
Selecting an Augmented Random Effects Model

A. L. Rukhin*

*Statistical Engineering Division National Institute of Standards and Technology
100 Bureau Dr. Gaithersburg, MD 20899 USA*

Received August 26, 2021; revised December 11, 2021; accepted December 11, 2021

Abstract—There are many collaborative studies where the data are discrepant while uncertainty estimates reported in each study cannot be relied upon. The classical commonly used random effects model explains this phenomenon by additional noise with a constant heterogeneity variance. This assumption may be inadequate especially when the smallest uncertainty values correspond to the cases which are most deviant from the bulk of data. An augmented random effects model for meta-analysis of such studies is offered. It proposes to think about the data as consisting of different classes with the same heterogeneity variance only within each cluster. The choice of the classes is to be made on the basis of the classical or restricted likelihood. We discuss the properties of the corresponding procedures which indicate the studies whose heterogeneity effect is to be enlarged. The conditions for the convergence of several iterative algorithms are given.

DOI: 10.3103/S1066530720040043

Keywords: heterogeneity, information criterion, maximum likelihood, meta-analysis

1. INTRODUCTION: HETEROSCEDASTIC META-ANALYSIS MODEL

In a research synthesis problem one has to combine several estimates of the quantity of interest μ . The popular random effects model (REM) postulates the form of these estimators $x_i, i = 1, \dots, N$,

$$x_i = \mu + \lambda_i + \epsilon_i. \quad (1)$$

Here, the parameter of primary focus μ is the common mean representing the treatment effect in biostatistics or the reference value in metrology. The measurement error ϵ_i of the i th study is supposed to have zero mean and the variance representing the within-study variability. The protocol requires that the participants who possibly use different measuring techniques report not only their estimates of μ , but also the estimates s_i^2 of this variance (within study uncertainty).

The term λ_i is commonly taken to have zero mean with some heterogeneity variance traditionally denoted by τ^2 . One can view λ_i as the additional noise imposed by Nature on the experiment, where one observes the results of independent individual studies all measuring the same overall effect μ . This variance can be estimated by one of the developed procedures [8, 10]. The classical fixed effects model (FEM) corresponding to the situation when $\tau^2 = 0$ provides a poor fit in many practical situations [1]. Indeed the heterogeneity in most research synthesis studies can be quite substantial (Thompson and Sharp, 1999). Moreover, some medical researchers believe that “examination of heterogeneity is perhaps the most important task in meta-analysis” [7, p. 30]. In many heterogeneous studies the assumption that all λ 's have the same dispersion seems to be violated if the smallest reported uncertainties correspond to the cases which are most deviant from the rest of data. Indeed this condition appears to be due mainly to mathematical expediency and limits the applicability of REM. We advocate a more detailed analysis by admitting several distinct heterogeneity variances whose larger values are assigned to the aberrant, outlying cases. The most homogeneous data subset gets zero heterogeneity variance.

*E-mail: andrew.rukhin@nist.gov

By independence of λ_i and ϵ_i , the variance of x_i cannot be smaller than the variance of ϵ_i typically estimable by s_i^2 although it commonly includes non-statistically derived components. In many situations there is doubt about the validity of these estimates especially when they are identified with the unknown uncertainties. Hoaglin [10] discusses possible dire consequences of such identification in the homogeneity hypothesis testing problem $\tau^2 = 0$. The case of REM allowing τ^2 to depend on the study was investigated by Rukhin (2019a) who suggested to use s_i only as a lower bound on the unknown i th uncertainty so that $\tau_i^2 \geq s_i^2$. In this work s_i^2 also serves as the lower bound for the unknown variance of x_i .

More generally one can entertain an augmented random effects model (AREM) which puts the studies into different classes or clusters according to the several values of the unknown heterogeneity variance, say, $\tau^2 = \tau_k^2 > 0, k = 1, \dots, K$. Then the cluster corresponding to $\tau_0^2 = 0$ plays a special role: it includes all conforming studies satisfying FEM. This class represents the largest consistent subset of all studies allowing to identify cases affected by “excess-variation” from likelihood calculations. Its definition is a controversial issue in metrology [3, 20, 4]. We illustrate our definition of this concept by two practical examples in Section 6. The situation when each heterogeneous cluster consists of just one element corresponds to the mentioned setting with s_i^2 representing a lower bound on the variance of x_i .

There is a body of work aimed at extending REM for meta-analysis needs. See Ohlssen et al., 2007, Lee and Thompson, 2008, or Kulinskaya and Olkin, 2014 for more flexible parametric and nonparametric models. The main objective of this paper is to suggest a methodology for selecting one of AREM models with distinct heterogeneity variances. We use maximum likelihood estimators of the mean and variances which provide likelihood-based information criteria studied numerically in Section 7. For this purpose in the next section iterative algorithms to find the maximum likelihood procedures and the conditions for their convergence are discussed. Section 3 contains a necessary condition for the global extremum, and Section 4 gives an example where s_i^2 are equal to illustrate that the traditional REM have smaller likelihood than some other AREM models. Section 5 discusses the algebraic difficulty of the likelihood equations. The proofs are collected in the Appendix.

2. MAXIMUM LIKELIHOOD ESTIMATORS

Assume that the data are represented by a series of independent but not equally distributed random variables $x_i \sim N(\mu, \sigma_i^2)$ with the unknown common mean μ and variances $\sigma_i^2, \sigma_i^2 = \tau_i^2 + s_i^2$, where $\tau_i^2 \geq 0$ is unknown and s_i^2 represents the within-study variance, $i = 1, \dots, N$. Supposing that there are no more than K different positive τ 's, the variance of x_i for i in the class I_k is $s_i^2 + \tau_k^2$. Clusters I_0, I_1, \dots, I_K define a partition of the set $\{1, \dots, N\}$ with I_k containing n_k elements, $n_0 + n_1 + \dots + n_K = N$. The total number of heterogeneous clusters K cannot exceed $N - n_0$, and there is no generality loss in assuming that $n_1 \geq \dots \geq n_K$. The special cluster I_0 corresponds to the vanishing heterogeneity variance $\tau_0^2 = 0$.

For given n_0 and K , the number of non-empty heterogeneity clusters is the *Stirling number of the second kind* $\left\{ \begin{matrix} N - n_0 \\ K \end{matrix} \right\}$ (Graham et al., 1994). Thus the total number of all different partitions is

$$\sum_{0 \leq n_0 \leq N} \left\{ \begin{matrix} N - n_0 \\ K \end{matrix} \right\} \binom{N}{n_0} = \left\{ \begin{matrix} N + 1 \\ K + 1 \end{matrix} \right\},$$

which allows I_0 to be empty. When $K = 1$, this number is $2^N - 1$, for $K = 2$ it is $(3^N - 2^{N+1} + 1)/2$, for $K = N - 1$, it is $N(N + 1)/2$. An implication is that in practice either K should be chosen to be small or close to N .

For given clusters I_0, I_1, \dots, I_K the (classical) log-likelihood function (times -2) is

$$\begin{aligned} & L(\mu, \tau_1^2, \dots, \tau_K^2; I_0, I_1, \dots, I_K) \\ &= L(\mu, \tau_1^2, \dots, \tau_K^2) = \sum_{k=0}^K \sum_{I_k} \left[\frac{(x_i - \mu)^2}{\tau_k^2 + s_i^2} + \log(\tau_k^2 + s_i^2) \right], \quad \tau_0^2 = 0. \end{aligned} \quad (2)$$

We call (2) the augmented random effects model (AREM.CL.K) likelihood. A standard argument shows that for given clusters the model AREM.CL.K is identifiable.

The form of the restricted likelihood function \tilde{L} (AREM.RL.K) is also well known:

$$\begin{aligned} \tilde{L}(\tau_1^2, \dots, \tau_K^2) &= \sum_{k=0}^K \sum_{I_k} \left[\frac{(x_i - \tilde{\mu})^2}{\tau_k^2 + s_i^2} + \log(\tau_k^2 + s_i^2) \right] \\ &+ \log \left(\sum_{k=0}^K \sum_{I_k} \frac{1}{\tau_k^2 + s_i^2} \right). \end{aligned} \quad (3)$$

Here the cluster weighted mean

$$\tilde{\mu} = \tilde{\mu}(\tau_1^2, \dots, \tau_K^2) = \sum_{k=0}^K \sum_{I_k} \frac{x_i}{\tau_k^2 + s_i^2} \left[\sum_{k=0}^K \sum_{I_k} \frac{1}{\tau_k^2 + s_i^2} \right]^{-1} \quad (4)$$

is the best linear unbiased estimator of μ which also minimizes (2) in μ for given τ 's.

