_i^2) \\ \text{II. } \theta_i | \alpha &\stackrel{\text{i.i.d.}}{\sim} g(\alpha), \quad i = 1, \dots, k, \end{aligned} \quad (29)$$

for $g(\alpha) = (g_1(\alpha), \dots, g_m(\alpha))$ and (α) an exponential family density as in (5).

The conditional density p_{ij} in (7) which the observations in stage I. of (29) are drawn from now denotes the density of a normal distribution, $\text{Normal}(\theta_i, \sigma_i^2)$.

An estimate \hat{g} of g is obtained as described in Section 2.1. Note that \hat{g} is a (counting) density by construction. The prior expectation and variance of θ_1 (assuming that they exist) under \hat{g} are thus given by

$$\mu_{\hat{g}} = \mathbb{E}[\theta_1] = \sum_{j=1}^m t_j \hat{g}_j \quad \text{and} \quad \tau_{\hat{g}}^2 = \text{Var}[\theta_1] = \sum_{j=1}^m (t_j - \mu_{\hat{g}})^2 \hat{g}_j, \quad (30)$$

and may be regarded as estimators for μ and τ^2 , the average (treatment) effect and its variance (note that since the θ_i are i.i.d. the expectation and variance of every θ_i are equal to those of θ_1).

Obvious defects of $\mu_{\hat{g}}$ and $\tau_{\hat{g}}^2$ as estimators for μ and τ^2 are the apparent lack of analytic expressions for their variances. Theoretically, the bootstrap could be employed as a method for estimating the variances of $\mu_{\hat{g}}$ and $\tau_{\hat{g}}^2$ as well as for constructing confidence intervals. However, evaluating statistical properties such as bias, MSE and coverage probabilities of these bootstrapped estimates turned out to be computationally infeasible (at least to the author of this report) due to the large number of required simulations.

We can nevertheless mimic the construction of a Wald-type CI by plugging $\tau_{\hat{g}}^2$ into the general form given by (27). The coverage probability of the resulting interval estimator can then be assessed in simulations, as is done below.

2.5 G -Modeling posterior inference

The posterior probabilities for each event $\{\theta_i = t_j : i = 1, \dots, k, j = 1, \dots, m\}$ based on the corresponding observation $\hat{\theta}_i$ are obtained via Bayes' rule,

$$\hat{g}_{ij} = \mathbb{P}\{\theta_i = t_j | \hat{\theta}_i\} = \frac{p_{ij}\hat{g}_j}{\sum_{j=1}^k p_{ij}\hat{g}_j}, \quad j = 1, \dots, m. \quad (31)$$

The entire posterior density for an individual θ_i can thus be represented as a vector

$$\hat{g}_i = (\hat{g}_{i1}, \dots, \hat{g}_{im})^\top. \quad (32)$$

Once the posterior \hat{g}_i is obtained, it is straightforward to generate random samples from it and compute posterior quantiles, thus allowing for the construction of credible intervals for each θ_i .

3 Simulation study

3.1 Study goals

The goal of the simulation study is to apply the g -modeling meta-analysis developed in Section 2.4 and compare its performance to the classical approaches based on ML- and REML-estimation across several scenarios.

The study follows Chung, Rabe-Hesketh, and Choi (2013) in several regards. The three meta-analytic estimation strategies are compared with respect to the following properties: The *proportion of zero-variance estimates* (i.e., $\hat{\tau}^2 = 0$), the *MSE (or RMSE)* and the *bias* of $\hat{\tau}_{\text{ML}}^2$, $\hat{\tau}_{\text{REML}}^2$ and $\tau_{\hat{g}}^2$, as well as the coverage probabilities of the 95% Wald-type CI's for μ obtained by plugging these different estimators for τ^2 into (27). No explicit rationale for using

MSE (or RMSE), bias and coverage probabilities seems necessary as these are universally accepted as sensible quantities for evaluating estimators.