Thus to find strictly positive maximum likelihood estimates $\hat{\tau}_k^2$, one has to solve equations

$$\frac{\partial L}{\partial \tau_k^2} = 0, \quad k = 1, \dots, K,$$

with positive definite matrix of second derivatives. These equations can be written in the form

$$\sum_{I_k} \frac{(x_i - \tilde{\mu})^2 - s_i^2}{(\tau_k^2 + s_i^2)^2} = \tau_k^2 \sum_{I_k} \frac{1}{(\tau_k^2 + s_i^2)^2}. \quad (5)$$

The following Theorem 2.1 gives an explicit form of the Hessian and offers an iterative algorithm motivated by (5).

Let a and d be K -dimensional vectors with the coordinates

$$a_k = 2^{1/2} \sum_{I_k} \frac{x_i - \tilde{\mu}}{(\tau_k^2 + s_i^2)^2} \left(\sum_m \sum_{I_m} \frac{1}{\tau_m^2 + s_i^2} \right)^{-1/2}$$

and

$$d_k = 2 \sum_{I_k} \frac{(x_i - \tilde{\mu})^2}{(\tau_k^2 + s_i^2)^3} - \sum_{I_k} \frac{1}{(\tau_k^2 + s_i^2)^2}, \quad k = 1, \dots, K.$$

Theorem 2.1. For AREM.CL.K the Hessian $H = (\partial^2 L / [\partial \tau_k^2 \partial \tau_\ell^2])_{k,\ell=1}^K$ has the form

$$H = \text{diag}(d) - aa^T. \quad (6)$$

With $\hat{\tau}_k^2$ substituted for τ_k^2 in the definition of a, d , and $\tilde{\mu}$ to get $\hat{a}, \hat{d}, \hat{\mu}$, the sufficient condition for the minimum of (2) to be attained is:

$$\sum_{k=1}^K \frac{\hat{a}_k^2}{\hat{d}_k} < 1, \quad \hat{d}_k > \hat{a}_k^2, \quad k = 1, \dots, K. \quad (7)$$

A necessary condition is that $\sum_{k=1}^K \hat{a}_k^2 / \hat{d}_k \leq 1$.

Provided that (7) holds and

$$\hat{d}_k - \hat{a}_k^2 < 2 \sum_{I_k} \frac{1}{(\hat{\tau}_k^2 + s_i^2)^2}, \quad k = 1, \dots, K, \quad (8)$$

the positive maximum likelihood estimators $\hat{\tau}_k^2, k = 1, \dots, K$ exist and can be determined by iterations as

$$\hat{\tau}_k^2 = \max \left[0, \sum_{I_k} \frac{(x_i - \hat{\mu})^2 - s_i^2}{(\hat{\tau}_k^2 + s_i^2)^2} \right] \left[\sum_{I_k} \frac{1}{(\hat{\tau}_k^2 + s_i^2)^2} \right]^{-1}. \tag{9}$$

There are many numerical methods like the Newton–Raphson rule which need only condition (7) for convergence. They are more reliable than the EM algorithm and can be used instead of (9) or (14) in the following Theorem 2.2. An attractive feature of iterations in Theorem 2.1 is that they decrease the value of (2) at each step (Rukhin, 2011). This fact matters since the likelihood may not be a unimodal function of τ^2 ; indeed the (polynomial) Eq. (5) can have several positive solutions. Therefore the choice of a good starting value is important.

To find the global optimizers $\hat{I}_0, \dots, \hat{I}_K$, enumerate all $\left\{ \begin{matrix} N + 1 \\ K + 1 \end{matrix} \right\}$ partitions I_0, \dots, I_K of the index set $\{1, \dots, N\}$ and for each partition determine the solution of (5) via Theorem 2.1. Failure to converge is interpreted as non-existence of solutions satisfying (8) resulting in rejection of the candidate clustering. The partition $\hat{I}_0, \dots, \hat{I}_K$ which provides the overall minimizer is taken as the final choice delivering the maximum likelihood estimators $\hat{\tau}_1^2, \dots, \hat{\tau}_K^2$ and $\hat{\mu}$. A necessary optimality condition is provided in Section 3.

In the simplest but important case, $K = 1$ (AREM.CL.K = 1), we have to choose two clusters, one with zero heterogeneity I_0 , another its heterogeneous complement I_1 . Then one has to determine the likelihood defined by the conditions: for $i \in I_1$, $\tau_i^2 = \tau^2$ with unknown but positive τ^2 , while I_0 corresponds to j 's for which $\tau_j^2 = 0$. The maximum likelihood estimator of μ is of the form (4) with $\tau^2 = \hat{\tau}^2$,

$$\hat{\mu} = \left[\sum_{I_0} \frac{x_j}{s_j^2} + \sum_{I_1} \frac{x_i}{\hat{\tau}^2 + s_i^2} \right] \left[\sum_{I_0} \frac{1}{s_j^2} + \sum_{I_1} \frac{1}{\hat{\tau}^2 + s_i^2} \right]^{-1},$$

and $\hat{\tau}^2 = \hat{\tau}^2(I_1)$ is a solution of the equation

$$\sum_{I_1} \frac{(x_i - \hat{\mu})^2}{(\hat{\tau}^2 + s_i^2)^2} = \sum_{I_1} \frac{1}{\hat{\tau}^2 + s_i^2}. \tag{10}$$

To get the μ -estimator for a fixed I_1 , it suffices to solve (10) choosing the true minimizer of $\min_{\mu, \tau^2} L(\mu, \tau^2)$ out of possibly several solutions with positive second derivative, i.e.,

$$\begin{aligned} & \left(\sum_{I_1} \frac{1}{\hat{\tau}^2 + s_i^2} + \sum_{I_0} \frac{1}{s_j^2} \right) \sum_{I_1} \frac{(x_i - \hat{\mu})^2}{(\hat{\tau}^2 + s_i^2)^3} - \left(\sum_{I_1} \frac{x_i - \hat{\mu}}{(\hat{\tau}^2 + s_i^2)^2} \right)^2 \\ & > \frac{1}{2} \left(\sum_{I_1} \frac{1}{\hat{\tau}^2 + s_i^2} + \sum_{I_0} \frac{1}{s_j^2} \right) \sum_{I_1} \frac{1}{(\hat{\tau}^2 + s_i^2)^2}. \end{aligned} \tag{11}$$

When $I_0 = \emptyset$, $I_1 = \{1, \dots, N\}$, one gets the setting of traditional REM with just one τ^2 and (9) presents a commonly used procedure to determine this parameter (e.g., Rukhin, 2019b). However in all examples the likelihood of this model cannot exceed that of models with non-empty I_0 (see Section 4). The seemingly new convergence conditions (7) and (8) mean that in this situation $\hat{a}_1^2 < \hat{d}_1 < 2 \sum_{I_1} (\hat{\tau}^2 + s_i^2)^{-2} + \hat{a}_1^2$. The likelihood of this case typically is smaller than that of some other AREM.CL.K = 1 models.

Another special case is the model $AREM.CL.K = N - n_0$, with lower bounded variances, $n_k \equiv 1, k \notin I_0$. In this situation

$$\hat{\mu} = \left[\sum_{k \notin I_0} \frac{x_k}{\hat{\tau}_k^2 + s_k^2} + \sum_{I_0} \frac{x_j}{s_j^2} \right] \left[\sum_{k \notin I_0} \frac{1}{\hat{\tau}_k^2 + s_k^2} + \sum_{I_0} \frac{1}{s_j^2} \right]^{-1}.$$

Then the iteration scheme of Theorem 2.1 can be reduced to a one-dimensional problem involving only $\hat{\mu}$.