However, the proportion of zero-variance estimates as a reasonable property with respect to which estimators for τ^2 can be compared seems to require some further justification. Section 2.2 briefly touched on the issue of zero-variance estimates which frequently occur in both ML- and REML-applications. An estimate of $\hat{\tau}_{\text{ML}}^2 = 0$ appears to be in conflict with the intuitive interpretation of stage II. in (16): by assuming model (16), a researcher expects variability in the true study effects $\theta_1, \dots, \theta_k$. Claiming $\hat{\tau}^2 = 0$ is equivalent to claiming that there is no variability in the study effects and that one and the same true effect generated the observable $\hat{\theta}_i$ in the k studies, each usually conducted at a different time, in a different place, by a different group of researchers, etc. Against this backdrop of variable situational factors, the stance that it is reasonable to model the observations from different studies as having been generated by one and the same true effect appears objectionable on scientific grounds (For completeness, the objectionable assumption is called the fixed- or common-effect assumption in the literature and it would provide for an alternative model to (16)). Therefore, it may be considered desirable that estimators for τ^2 avoid zero-variance estimates for τ^2 . The estimator τ_g^2 achieves this by construction (precluding the case $\hat{g}_j = 0$ for every $j = 1, \dots, m$).

3.1.1 Study Design

The study design draws on Chung, Rabe-Hesketh, and Choi (2013) Brockwell and Gordon (2001). For each number of studies $k \in \{10, 30, 100\}$ in a meta-analysis, k true study-level effects $\theta_1, \dots, \theta_k$ were drawn from a $\text{Normal}(\mu = 0.5, \tau^2)$ distribution, $\tau^2 \in \{0.01, 0.05, 0.1\}$. For each θ_i an observation $\hat{\theta}_i$ was drawn from a normal distribution with mean θ_i and variance σ_i^2 drawn from a $0.25 \cdot \chi^2(1)$ distribution truncated at 0.009 below and 0.6 above. The truncated χ^2 -distribution for the σ_i^2 was first used by Brockwell and Gordon (2001) due to its apparent ‘consistency’ with the distribution of sample variances of empirical *log-odds ratios*. One can thus think of the simulated θ_i ’s as true log-odds ratios and the $\hat{\theta}_i$ ’s as sample estimates with variances equal to the simulated σ_i^2 . Log-odds ratios (or odds ratios) are ubiquitous in medical research.

For each combination of k and τ^2 , 1,000 replications of the simulation were carried out. The normal random effects model (16) was fit to each replication using both ML- and REML-estimation as described in Section 2.2. The g -modeling equivalent given by 29 was fit as described in Section 2.4 with the following specifications: Θ was given by discretizing the interval between -1 and 1.5 into 100 equally spaced points; i.e., $\Theta = \{-1, -0.974, \dots, 1.5\}$.

The structure matrix Q was obtained as described in Section 2.1. In particular, a matrix of natural cubic spline bases with 5 degrees of freedom for the points in Θ was augmented by a column of 1’s, resulting in a 100×6 structure matrix Q . This is the default approach used in the R package `deconvolveR` Narasimhan and Efron (2020), which was specifically developed for the g -modeling approach outlined in Section 2.1.

Values of 0.05, 0.2, and 0.6 for the penalty factor c_0 under conditions of $\tau^2 = 0.01, 0.05$ and 0.1, respectively, were found to return generally favorable results in terms of the bias-variance trade-off for estimating g . The rationale underlying different choices of c_0 for each τ^2 may be stated as follows: The bias and variance of \hat{g} as an estimator can be gauged by increasing or decreasing the factor c_0 . An increase in c_0 ‘flattens’ the prior towards a uniform distribution, thus making the bias more pronounced. At the same time, the variance of \hat{g} viewed as an *estimator* for g is reduced. Flattening the prior however also *increases* the variance τ_g^2 of a particular realization of \hat{g} .¹ The dependence of τ_g^2 on c_0 also affects the width (and hence the coverage probability) of the Wald-type CIs for μ , which increases as τ_g^2 increases.