Theorem 2.2. *A sufficient condition for the minimum in $AREM.CL.K = N - n_0$ is that for any $k \notin I_0$,*

$$\frac{2}{\hat{\tau}_k^2 + s_k^2} < \sum_{\ell \notin I_0} \frac{1}{\hat{\tau}_\ell^2 + s_\ell^2} + \sum_{I_0} \frac{1}{s_j^2} < \sum_{I_0} \frac{2}{s_j^2}. \tag{12}$$

A non-strict version of (12) and inequalities

$$(x_k - \hat{\mu})^2 > s_k^2, \quad k \notin I_0, \tag{13}$$

form necessary conditions. The iteration scheme to find the maximum likelihood estimator

$$\hat{\mu} = \left[\sum_{k \notin I_0} \frac{x_k}{(x_k - \hat{\mu})^2} + \sum_{I_0} \frac{x_j}{s_j^2} \right] \left[\sum_{k \notin I_0} \frac{1}{(x_k - \hat{\mu})^2} + \sum_{I_0} \frac{1}{s_j^2} \right]^{-1} \tag{14}$$

converges if (12) holds. Under condition (13) the maximum likelihood estimators $\hat{\tau}_k^2 = (x_k - \hat{\mu})^2 - s_k^2, k \notin I_0$, are positive.

Similar results can be derived for restricted maximum likelihood estimators of τ^2 with K simultaneous equations

$$\sum_{i \in I_k} \frac{(x_i - \tilde{\mu})^2}{(\tilde{\tau}_k^2 + s_i^2)^2} = \sum_{i \in I_k} \frac{1}{\tilde{\tau}_k^2 + s_i^2} - \sum_{i \in I_k} \frac{1}{(\tilde{\tau}_k^2 + s_i^2)^2} \left(\sum_m \sum_{i \in I_m} \frac{1}{s_i^2 + \tilde{\tau}_m^2} \right)^{-1}. \tag{15}$$

Define the K -dimensional vectors b and c by their coordinates

$$b_k = \sum_{I_k} \frac{1}{(\tau_k^2 + s_i^2)^2} \left(\sum_m \sum_{I_m} \frac{1}{\tau_m^2 + s_i^2} \right)^{-1}$$

and

$$c_k = 2 \sum_{I_k} \frac{(x_i - \tilde{\mu})^2}{(\tau_k^2 + s_i^2)^3} - \sum_{I_k} \frac{1}{(\tau_k^2 + s_i^2)^2} + \sum_{I_k} \frac{2}{(\tau_k^2 + s_i^2)^3} \left(\sum_m \sum_{I_m} \frac{1}{\tau_m^2 + s_i^2} \right)^{-1},$$

$k = 1, \dots, K$. By substituting the restricted maximum likelihood estimators $\tilde{\tau}_k^2$, one gets vectors \tilde{b}, \tilde{c} as well as \tilde{a} .

Theorem 2.3. *For $AREM.RL.K$ the Hessian corresponding to (15) has the form*

$$\tilde{H} = C - \tilde{a}\tilde{a}^T - \tilde{b}\tilde{b}^T, \tag{16}$$

$C = \text{diag}(\tilde{c})$. *A sufficient condition for the minimum of $\tilde{L}(\tau_1^2, \dots, \tau_K^2)$ in (3) to be attained at $\tilde{\tau}_k^2, k = 1, \dots, K$ is that for all k*

$$\tilde{a}_k^2 + \tilde{b}_k^2 < \tilde{c}_k \tag{17}$$

and

$$\sum_k \frac{\tilde{a}_k^2 + \tilde{b}_k^2}{\tilde{c}_k} < \min \left[1 + \sum_k \frac{\tilde{a}_k^2}{\tilde{c}_k} \sum_k \frac{\tilde{b}_k^2}{\tilde{c}_k} - \left(\sum_k \frac{\tilde{a}_k \tilde{b}_k}{\tilde{c}_k} \right)^2, 2 \right]. \tag{18}$$

For any minimizer non-strict inequalities in (17) and (18) must hold. Under these conditions an iteration scheme to find positive restricted maximum likelihood estimators $\hat{\tau}_k^2, k = 1, \dots, K$,

$$\hat{\tau}_k^2 = \max \left\{ 0, \sum_{I_k} \frac{(x_i - \tilde{\mu})^2 - s_i^2 + [\sum_m \sum_{I_m} (\hat{\tau}_m^2 + s_i^2)^{-1}]^{-1}}{(\hat{\tau}_k^2 + s_i^2)^2} \right\} \left[\sum_{I_k} \frac{1}{(\hat{\tau}_k^2 + s_i^2)^2} \right]^{-1},$$

$\tilde{\mu} = \tilde{\mu}(\hat{\tau}_1^2, \dots, \hat{\tau}_K^2)$, converges if (17) and (18) are valid.

Specification of Theorem 2.3 for $\text{AREM.RL.K} = 1$ and $\text{AREM.RL.K} = N - n_0$ is given in Section 9.4; heuristic discussion of the conditions in Theorems 2.1–2.3 is postponed until Section 4.

3. CONDITION FOR GLOBAL EXTREMUM

In this section we give a necessary condition for attaining the global extremum of (2) over all clusters I_0, I_1, \dots, I_K .

Given the partition I_0, I_1, \dots, I_K , consider $L(\tilde{\mu}, \tau_1^2, \dots, \tau_K^2)$ as a function of $\tau_1^2, \dots, \tau_K^2$. Let

$$S = S(\tau_1^2, \dots, \tau_K^2) = \sum_m \sum_{I_m} (\tau_m^2 + s_i^2)^{-1}, \tag{19}$$

so that $\tilde{\mu} = \sum_m \sum_{I_m} x_i (\tau_m^2 + s_i^2)^{-1} / S$.

For a fixed $q = 0, \dots, K$, choose any $n \in I_q$ which is moved from I_q to I_p and put $\bar{I}_q = I_q \setminus \{n\}$ with the set I_p ($p \neq q$) replaced by $\bar{I}_p = I_p \cup \{n\}$. Let $\bar{I}_k = I_k, k \neq p, q$, and denote the similar modification of $\tilde{\mu}$ by

$$\bar{\mu} = \sum_m \sum_{\bar{I}_m} \frac{x_i}{\tau_m^2 + s_i^2} \left[\sum_m \sum_{\bar{I}_m} \frac{1}{\tau_m^2 + s_i^2} \right]^{-1}, = \frac{x_n \Delta + \tilde{\mu} S}{\Delta + S}, \tag{20}$$

$\Delta = (\tau_p^2 + s_n^2)^{-1} - (\tau_q^2 + s_n^2)^{-1}$. In this notation one has for any $\tau_k^2, k = 1, \dots, K$

$$\sum_k \sum_{\bar{I}_k} \frac{(x_i - \bar{\mu})^2}{\tau_k^2 + s_i^2} = \sum_k \sum_{I_k} \frac{(x_i - \tilde{\mu})^2}{\tau_k^2 + s_i^2} + \frac{\Delta S (x_n - \tilde{\mu})^2}{\Delta + S}. \tag{21}$$

To prove (21) write the sum in its left-hand side as

$$\sum_k \sum_{I_k} \frac{(x_i - \bar{\mu})^2}{\tau_k^2 + s_i^2} + \Delta (x_n - \bar{\mu})^2 = \sum_k \sum_{I_k} \frac{(x_i - \tilde{\mu})^2}{\tau_k^2 + s_i^2} - S (\tilde{\mu} - \bar{\mu})^2 + \frac{\Delta S^2 (x_n - \tilde{\mu})^2}{(\Delta + S)^2}$$

and employ (20) to simplify.

This formula and the representation (2) of the likelihood function lead to the following result.

Theorem 3.1. *If in AREM.CL.K the clusters I_0, I_1, \dots, I_K with estimates $\hat{\tau}_1^2, \dots, \hat{\tau}_K^2$ and $\hat{\mu}$ provide the global minimum of (2), then for any $n \in I_q$ and $p, 0 \leq p \neq q \leq K$, one must have*

$$\frac{\Delta S (x_n - \hat{\mu})^2}{\Delta + S} \geq \log \left(\frac{\hat{\tau}_q^2 + s_n^2}{\hat{\tau}_p^2 + s_n^2} \right). \tag{22}$$

Here S is defined by (19) with $\hat{\tau}_1^2, \dots, \hat{\tau}_K^2$ replacing $\tau_1^2, \dots, \tau_K^2$.

According to Theorem 3.1 the distance between elements of heterogeneity clusters $I_k, k \geq 1$ and $\hat{\mu}$ measured relative to τ_k^2 cannot be too small. Therefore the optimal homogeneity cluster I_0 must consist of data points which are fairly close to the consensus estimate $\hat{\mu}$.

4. EQUAL VARIANCES

To elucidate the conditions of Theorems 2.1–2.3 in this section we look at the simplest case $K = 1$ when all heterogeneity is in x 's, i.e., s_i^2 are equal, $s_i^2 \equiv s^2$. Then $\hat{\mu} = \bar{x} = \sum x_i/N$, $\hat{\tau}^2 = \max[0, \sum_i (x_i - \bar{x})^2/N - s^2]$. Provided that $\hat{\tau}^2 > 0$,

$$\mathcal{L}(\emptyset, \{1, \dots, N\}) = N[1 + \log(\hat{\tau}^2 + s^2)],$$

which can be assumed smaller than $\mathcal{L}(\{1, \dots, N\}, \emptyset)$. For a heterogeneity cluster $I = I_1$ of cardinality M ,

$$\bar{\mu} = \frac{\sum_{i \notin I} x_i/s^2 + \sum_{i \in I} x_i/(\tau^2 + s^2)}{(N - M)/s^2 + M/(\tau^2 + s^2)} = (1 - \omega)\bar{x}_{I^c} + \omega\bar{x}_I,$$

where $\omega = M(\tau^2 + s^2)^{-1}/[(N - M)s^{-2} + M(\tau^2 + s^2)^{-1}]$, $\omega \leq \omega_0 = M/N$, $\bar{x}_I = \sum_{i \in I} x_i/M$, $\bar{x}_{I^c} = \sum_{i \notin I} x_i/(N - M)$, $N \geq 3$.

By putting $\alpha_I = \sum_{i \in I} (x_i - \bar{x}_I)^2/(Ms^2)$, $\beta_I = \sum_{i \notin I} (x_i - \bar{x}_{I^c})^2/[(N - M)s^2]$, $\gamma_I = (\bar{x}_I - \bar{x}_{I^c})^2/s^2$, it is convenient to work with the difference between likelihoods

$$\begin{aligned} [L(\tilde{\mu}, \tau^2(\omega); I^c, I) - \mathcal{L}(\emptyset, \{1, \dots, N\})]/N &= \omega_0 \log \left(\frac{\omega_0(1 - \omega)}{(1 - \omega_0)\omega} \right) \\ &+ \frac{(1 - \omega_0)}{1 - \omega} [\alpha_I\omega + \beta_I(1 - \omega) + \gamma_I\omega(1 - \omega)] - 1 - \log(v). \end{aligned} \tag{23}$$

Here

$$v = \omega_0\alpha_I + (1 - \omega_0)\beta_I + \omega_0(1 - \omega_0)\gamma_I = \frac{\sum_i (x_i - \bar{x})^2}{Ns^2}. \tag{24}$$

The quantities a_1 and d_1 from Theorem 2.1 are such that

$$\begin{aligned} a_1^2 &= \frac{2M(1 - \omega)^2\omega s^2\gamma_I}{(\tau^2 + s^2)^3} = \frac{2(N - M)\omega^2(1 - \omega)\gamma_I}{(\tau^2 + s^2)^2}, \\ d_1 &= \frac{2(N - M)\omega[\alpha_I + \gamma_I(1 - \omega)^2]}{(1 - \omega)(\tau^2 + s^2)^2} - \frac{M}{(\tau^2 + s^2)^2}. \end{aligned}$$

The condition (7) there means that at maximum likelihood solution

$$\alpha_I + \gamma_I(1 - \omega)^3 \geq \frac{\omega_0(1 - \omega)}{2(1 - \omega_0)\omega}.$$

which means that $\alpha_I \geq [\omega_0(2\omega - 1)(1 - \omega)][2(1 - \omega_0)\omega^2]$ and then the iteration algorithm is convergent. The condition 12 of Theorem 2.2 is much simpler, $\omega < \frac{1}{2}$, indicating that the relative weight of increased variance cases cannot exceed one-half.

The restricted likelihoods (up to a constant term) are $\tilde{\mathcal{L}}(\emptyset, \{1, \dots, N\}) = (N - 1)[1 + \log(Nv/(N - 1))]$, $v > (N - 1)/N$, and

$$\begin{aligned} \tilde{L}(\tilde{\mu}, \tau^2(\omega); I^c, I) &= \log \left(\frac{1 - \omega_0}{1 - \omega} \right) \\ &+ N \left[\omega_0 \log \left(\frac{\omega_0(1 - \omega)}{(1 - \omega_0)\omega} \right) + \frac{(1 - \omega_0)}{1 - \omega} [\alpha_I\omega + \beta_I(1 - \omega) + \gamma_I\omega(1 - \omega)] \right]. \end{aligned}$$

In Theorem 2.3 one has $b_1 = \omega^2(\tau^2 + s^2)^{-2}$ and

$$c_1 = b_1 + \frac{2\omega}{(\tau^2 + s^2)^2},$$

so that the condition (17) means that

$$\alpha_I \geq \frac{(1 - \omega)(M - 2\omega + \omega^2)}{2(N - M)\omega}.$$

Since $K = 1$, (18) is equivalent to (17) and under this condition the iteration algorithm converges.

The situation with equal uncertainties is also helpful to find out if the classical random effects model ($I_0 = \emptyset$) can have a higher likelihood than AREM.CL.K with a non-empty homogeneous data set. Define for any K

$$\mathcal{L}(I_0, \dots, I_K) = \min_{\tau_1^2, \dots, \tau_K^2} L(\tilde{\mu}, \tau_1^2, \dots, \tau_K^2; I_0, \dots, I_K). \tag{25}$$

Conjecture. If \mathcal{L} is defined by (25) and $K = 1$, then

$$\mathcal{L}(\emptyset, \{1, \dots, N\}) \geq \min_{I_0 \neq \emptyset} \mathcal{L}(I_0, I_0^c). \tag{26}$$

This conjecture holds in many practical examples. It is a feature of the model with lower bounded variances (Rukhin, 2019a). We confirm (26) here for the equal variances case for all $N \geq 3, K = 1$.

For $0 < \omega \leq \omega_0$, and $v \geq 1$ given in (24) we define $F_I(\omega; \alpha_I, \beta_I, \gamma_I)$ as the left hand side of (23),

$$F_I(\omega; \alpha_I, \beta_I, \gamma_I) = \omega_0 \log \left(\frac{\omega_0(1 - \omega)}{(1 - \omega_0)\omega} \right) + \frac{(1 - \omega_0)}{1 - \omega} [\alpha_I \omega + \beta_I(1 - \omega) + \gamma_I \omega(1 - \omega)] - 1 - \log(v). \tag{27}$$

The goal is to show that for some $M, 1 \leq M < N$ there is I such that $\min_{0 < \omega \leq \omega_0} F_I(\omega; \alpha_I, \beta_I, \gamma_I) < 0$. One has $F_I(\omega_0; \alpha_I, \beta_I, \gamma_I) = v - \log(v) - 1 > 0$ and $\lim_{\omega \rightarrow 0} F_I(\omega; \alpha_I, \beta_I, \gamma_I) = +\infty$. Thus the sought minimum cannot be attained close to the boundary.

For fixed I the desired minimizer ω satisfies the cubic equation

$$\alpha_I + \gamma_I(1 - \omega)^2 = \frac{\omega_0(1 - \omega)}{(1 - \omega_0)\omega}. \tag{28}$$

If (28) does not have a solution ω in the interval $(0, \omega_0)$, for which the second derivative is positive then $\min F_I(\omega; \alpha_I, \beta_I, \gamma_I) = 0$ and the homogeneity cluster I cannot have larger likelihood than \emptyset . The existence of a solution implies that $\alpha_I + \gamma_I(1 - \omega_0)^2 > 1$.

One can express α_I through ω and γ_I to get the inequality in terms of these two variables

$$\omega_0 \log \left(\frac{\omega_0(1 - \omega)}{(1 - \omega_0)\omega} \right) + (1 - (\omega_0)\gamma_I\omega^2 - 1) - \log(v) \leq 0, \tag{29}$$

which for fixed ω allows to find the region $\alpha_I, \beta_I, \gamma_I$, where (27) holds.