Choosing the ‘right’ value for c_0 thus requires considerable subject-matter insight into the problem at hand. For instance, if substantive background knowledge allows the researcher to expect a relatively large amount of variation in the true study effects, then the bias introduced by increasing c_0 may be negligible as in this case a “flatter” prior may still be a good approximation to the true prior. The goal of minimizing the variance of the estimator \hat{g} takes precedence over the goal of minimizing bias. On the other hand, if the between-study variation is expected to be small a priori, it may be wise to keep c_0 close to 0, as now the true prior distribution is expected to be more concentrated around a single value. Flattening \hat{g} (even by just a small amount) may cause it to deviate considerably from the true prior. The more concentrated the prior distribution, the smaller its variance τ^2 . Hence, the variances $\sigma_i^2 + \tau^2$ of the observations $\hat{\theta}_i$ also become smaller. This decrease at the level of the observations carries over to the estimator \hat{g} and the need to keep its variance in check becomes insignificant in comparison to the need of minimizing the bias of \hat{g} .

All analyses were carried out in *Julia* (Bezanson et al. 2017). In particular the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm as implemented in the `Optim.jl` package [?] was used to estimate $\alpha \in \mathbb{R}^6$ in the g -modeling and τ^2 and “classical” (maximum-likelihood and restricted maximum likelihood) approaches to meta-analysis, respectively.

3.2 Results

All results concerning the g -modeling approach are based on a bias-corrected version of the estimates of g . In particular if \hat{g} denotes the (uncorrected) estimate for g obtained by plugging $\hat{\alpha}$ into $g(\alpha)$, the bias corrected estimate is given by

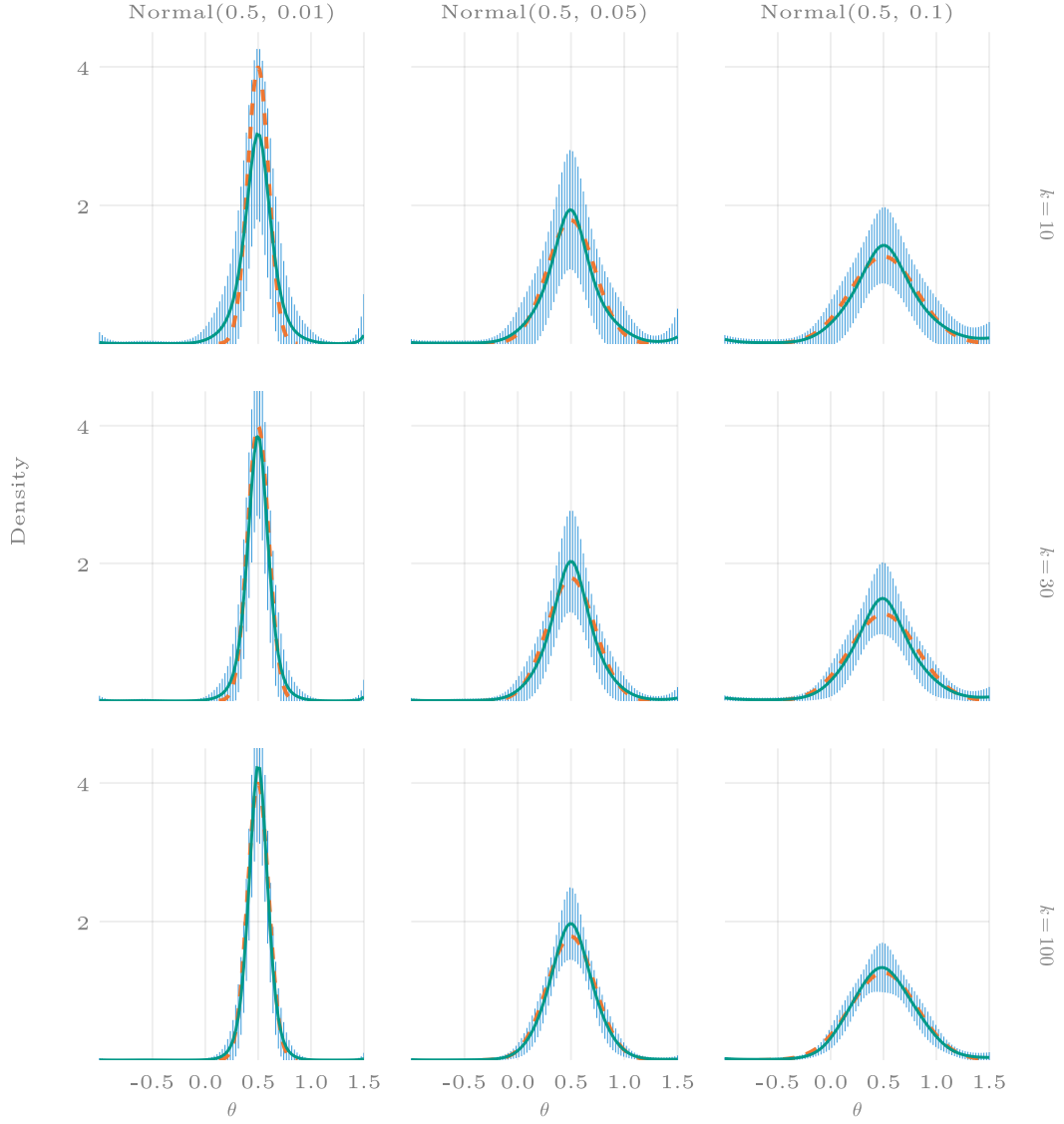
$$\hat{g}_{\text{BC}} = \hat{g} - \text{Bias}(\hat{g})$$