Indeed it suffices to look at the simplest situation when $M = N - 1$ in which case $b_I = 0$. If $\min_{\omega} \sum_I F_I(\omega, v) \leq 0$ then $\sum_I \min_{\omega} F_I(\omega, v) \leq 0$ so that for some $I, F_I(\omega, v) \leq 0$. To evaluate $\min_{\omega} \sum_I F_I(\omega, v)$ averages of α_I and γ_I are needed. These quantities can be obtained by summing over all M -element sets I ,

$$\omega_0 \sum_I \alpha_I = \binom{N}{M} \frac{(M - 1)v}{(N - 1)},$$

$$(1 - \omega_0) \sum_I \beta_I = \binom{N}{M} \frac{(N - M - 1)v}{(N - 1)},$$

which vanishes when $M = N - 1$, and

$$\omega_0(1 - \omega_0) \sum_I \gamma_I = \binom{N}{M} \frac{v}{(N - 1)}.$$

Thus for $M = N - 1$ the inequality $\min_{\omega} \sum_I F_I(\omega, \alpha_I, 0, \gamma_I) < 0$ means that (27) holds with $\gamma_I = v/[(N - 1)\omega_0(1 - \omega_0)] = v[N/(N - 1)]^2$, $\alpha_I = (N - 2)v/[(N - 1)\omega_0] = vN(N - 2)/(N - 1)^2$. The direct calculation given in Section 9.4 implies that this inequality holds for all $N \geq 3$. Indeed the region where (29) does not hold is convex and $\alpha_I^0 = N(N - 2)/(N - 1)^2$ and $\gamma_I^0 = [N/(N - 1)]^2$ provide its

extreme point. Thus if $\alpha_I - \alpha_I^0 \geq \lambda_N(\gamma_I - \gamma_I^0)$, (29) holds. Here λ_N denotes the slope of the boundary at (α_I^0, γ_I^0) . It is shown in the Appendix that $\lambda_N = (N^2 - 2N - 2 + (N - 1)\sqrt{N^2 - N - 2})[N(N + 2)]^{-1}$. Thus for $N \geq 3$, $\max_I \alpha_I - \alpha_I^0 \geq \lambda_N(\min_I \gamma_I - \gamma_I^0)$, and one can determine the optimal cluster I from the condition $I^c = \{\arg \min_k \gamma_k\}$.

The counterpart of (27) for the restricted likelihood (up to a constant term) is

$$\frac{N - 1}{N} \left[\left(\frac{Nv}{N - 1} \right) \log \left(\frac{Nv}{N - 1} \right) - 1 \right] \leq \frac{1}{N} \log \left(\frac{1 - \omega}{1 - \omega_0} \right) + \omega_0 \log \left(\frac{(1 - \omega_0)\omega}{\omega_0(1 - \omega)} \right) + \frac{(\omega_0 - \omega)(N\omega_0 + \omega)}{N\omega} - (\omega_0 - \omega)^2 c_I,$$

where

$$\frac{\alpha_I}{1 - \omega} + (1 - \omega)\gamma_I = \frac{\omega_0}{(1 - \omega_0)\omega} + \frac{1}{N(1 - \omega_0)}.$$

However this inequality typically does not hold. The simplest example is given by a centered data set: $N = 3, x_1 - \bar{x} = -5/(2\sqrt{3}), x_2 - \bar{x} = 1/(2\sqrt{3}), x_3 - \bar{x} = 2/\sqrt{3}, s^2 = 1$. Then the cluster $\{1, 2, 3\}$ provides the restricted likelihood solution with $\hat{\tau}^2 = 7/4$. In contrast the maximum likelihood method chooses heterogeneity cluster $I_1 = \{2, 3\}$ with a smaller variance estimate $\hat{\tau}^2 = 0.73$.

Now we formulate the main results of this section.

Theorem 4.1. *When $s_i^2 \equiv s^2$, for the optimal cluster I , $\omega = M(\tau^2 + s^2)^{-1}/[(N - M)s^{-2} + M(\tau^2 + s^2)^{-1}]$ solves cubic equation (28). For all $N \geq 3$, (26) holds when $M = N - 1$, with $I^c = \{i\}, i = \arg \min_k \gamma_k$. For the restricted likelihood function the corresponding inequality does not hold.*

The estimates of τ^2 obtained for $I_0 \neq \emptyset$ minimizing $\mathcal{L}(I_0, I_0^c)$ are typically (much) larger than those derived from AREM. $CL.K = 1$. Indeed the inequality (27) cannot be true for $\omega_1 = \omega_0/[\omega_0 + (1 - \omega_0)v] < \omega_0$, as $\omega_0(1 - \omega_1)/[(1 - \omega_0)\omega_1] = v$. Thus the minimizer ω in (27) cannot exceed ω_1 which corresponds to $\hat{\tau}^2$.

In the general case of arbitrary uncertainties one cannot restrict attention to the case $M = N - 1$. As a matter of fact the cluster leading to a larger likelihood value can correspond to any $M, M = 1, \dots, N - 1$. According to Theorem 3.1 all elements of the heterogeneity class must be far away from the consensus mean estimate.

5. MAXIMUM LIKELIHOOD DEGREE

The goal here is to give the formula for the degree of (polynomial) likelihood equations (11) representing the algebraic complexity of the problem. In the situation of the previous section this degree is three. Our formula is based on a different representation of the likelihood function (2),

$$L(\tilde{\mu}, \tau_1^2, \dots, \tau_K^2) = \sum_{0 \leq k, \ell \leq K} \sum_{i \in I_k, j \in I_\ell} \frac{(x_i - x_j)^2}{(\tau_k^2 + s_i^2)(\tau_\ell^2 + s_j^2)} \left[\sum_k \sum_{I_k} \frac{1}{\tau_k^2 + s_i^2} \right]^{-1} + \log \left(\prod_{k=0}^K \prod_{I_k} (\tau_k^2 + s_i^2) \right),$$

which follows from the Lagrange identity. It will be assumed that all x 's in each cluster and all s 's are distinct (i.e., that the data is *generic*). This is a usual condition imposed when studying the maximum likelihood degree (cf. Gross et al., 2012) for a homogeneous variance component problem.

To represent the solutions of likelihood equations as those of polynomial equations let

$$P = P(\tau_1^2, \dots, \tau_K^2; I_0, I_1, \dots, I_K) = \prod_{k=1}^K \prod_{i \in I_k} (\tau_k^2 + s_i^2)$$

be a polynomial of degree $n_1 + \dots + n_K = N - n_0 > 0$. Then with $S_0 = \sum_{j \in I_0} s_j^{-2}$,

$$\sum_{I_0} \frac{1}{s_j^2} + \sum_{k=1}^K \sum_{I_k} \frac{1}{\tau_k^2 + s_i^2} = S_0 + \sum_{k=1}^K \frac{\partial \log P}{\partial \tau_k^2} = S_0 + \sum_{k=1}^K \frac{P'_k}{P}.$$

We also define two polynomials

$$Q = \sum_{0 \leq k, \ell \leq K} \sum_{i \in I_k, j \in I_\ell} \frac{(x_i - x_j)^2 P(\tau_1^2, \dots, \tau_K^2)}{(\tau_k^2 + s_i^2)(\tau_\ell^2 + s_j^2)},$$

$\tau_0^2 = 0$, and

$$R = S_0 P + \sum_{k=1}^K P'_k.$$

The degree of Q is $N - \max(n_0, 2)$, that of R is $N - \max(n_0, 1)$.

The scaled likelihood function L from (2) takes the form

$$L = \frac{Q}{R} + \log P + \sum_{j \in I_0} \log s_j^2,$$

and the likelihood equations for $\tau_\ell^2, \ell = 1, \dots, K$ become

$$Q'_\ell P R - Q P R'_\ell + P'_\ell R^2 = 0. \tag{30}$$

For the restricted likelihood function $\tilde{L} = Q/R + \log R$, these equations are simpler

$$Q'_\ell R - Q R'_\ell + R R'_\ell = 0. \tag{31}$$

Combining these facts with information about the degree of P, Q , and R , one gets the next result.

Theorem 5.1. *The degree DL of the polynomial Eqs. (30) has the form*

$$DL = \begin{cases} 3N - 3n_0 - 1 & n_0 \geq 1 \\ 3N - 3 & n_0 = 0, \end{cases}$$

the degree RL of the Eqs. (31) is

$$RL = \begin{cases} 2N - 2n_0 - 1 & n_0 \geq 1 \\ 2N - 3 & n_0 = 0. \end{cases}$$

6. TWO PRACTICAL EXAMPLES

In many applications the smallest reported values s_k^2 often seem to belong to the outlying cases. Two examples are provided in this section; many more can be found in metrology literature.

6.1. Key Comparison CCL-K1

Length key comparison CCL-K1 was carried out to compare the deviations from nominal length of steel and tungsten carbide gauge blocks (Thalmann, 2002). The measurement results for one of these blocks, namely, the deviations from a tungsten carbide gauge blocks of nominal length 1mm by eleven participating national institutes are given in Table 1.