where $\text{Bias}(\hat{g})$ is given by (15). Since $\text{Bias}(\hat{g})$ is just an estimate of the true bias, some of the values (i.e., probabilities) in the density vector \hat{g}_{BC} might turn out to be negative. However, these negative values were observed only low-probability regions of \hat{g}_{BC} and appeared to be

¹Notice the distinction between the two types of variances mentioned in the last two sentences: In the first sentence, variance refers to the variation of the estimator \hat{g} in hypothetical repetitions of the experiment, whereas in the second sentence the term variance refers to τ_g^2 - an estimate of τ^2 .

miniscule in magnitude. In effect, these negative ‘probabilities’ were equal to zero thus suggesting the ad-hoc solution of replacing them by a small positive probability of 10^{-32} .

The prior density estimates are displayed in Figure 3.2. The solid blue curves connect the average densities over the 1,000 replications at each of the 100 values Θ sits on ($\Theta = \{-1, -0.974, \dots, 1.5\}$). The vertical error-bars correspond to ± 1 standard error (computed as the standard deviation of the 1,000 $\hat{g}_j, j = 1, \dots, m$) of these averages. For $\tau^2 = 0.01$, the g -modeling densities are clearly wider than the densities of the true distributions (dashed orange curves) from which the study-level θ_i ’s were drawn. Consequently, $\tau_{\hat{g}}$ overestimated the true τ^2 on average (even though c_0 was close to 0 and thus the bias was relatively low). As k increased, the shapes of the average g -modeling estimates seemed to approach that of the true prior distribution, $\text{Normal}(0.5, 0.01)$. Similar behavior can be observed for $\tau^2 = 0.05$ and $\tau^2 = 0.1$, albeit the average g -modeling estimates resembled the shapes of the corresponding true prior distribution much closer than did the estimates for $\tau^2 = 0.01$, even for small k . Moreover, the variance of the estimator \hat{g} decreased with increasing k as is indicated by the standard errors.



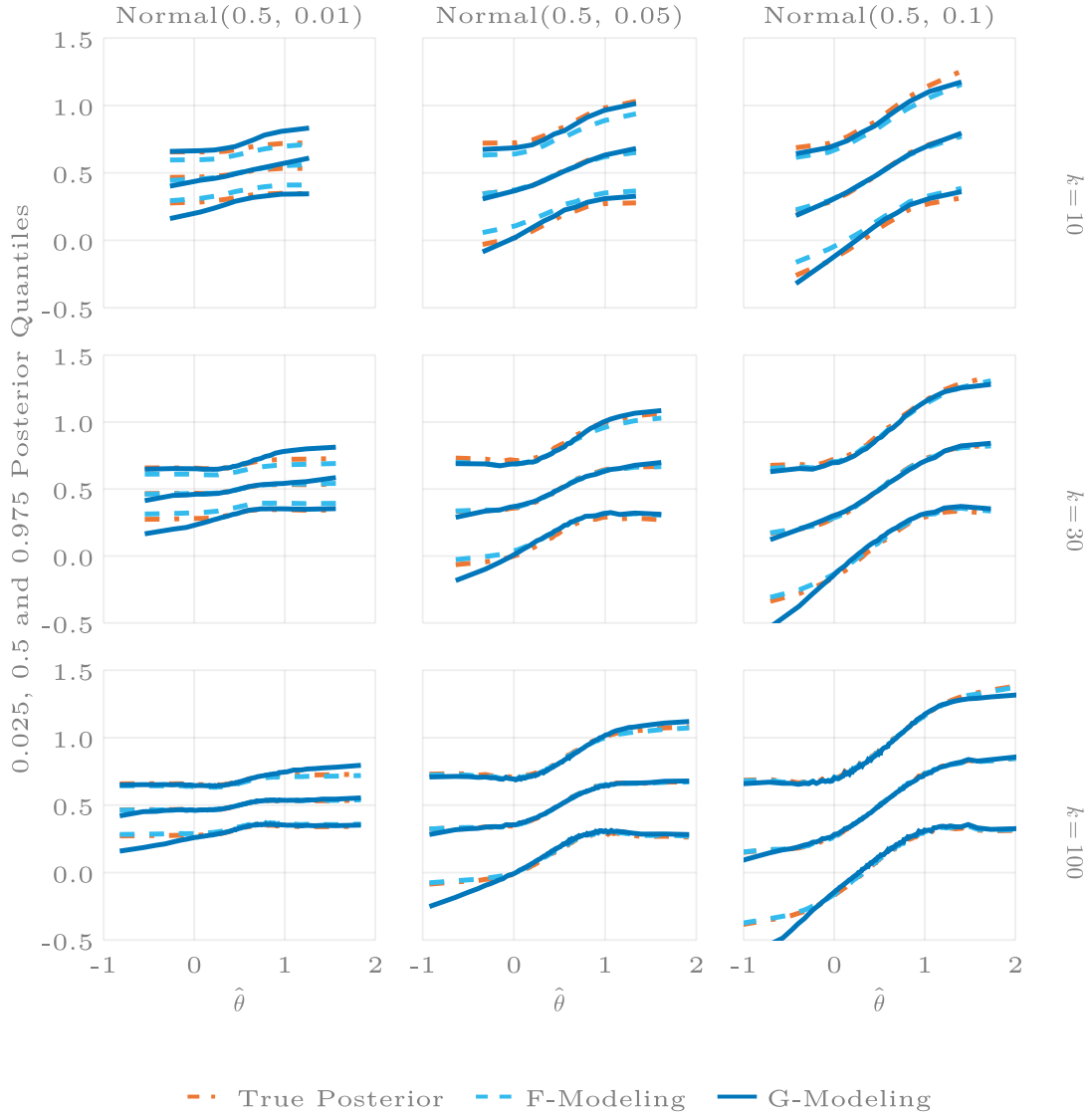
G -modeling estimates of prior distribution of study-level effects θ . The dashed orange curves indicate the true prior distribution from which study-level effects were drawn, the green curves connect the average densities of each of 100 equally spaced points over 1,000 simulation replicates. The upper and lower endpoints of the blue vertical bars correspond ± 1 standard error from the mean of the density estimates at each point. k denotes the number of study-level effects that were drawn (i.e., the number of studies in a meta-analysis).

The top row of Figure 3.2 illustrates the frequencies with which the different estimators $\tau_{\hat{g}}^2$ returned the substantively questionable estimate of a zero variance (i.e., $\tau^2 = 0$). Unsurpris-

ingly τ_g^2 never estimated τ^2 to be zero. Figure 3.2 nevertheless demonstrates that zero-variance estimates occurred quite often when using the ML- or REML-estimator, even when the number of studies in the meta-analysis was exceptionally large at $k = 100$.

The second and third row of Figure 3.2 show that τ_g^2 had the largest Bias of the three estimator across all combinations of k and τ^2 considered in the simulation study. However, τ_g^2 actually beat the REML- and the ML-estimators in terms of MSE for $k = 10$ and $\tau^2 = 0.1$ (note that Figure 3.2 shows the *root mean squared error* which is just $\sqrt{\text{MSE}}$).