When $K = 1$, according to both maximum likelihood estimators $\hat{I}_1 = \tilde{I}_1 = \{6, 7, 8\}$ increasing the uncertainties of these three labs to 23.61, 23.95, and 24.27 nm (or to 23.81, 24.12, and 24.47 nm) with $\hat{\tau} = 22.75$ nm, $\hat{\mu} = 18.07$ nm. For the restricted maximum likelihood estimator $\tilde{\tau} = 22.76$ nm, $\tilde{\mu} = 18.09$ nm.

Cox (2007) analyzed the same data set with the conclusion that two laboratories 6 and 7 are to be excluded from the largest consistent subset which he defined to be formed by the studies for which

Table 1. CCL-K1 data on deviations from nominal length of a gauge block (tungsten carbide) in nm units

i	1	2	3	4	5	6	7	8	9	10	11
x_i	15	15	30	18	24	-9	-9	33	12.5	8.8	21
s_i	9	14	10	13	9	7	8	9	8.6	10	5.4

Table 2. Relative air speed data (nominal speed 2 m/s)

x_i	0.9993	1.0090	1.0196	1.0270
s_i	0.0026	0.0027	0.0017	0.0025

$\sum_j (x_j - \tilde{\mu})^2 / s_j^2$ does not exceed the critical point of χ^2 -distribution for 0.05 significance level. This approach which gives the consensus deviation as 20.3 nm has been criticized by Toman and Possolo, 2009 and Elster and Toman, 2010.

The maximum of the likelihood function is not attained at $I_1 = \{6, 7\}$, so that a better definition of the largest consistent subset is $\hat{I}_0 = \hat{I}_1^c = \{1, 2, 3, 4, 5, 9, 10, 11\}$. The same homogeneity cluster persists when $K = 2$. The estimate $\hat{\tau}_1 = 22.55$ nm is then replaced by $\hat{\tau}_1 = 11.07$ nm and $\hat{\tau}_2 = 26.70$ nm. The classical maximum likelihood estimator of τ based on the cluster \hat{I}_0 above is positive, but small: 0.056 nm; the commonly used DerSimonian–Laird estimator of this parameter vanishes so that \hat{I}_0 looks homogeneous.

Larger values of K do not lead to substantial gains in the likelihood functions. The best (classical likelihood) choice for $K = 3$ recommends to remove from \hat{I}_0 lab 3, i.e., to take $\{1, 2, 4, 5, 9, 10, 11\}$ as a new consistent subset; that for the restricted likelihood is $\{1, 2, 3, 4, 5, 9, 11\}$.

6.2. CIPM CCQM. FF-K3 Study

Another example of discrepant data is the international fluid flow comparisons of air speed measurement (CCM.FF-K3) (Terao et al., 2007). An ultrasonic anemometer chosen as a transfer standard was circulated between four national metrology institutes who reported calibration results at certain speeds. The (dimensionless) data given in the Table 2 represents the ratio of the laboratory's reference air speed to the one measured by the transfer standard.

Since the data is aberrant, the organizers of this study decided to use the median as an estimator of μ . When $K = 1$, the maximum likelihood estimator is $\hat{I}_0 = \{3\}$, $\hat{I}_1 = \{1, 2, 4\}$, $\hat{\mu} = 1.0193$ while the restricted maximum likelihood estimator gives $\tilde{I}_0 = \emptyset$, $\tilde{I}_1 = \{1, 2, 3, 4\}$, $\tilde{\mu} = 1.0138$. All increased uncertainties are about 0.0137 (MLE) ($\hat{\tau}^2 = 1.81 \times 10^{-4}$) or 0.0118 (REML) ($\tilde{\tau}^2 = 1.49 \times 10^{-4}$), respectively. If $K = 2$, the maximum likelihood estimator practically remains the same, $\hat{\mu} = 1.020$, but the negative likelihood (2) is -36.35 (attained at $\hat{I}_0 = \emptyset$, $\hat{I}_1 = \{1, 2\}$, $\hat{I}_2 = \{3, 4\}$) while the corresponding value is -35.45 for $K = 1$.

For the restricted maximum likelihood estimator $\tilde{I}_0 = \{3\}$, $\tilde{I}_1 = \{2\}$, $\tilde{I}_2 = \{1, 4\}$, $\tilde{\mu} = 1.020$, $\tilde{\tau}_1 = 0.0199$, $\tilde{\tau}_2 = 0.0087$, with the first increased uncertainty larger than that for $K = 1$, but the second one smaller. The negative restricted likelihood (3) decreases from -22.89 to -23.52 . For $K = 2$ both μ -estimators evaluated according to different likelihood methods turn out to be practically equal to 1.0196 (larger than the median 1.0143).

Table 3 summarizes the results. It indicates that from the view of information criteria AREM with $K = 2$ provides the best fit to the data. Then maximum likelihood estimators coincide, and $\hat{I}_0 = \{3\}$, $\hat{I}_1 = \{2\}$, $\hat{I}_2 = \{1, 4\}$. The value $K = 3$ gives only small gains in the likelihood.

Table 3. Likelihoods of different AREM models and information criterion numbers

Model	FEM	REM	AREM.K = 1	AREM.K = 2	AREM.K = 3
dimension	1	2	2	3	4
CL	23.12	-32.47	-35.45	-36.35	-36.45
RL	36.70	-22.10	-22.89	-23.52	-23.62
$AIC_{t,extrm}AREM.CL$	25.12	-28.47	-31.45	-30.35	-28.45
$AIC_{t,extrm}AREM.RL$	38.70	-18.10	-18.89	-17.52	-15.62
$BIC_{t,extrm}AREM.CL$	24.51	-29.69	-32.68	-32.19	-30.91
$BIC_{t,extrm}AREM.RL$	38.10	-19.33	-20.11	-19.37	-18.07

7. SIMULATION RESULTS FOR INFORMATION QUANTITIES

The quantities (25) provide statistics for Akaike's information criterion (AIC) or for the Bayesian information criterion (BIC) based on the classical or restricted likelihood (Claeskens and Hjort, 2008). Thus one can get the information criteria numbers for these models, e.g.,

$$AIC_{AREM.CL} = \mathcal{L}(I_0, \dots, I_K) + 2(K + 1)$$

and

$$BIC_{AREM.CL} = \mathcal{L}(I_0, \dots, I_K) + (K + 1) \log N.$$

In the corresponding formulas for $AIC_{AREM.RL}$ and $BIC_{AREM.RL}$ through $\tilde{\mathcal{L}}(I_0, \dots, I_K)$ one has to replace in the second term $K + 1$ by K . These numbers are employed in Section 6 in two practical examples. Here we compare them numerically for $AREM.K = 1$ and $AREM.K = N - n_0$.

To compare the properties of the considered likelihood procedures we performed Monte Carlo simulations involving AREM when $I_0^c = \{1, \dots, M\}$ with $M = 1$ or 2 , $N = 5$, $K = 1$. In these models $x_i \sim N(0, \tau^2 + s_i^2)$, $i = 1, \dots, M$ and $x_j \sim N(0, s_j^2)$ for $j \geq M + 1$. The variances s_k^2 were obtained from realizations of standard exponential random variables for 50 000 Monte Carlo runs.

Figure 1 displays the percentage of correct identification by information criterions $AIC_{AREM.CL}$ and $AIC_{AREM.RL}$. The Bayes information criterions $BIC_{AREM.CL}$ and $BIC_{AREM.RL}$ are not reported as their values differ only by a constant.

This quantity ranges from 0.12 ($\tau^2 = 0$) to 0.31 ($\tau^2 = 3$) for AREM.CL when $M = 1$ which is reasonable in view of the total number of models ($2^N - 1 = 31$). When $M = 2$ it grows from 0.02 to 0.17. The probabilities of the correct choice behave similarly for $K = N - n_0$ but are somewhat smaller e.g., when $M = 1$ they increase from 0.09 to 0.21, and from 0.02 to 0.16 ($M = 2$).

Neither classical nor restricted likelihood procedures perform well in the case of the traditional REM ($M = N$, $K = 1$) favoring AREM with smaller M .