The fourth row of Figure 3.2 shows the observed coverage probabilities of 95% Wald-type CIs for μ constructed using equation (27). For the ML- and REML-estimators, coverage probabilities were below the nominal 95% level in all situations, with the ML-estimator showing the largest deviation at $k = 10$ and $\tau^2 = 0.10$ (coverage below 90%). CIs based on τ_g^2 on the other hand were closest to the nominal level in most cases, thus performing consistently better than ML- and REML-based estimators.



Proportion of zero-variance estimates, root mean squared error (RMSE), and bias for τ^2 and coverage probabilities of 95% confidence intervals for μ over 1,000 replications.

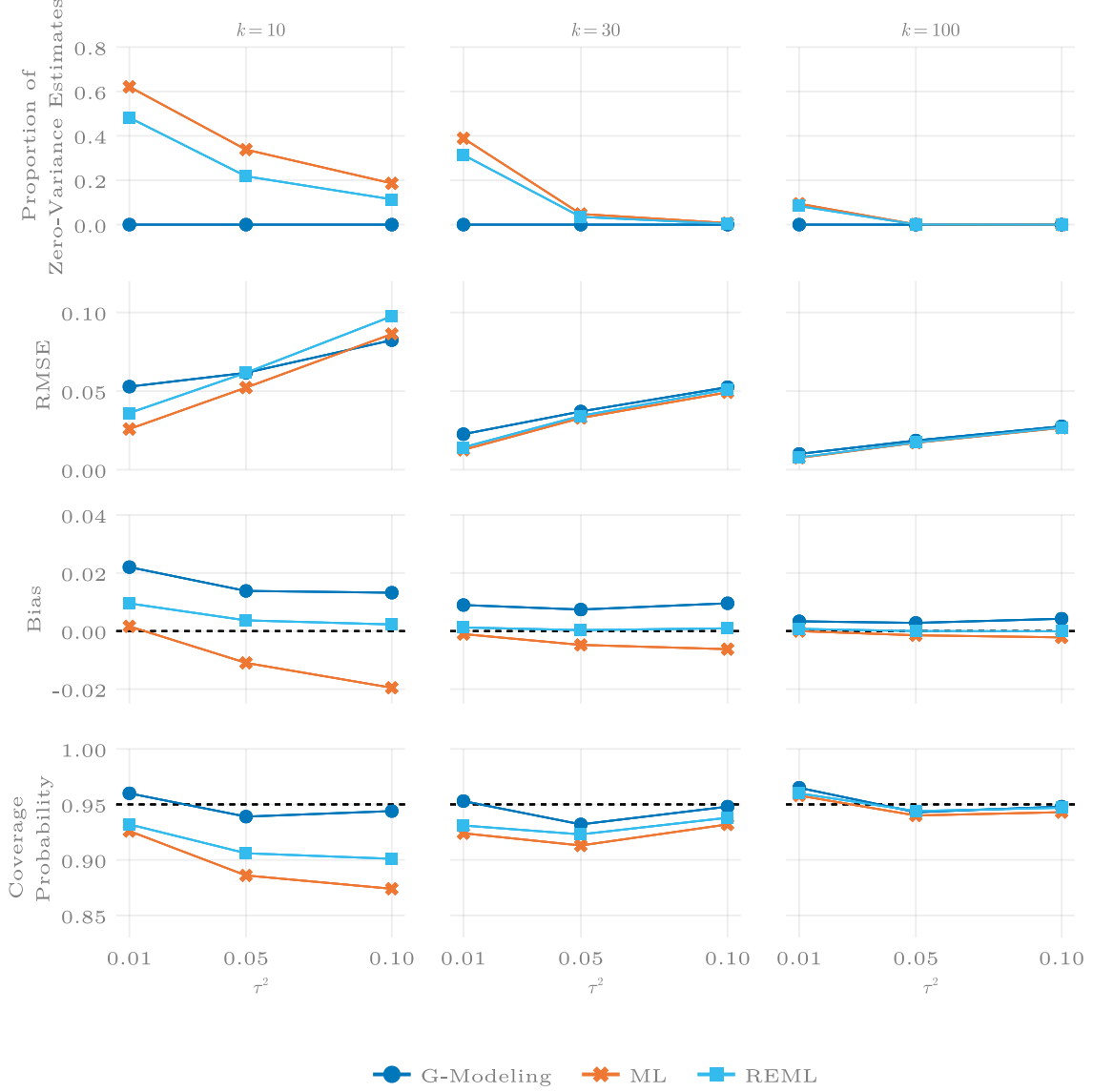
Figure 3.2 summarizes the posterior distributions for the individual θ_i 's by averaged 0.025, 0.5 and 0.975 quantiles. The individual curves were created as follows: For a given combination of k and τ^2 let $\hat{\theta}_{(s:i)}$ denote the i th *ordered* value of the simulated observations $\hat{\theta}_{s:1}, \dots, \hat{\theta}_{s:k}$ in simulation $s \in \{1, \dots, 1000\}$. The *inverse-variance* weighted mean over all 1,000 simulations

of the i th ordered observation is given by

$$\bar{\hat{\theta}}_{(i)} = \frac{\sum_{s=1}^{1000} \hat{\theta}_{s:i} / \sigma_{(s:i)}^2}{\sum_{s=1}^{1000} 1 / \sigma_{(s:i)}^2}$$

where $\sigma_{(s:i)}^2$ denotes the variance of the i th ordered observation. Similarly, let $\hat{g}_{(s:i)}$ denote the g -modeling posterior density for the i th ordered true effect $\theta_{(s:i)}$ in simulation s . The 0.025-, 0.5 and 0.975-quantiles of $\hat{g}_{(s:i)}$ were approximated by the corresponding sample quantiles observed in 2,000 random draws from $\hat{g}_{(s:i)}$. The interval between the 0.025- and the 0.975-quantile thus gives an approximation to the 95% central credible interval for $\theta_{(s:i)}$. These sample quantiles were averaged over the 1,000 simulations. Consequently, each weighted mean ordered observation $\bar{\hat{\theta}}_{(i)}$ was associated with a triple of averaged quantiles, $(q_{(i)}^{0.025}, q_{(i)}^{0.5}, q_{(i)}^{0.975})$. Analogous triples of averaged quantiles were calculated for the f -modeling-based posterior distribution and the true posterior distribution. These were obtained by plugging the estimate $\hat{\tau}_{\text{REML}}^2$ and the true τ^2 into (28), respectively.

In Figure 3.2 these different triples of average posterior quantiles are plotted against the $\bar{\hat{\theta}}_{(i)}$'s with the curves closest to the bottom of the graph connecting the 0.025 quantiles, the middle curves connecting the 0.5 quantiles and the curves closest to the top connecting the 0.975 quantiles.



Comparison of average posterior quantiles over 1,000 simulation replicates for each combination of number of studies k and distribution $\text{Normal}(0.5, \tau^2)$. Curves closest to the bottom correspond to averages of 0.025 quantiles, curves closest to the top correspond to averages of 0.975 quantiles, curves in between correspond to averages of 0.5 quantiles.

Visually, the g -modeling posterior quantiles (solid blue curves in Figure 3.2) seem to match the true posterior quantiles (dashed orange curves) quite well, even for small k . The biggest disagreement between the f -modeling (dashed blue curves) and g -modeling quantiles was observed for the combination $k = 10$ and $\tau^2 = 0.01$: Considering only the 0.025 and 0.975

quantiles, we see that f -modeling consistently underestimated these quantiles, thus indicating that on average the f -modeling 95% credible intervals for the θ_i 's were narrower than the true 95% credible intervals. For the g -modeling quantiles, the opposite was the case: g -modeling 95% credible intervals tended to overestimate the true 95% credible intervals.

4 Further Remarks

The aim of this investigation was to develop an approach for conducting a meta-analysis based on the g -modeling strategy to empirical Bayes and to compare its performance to the classical maximum- and restricted maximum-likelihood-based random-effects meta-analyses. The results of the simulation study suggest that g -modeling offers a viable strategy for carrying out an entire meta-analysis.

In particular, zero-variance estimates are avoided which not only results in more plausible inferences about the variation of true study effects but also improves upon the coverage probability of the Wald-type CIs constructed using (27). In addition, posterior quantiles for the individual study effects based on g -modeling closely matched those of the true posterior distributions. This suggests that g -modeling gives a reasonable answer to the question of estimating multiple quantities in parallel.