8. DISCUSSION

This work proposes a class of models to handle discrepant heterogeneous data in research synthesis. The traditional random effects model is extended to allow for different heterogeneity variance values. The algorithms for determination of these values along with the convergence conditions are presented. The procedures are based on the Gaussian likelihood although this distribution can be replaced by a non-normal location/scale parameter family. However the underlying density cannot be well estimated because usually the data is too scarce.

The suggested approach prescribes additional error to some outlying studies whose summary results are given larger uncertainties but still enter the final answer. Typically the uncertainty enlargements apply only to a few cases. The straightforward iterative numerical algorithms to evaluate the maximum likelihood estimators can be employed for small/moderate number of studies. If there are many studies,

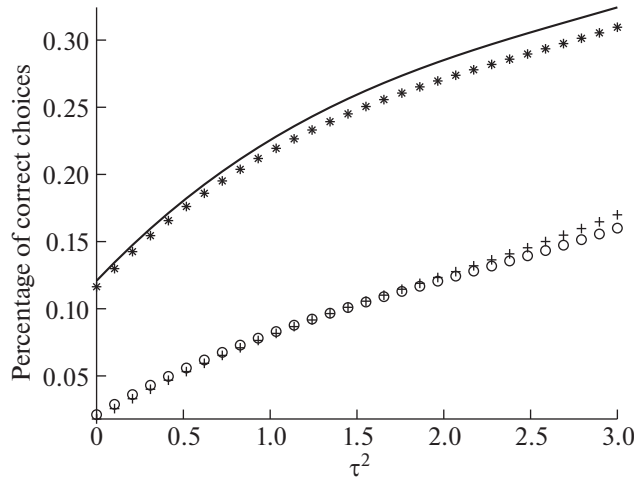


Fig. 1. The percentage of correct model identification as a function of τ^2 for AIC based on classical likelihood ($M = 1$, solid line; $M = 2$, line marked by +) and on restricted likelihood ($M = 1$, line marked by *; $M = 2$, line marked by circles).

the likelihood calculations become impractical but the present methodology still can be used provided that the sizes n_k of heterogeneity clusters $I_k, k = 1, \dots, K$ are small.

We do not recommend more than two clusters unless there are good practical reasons to believe in so many categories. Then the $K = N - n_0$ model can be implemented. The largest consistent subset obtained when $K = 1$ or 2 as a rule gets smaller if K increases while the chance of the correct model identification diminishes.

Appendix

PROOF OF THEOREM 2.1

Proof. Since for $\tilde{\mu}$ defined by (4)

$$\frac{\partial \tilde{\mu}}{\partial \tau_k^2} = - \sum_{I_k} \frac{x_i - \tilde{\mu}}{(\tau_k^2 + s_i^2)^2} \left(\sum_m \sum_{I_m} \frac{1}{\tau_m^2 + s_i^2} \right)^{-1},$$

the diagonal elements h_{kk} of the Hessian are $\partial^2 L / \partial \tau_k^4 = d_k - a_k^2$, its off-diagonal elements have the form: $h_{k\ell} = \partial^2 L / [\partial \tau_k^2 \partial \tau_\ell^2] = -a_k a_\ell, k \neq \ell$. Thus (6) holds, and the matrix

$$\hat{H} = \text{diag}(\hat{d}) - \hat{a}\hat{a}^T$$

is positive definite if and only if (7) is valid.

The vector function $\psi = (\psi_1, \dots, \psi_K)$ of $\tau_1^2, \dots, \tau_K^2$, with

$$\psi_k = \sum_{I_k} \frac{(x_i - \mu)^2 - s_i^2}{(\tau_k^2 + s_i^2)^2} \left[\sum_{I_k} \frac{1}{(\tau_k^2 + s_i^2)^2} \right]^{-1} = \tau_k^2 - \left[\sum_{I_k} \frac{1}{(\tau_k^2 + s_i^2)^2} \right]^{-1} \frac{\partial L}{\partial \tau_k^2}$$

defines iterations (9): $\psi(\tau_1^2, \dots, \tau_K^2) = (\tau_1^2, \dots, \tau_K^2)$. In these iterations $\mu = \tilde{\mu}(\tau_1^2, \dots, \tau_K^2)$, as in (4). The conditions (7) and (8) mean that the spectral radius of the Jacobian ψ' evaluated at $\hat{\tau}_1^2, \dots, \hat{\tau}_K^2$ is smaller than 1 which is a sufficient condition for the convergence of iteration process (9) defined by ψ (cf. Ortega and Rheinboldt, 2000). \square

PROOF OF THEOREM 2.2

Proof. According to (5) one has for stationary points $(x_k - \hat{\mu})^2 = \hat{\tau}_k^2 + s_k^2$. It follows that $\hat{d}_k = (\hat{\tau}_k^2 + s_k^2)^{-2}$ and $\hat{a}_k^2/\hat{d}_k = 2(\hat{\tau}_k^2 + s_k^2)^{-1}[\sum_\ell(\hat{\tau}_\ell^2 + s_\ell^2)^{-2} + \sum_j s_j^{-2}]^{-1}$. Therefore condition (7) in this situation means that (12) is valid.

The (scalar) iteration function is now given by (14) and its (positive) derivative at $\hat{\mu}$,

$$\left[\sum_k \frac{2}{\hat{\tau}_k^2 + s_k^2} \right] \left[\sum_k \frac{1}{\hat{\tau}_k^2 + s_k^2} + \sum_j \frac{1}{s_j^2} \right]^{-1}$$

is smaller than 1 if and only if the first inequality in (12) holds. □

PROOF OF THEOREM 2.3 AND COMMENTS

Proof. Now the elements of the Hessian \tilde{H} have the form

$$\tilde{h}_{k\ell} = \delta_{k\ell}\tilde{c}_\ell - \tilde{a}_k\tilde{a}_\ell - \tilde{b}_k\tilde{b}_\ell, \quad k, \ell = 1, \dots, K,$$

where $\delta_{k\ell}$ is the Kronecker symbol. Thus (16) holds. Provided that $\tilde{c}_k > 0$, \tilde{H} is positive definite if and only if the similar matrix $I - C^{-1/2}\tilde{a}\tilde{a}^T C^{-1/2} - C^{-1/2}\tilde{b}\tilde{b}^T C^{-1/2}$, is. The two eigenvalues λ of this matrix, which are different from 1, solve the equation

$$(1 - \lambda)^2 - (\tilde{a}^T C^{-1}\tilde{a} + \tilde{b}^T C^{-1}\tilde{b})(1 - \lambda) + (\tilde{a}^T C^{-1}\tilde{a})(\tilde{b}^T C^{-1}\tilde{b}) - (\tilde{a}^T C^{-1}\tilde{b})^2 = 0,$$

and they are positive when and only when (18) holds. □

For AREM.RL. $K = 1$ the restricted maximum likelihood estimator of τ^2 satisfies (15) and the inequality $\tilde{a}_1^2 + \tilde{b}_1^2 < \tilde{c}_1$ which implies (18).

For AREM.RL. $K = N - n_0$ all stationary points $z_k = z_k(\mu, \zeta) = [(x_k - \mu)^2 + 1/(\zeta + \sum s_j^{-2})]^{-1}$ depend only on two variables μ and $\zeta = \sum(\tilde{\tau}_k^2 + s_k^2)^{-1}$. The two-dimensional iteration function $\Psi = (\mu, \zeta)$ in these variables,

$$\mu = \frac{\sum_k x_k z_k + \sum_j x_j s_j^{-2}}{\sum_k z_k + \sum_j s_j^{-2}}, \quad \zeta = \sum z_k,$$

defines an iteration method which converges under conditions (17) and (18). Indeed in this case $\tilde{c}_k = (\tilde{\tau}_k^2 + s_k^2)^{-2}$, and for 2×2 Jacobian Ψ' evaluated at $\tilde{\tau}_1^2, \dots, \tilde{\tau}_K^2$ one has

$$\text{tr}(\Psi') = \sum_k \frac{\tilde{a}_k^2 + \tilde{b}_k^2}{\tilde{c}_k},$$

$$\det(\Psi') = \sum_k \frac{\tilde{a}_k^2}{\tilde{c}_k} \sum_k \frac{\tilde{b}_k^2}{\tilde{c}_k} - \left(\sum_k \frac{\tilde{a}_k \tilde{b}_k}{\tilde{c}_k} \right)^2.$$

Therefore (17) implies that (positive) diagonal elements of Ψ' cannot exceed 1 while (18) means that $\text{tr}(\Psi') < 1 + \det(\Psi')$. Then the spectral radius of Ψ' is smaller than 1.

PROOF OF THEOREM 4.1

Proof. For fixed $N, M = N - 1, I$ let $\Phi(\alpha, \gamma) = F_I(\omega, \alpha, 0, \gamma)$ where $\omega = \omega(\alpha, gl)$ is the solution of (28). We start with the derivatives

$$\omega'_\alpha = \frac{\omega^2}{2\gamma(1 - \omega)\omega^2 - \omega_0((1 - \omega_0)^{-1}),}$$

$$\omega'_\gamma = (1 - \omega)^2 \omega'_\alpha.$$

Since $\partial F_I(\omega, \alpha, 0, \gamma)/\partial \omega = 0$, one has

$$\begin{aligned}\Phi'_\alpha &= \frac{\partial \Phi}{\partial \alpha} = \frac{(1 - \omega_0)\omega}{1 - \omega} - \frac{\omega_0}{v}, \\ \Phi'_\gamma &= (1 - \omega_0)\omega - \frac{\omega_0(1 - \omega_0)}{v}.\end{aligned}$$

The second derivatives depend on $\omega'_\alpha, \omega'_\gamma$,

$$\begin{aligned}\Phi''_{\alpha\alpha} &= \frac{(1 - \omega_0)\omega'_\alpha}{(1 - \omega)^2} + \frac{\omega_0}{v^2}, \\ \Phi''_{\alpha\gamma} &= (1 - \omega_0)\omega'_\alpha + \frac{\omega_0^2(1 - \omega_0)}{v^2}, \\ \Phi''_{\gamma\gamma} &= (1 - \omega_0)(1 - \omega)^2\omega'_\alpha + \frac{\omega_0^2(1 - \omega_0)^2}{v^2}.\end{aligned}$$

If the function $\phi(\gamma)$ is defined by the equation $\Phi(\phi(\gamma), \gamma) = 0$ then

$$\phi'(\gamma) = -\frac{\Phi'_\gamma}{\Phi'_\alpha},$$

and $\phi'(\gamma)$ vanishes when $\omega = \omega_0/v$. Thus $\phi(\gamma)$ increases when $\gamma \leq \tilde{\gamma}$ and then (sharply) decreases until it vanishes at $\tilde{\gamma}$ such that

$$\omega_0 \log \left(\frac{(1 - \omega_0)\omega}{\omega_0(1 - \omega)} \right) + (1 - \omega_0)(\tilde{\gamma}\omega^2 - 1) = \log(\omega_0(1 - \omega_0)\tilde{\gamma}),$$

where

$$\omega(1 - \omega) = \frac{\omega_0}{(1 - \omega_0)\tilde{\gamma}}.$$

(As N increases $\tilde{\gamma}$ grows up very fast: $\log \tilde{\gamma} \sim N \log N$. This asymptotics also governs $\tilde{\gamma}$.)

The second derivative

$$\begin{aligned}\phi''(\gamma) &= -\frac{(\Phi'_\gamma)^2\Phi''_{\alpha\alpha} - 2\Phi'_\alpha\Phi'_\gamma\Phi''_{\alpha\gamma} + (\Phi'_\alpha)^2\Phi''_{\gamma\gamma}}{(\Phi'_\alpha)^3} \\ &= -\frac{1}{(\Phi'_\alpha)^3} \left[\frac{(1 - \omega_0)\omega'_\alpha}{(1 - \omega)^2} \left((1 - \omega_0)\omega - \frac{\omega_0[1 - \omega_0 - (1 - \omega)^2]}{v^2} \right)^2 \right. \\ &\quad \left. + \frac{\omega_0^2(\omega_0 - \omega)^2\omega^2(1 - \omega)^2}{v^2} \right]\end{aligned}$$

is negative when $\gamma^0 = (N/(N - 1))^2 \leq \gamma \leq \tilde{\gamma}$. Because of concavity of function $\phi(\gamma)$ the region where $\Phi(\alpha, \gamma) \geq 0, \gamma^0 \leq \gamma \leq \tilde{\gamma}$ is convex.

It remains to evaluate the Hessian of Φ at (α^0, γ^0) ,

$$\begin{aligned}\Phi''_{\alpha\alpha} &= -\frac{\omega_0^2(N + 2)}{(N + 1)(N - 2)}, \\ \Phi''_{\alpha\gamma} &= \frac{\omega_0^2(N^2 - 2N - 2)}{(N + 1)N(N - 2)}, \\ \Phi''_{\gamma\gamma} &= \frac{\omega_0^2(N^2 - N - 3)}{(N + 1)N^2(N - 2)},\end{aligned}$$

which leads to the formula for the slope

$$\lambda_N = \frac{(N^2 - 2N - 2 + (N - 1)\sqrt{N^2 - N - 2})}{N(N + 2)}.$$

□

REFERENCES

1. M. Borenstein, L. Hedges, J. Higgins, and H. Rothstein, *Introduction to Meta-Analysis* (Wiley, New York 2009).
2. G. Claeskens and N. Hjort, *Model Selection and Model Averaging* (Cambridge University Press, Cambridge, UK 2008).
3. M. G. Cox “The evaluation of key comparison data: determining the largest consistent subset,” *Metrologia* **44**, 187–200 (2007)
4. C. Elster and B. Toman, “Analysis of key comparisons: estimating laboratories biases by a fixed effects model using Bayes model averaging,” *Metrologia* **47**, 113–119 (2010).
5. R. L. Graham, D. E. Knuth, and D. Patashnik, *Concrete Mathematics: A Foundation for Computer Science* (2nd ed. Addison Wesley, Reading, MA 1994).
6. E. Gross, M. Drton, and S. Petrovic, “Maximum likelihood degree of variance component models,” *EJS* **6**, 983–1016 (2012).
7. A. B. Haidich, “Meta-analysis in medical research,” *Hippokratia* **14** (Suppl. 1), 29–37 (2010).
8. J. Higgins, S. Thompson, and D. J. Spiegelhalter, “A re-evaluation of random effects meta-analysis,” *J. Royal Statist. Soc. Ser A.* **172**, 137–159 (2009).
9. D. C. Hoaglin, “Misunderstandings about Q and ‘Cochran’s Q test’ in meta-analysis,” *Statist. Med.* **35**, 485–495 (2016).
10. E. Kulinskaya and I. Olkin, “An overdispersion model in meta-analysis,” *Statist. Model.* **14**, 49–76 (2014).
11. D. Langan, J. Higgins, D. Jackson, J. Bowden, A. A. Veroniki, W. Kontopantelis, W. Viechtbauer and M. Simmonds “A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses,” *Res. Synth. Meth.* **10**, 83–98 (2019).
12. K. Lee and S. Thompson “Flexible parametric models for random-effects distributions,” *Statist. Med.* **27**, 418–34 (2008).
13. D. I. Ohlssen, L. D. Sharpless, and D. J. Spiegelhalter, “Flexible random-effects models using Bayesian semi-parametric models: Applications to institutional comparisons,” *Statist. Med.* **26**, 2088–112 (2007).
14. J. M. Ortega and W. C. Rheinboldt, *Iterative Solutions of Nonlinear Equations in Several Variables* (SIAM, Philadelphia 2000). <https://doi.org/10.1137/1.9780898719468>
15. A. L. Rukhin, “Maximum likelihood and restricted maximum likelihood solutions in multiple-method studies,” *J. Res. Nat. Inst. Stand. Techn.* **116**, 539–56 (2011).
16. A. L. Rukhin, “ Estimating heterogeneity variances to select a random effects model,” *J. Statist. Plan. Inf.* **202**, 1–13 (2019a).
17. A. L. Rukhin, “ Homogeneous data clusters in interlaboratory studies,” *Metrologia* **56**, 874–882 (2019b).
18. Y. Terao, M. van der Beek, T. T. Yeh, and H. Müller, “Final report on the CIPM air speed key comparison (CCM.FF-K3),” *Metrologia* **44**, Tech. Supp. 07009 (2007).
19. R. Thalmann, “CCL key comparison: calibration of gauge blocks by interferometry,” *Metrologia* **39**, 165–177 (2002).
20. S. Thompson and S. Sharp, “Explaining heterogeneity in meta-analysis: a comparison of methods,” *Statist. Med.* **18**, 2693–708 (1999).
21. B. Toman and A. Possolo, “Laboratory effects models for interlaboratory comparisons,” *Accred. Qual. Assur.* **14**, 553–563 (2009).