Two additional aspect of meta-analysis that have not been discussed so far pertain to the task of predicting the true effects of future studies, θ_{new} (e.g., Higgins, Thompson, and Spiegelhalter 2009) and assuming a non-normal or mixture distribution for the true study effects (i.e., the prior distribution). Concerning the former, numerous approaches to the construction of $(1 - \alpha)\%$ prediction intervals for θ_{new} exist. G -modeling seems to offer a particularly simple method: by simulating draws from the estimated prior distribution \hat{g} we are essentially simulating future study effects θ_{new} . It is then straightforward to determine the $\alpha/2$ and the $(1 - \alpha/2)$ sample quantiles of these draws and consider the interval bounded by these quantiles as an estimate of a $(1 - \alpha)\%$ prediction interval for a future true study effect (which is not the same as a future estimate).

As far as modelling the prior distribution as non-normal is concerned, g -modeling seems to offer an advantage over conventional meta-analytic methods. The normal random effects model assumes a normal distribution for the prior distribution of true study effects. This assumption may be called into question. G -modeling may be more flexible in this regard, allowing for non-normal (e.g., multimodal) prior distributions to be incorporated in the model.

References

- Bezanson, Jeff, Alan Edelman, Stefan Karpinski, and Viral B Shah. 2017. "Julia: A Fresh Approach to Numerical Computing." *SIAM Review* 59 (1): 65–98. <https://doi.org/10.1137/141000671>.

- Brockwell, Sarah E., and Ian R. Gordon. 2001. "A Comparison of Statistical Methods for Meta-Analysis." *Statistics in Medicine* 20 (6): 825–40. <https://doi.org/10.1002/sim.650>.
- Carroll, Raymond J., and Peter Hall. 1988. "Optimal Rates of Convergence for Deconvolving a Density." *Journal of the American Statistical Association* 83 (404): 1184–86. <https://doi.org/10.1080/01621459.1988.10478718>.
- Chung, Yeojin, Sophia Rabe-Hesketh, and In-Hee Choi. 2013. "Avoiding Zero Between-Study Variance Estimates in Random-Effects Meta-Analysis." *Statistics in Medicine* 32 (23): 4071–89. <https://doi.org/10.1002/sim.5821>.
- Efron, Bradley. 2014. "Two Modeling Strategies for Empirical Bayes Estimation." *Statistical Science* 29 (2): 285–301. <https://doi.org/10.1214/13-STS455>.
- . 2016. "Empirical Bayes Deconvolution Estimates." *Biometrika* 103 (1): 1–20. <https://doi.org/10.1093/biomet/asv068>.
- Efron, Bradley, and Trevor Hastie. 2016. *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Institute of Mathematical Statistics Monographs. New York, NY: Cambridge University Press.
- Harville, David A. 1977. "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." *Journal of the American Statistical Association* 72 (358): 320–38. <https://doi.org/10.2307/2286796>.
- Higgins, Julian P. T., Simon G. Thompson, and David J. Spiegelhalter. 2009. "A Re-Evaluation of Random-Effects Meta-Analysis." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172 (1): 137–59. <https://doi.org/10.1111/j.1467-985X.2008.00552.x>.
- Lehmann, E. L., and George Casella. 1998. *Theory of Point Estimation*. 2nd ed. Springer Texts in Statistics. New York: Springer.
- Narasimhan, Balasubramanian, and Bradley Efron. 2020. "deconvolveR: A G-Modeling Program for Deconvolution and Empirical Bayes Estimation." *Journal of Statistical Software* 94 (September): 1–20. <https://doi.org/10.18637/jss.v094.i11>.
- Raudenbush, Stephen W., and Anthony S. Bryk. 1985. "Empirical Bayes Meta-Analysis." *Journal of Educational Statistics* 10 (2): 75–98. <http://www.jstor.org/stable/1164836>.
- Viechtbauer, Wolfgang. 2005. "Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model." *Journal of Educational and Behavioral Statistics* 30 (3): 261–93. <https://doi.org/10.3102/10769986030003261>.