

Simulate effect sizes

Han Bossier

11/3/2019

Introduction

In this report, we look at estimators for the population effect at the meta-analysis level. We generate effect sizes at the study level and then fit a random effects model. We estimate the between study heterogeneity using the DerSimonian & Laird estimator, the Hedges' estimator and Restricted Maximum Likelihood. These estimates are used to estimate the population effect. We will increase the amount of between-study heterogeneity to see its effect on the estimates of the population effect.

Data generating model

Consider an observed (standardized) effect Y_i for study $i = 1, \dots, K$. We assume the following model:

$$Y_i = X\beta + u_i + \varepsilon_i,$$

where

$$u_i \sim N(0, \sigma_*^2) \quad \varepsilon_i \sim N(0, \sigma_i^2).$$

Here β represents the population effect, X the design matrix at the meta-analysis level, u a random effect and ε a within-study error. Note that X will be a vector of length K with 1's.

Model parameters

We set $\beta = 0.9$ and let the between-study variance (σ_*^2) increase in 4 Monte-Carlo simulation settings. More particularly, we have: $\sigma_*^2 \in \{0, 1, 4, 9\}$. We set the number of studies K to 500.

```
# True ES
Tes <- 0.9
# Number of studies
nstud <- 500
# X vector of 1's
X <- matrix(1, nrow = nstud)
# Between study variability
sigm2U <- c(0, 1, 4, 9)
```

To select the within-study variance components of each study, we do the following. First we sample a number of subjects for each study. This is a number between 100 and 300. Using this sample size and the true value for β (the population effect size), we calculate the true within-study variance for each study (i.e. σ_i^2).

```
# Sample number of subjects
nsubS <- sample(x = 100:300, size = nstud, replace = TRUE)
# Alternative: each study the same sample size
#nsubS <- rep(30, nstud)
# Calculate true value for the variance
sigm2E <- NeuRRoStat::varHedgeT(g = Tes, N = nsubS)
```

Note that we have a large number of subjects and studies which is needed to see if we have unbiased estimates.

Model estimation

We use the **R** package `metafor` to fit the random-effects models. To estimate σ_*^2 , we will use the DerSimonian & Laird estimator, the Hedges' estimator and a Restricted Maximum Likelihood approach. Note that the model fitting procedure uses weights where these are the inverse of the sum of the within- and estimated between-study variability. Furthermore, we use the `NeuRroStat` package to estimate σ_*^2 with the DerSimonian & Laird estimator. However, I checked its result and it is equal to the one using the `metafor` package.

Monte-Carlo

Settings

We set the number of simulations to 10.00.

```
# Number of simulations
nsim <- 1000
```

We loop over the 4 possible values for the between-study heterogeneity. The following code in **R** is an example of one simulation iteration for one value for σ_*^2 (e.g. $j = 1$). We generate K effect sizes, then calculate the within-study variance (using the formula of Hedges) of each effect size.

```
# Generate data
Y <- X %*% Tes + rnorm(n = nstud, mean = 0, sd = sqrt(sigm2U[j])) +
  rnorm(n = nstud, mean = 0, sd = sqrt(sigm2E))

# Weights of each study
vi <- NeuRroStat::varHedgeT(g = Y, N = nsubS)
```

Execution

Here we run the simulation.

```
# Empty data frame
results <- data.frame() %>%
  as_tibble()

# For loop over the simulations
for(i in 1:nsim){
  # For loop over the amount of between-study heterogeneity
  for(j in 1:length(sigm2U)){
    # Generate data
    Y <- X %*% Tes + rnorm(n = nstud, mean = 0, sd = sqrt(sigm2U[j])) +
      rnorm(n = nstud, mean = 0, sd = sqrt(sigm2E))

    # Weights of each study
    vi <- NeuRroStat::varHedgeT(g = Y, N = nsubS)

    # Fit the model using rma + HE or REML estimator for tau2
    fitHE <- metafor::rma(yi = Y, vi = vi, method = 'HE')
    fitREML <- metafor::rma(yi = Y, vi = vi, method = 'REML')

    # Between study variability: DL estimator
    tau2_DL <- NeuRroStat::tau(Y = Y, W = (1/vi), k = length(Y))
    # HE estimator
```

```

tau2_HE <- c(fitHE$tau2)
# REML estimator
tau2_REML <- c(fitREML$tau2)

# Weights
W <- diag(c(1/(vi + tau2_DL)))

# Estimate for B using various estimators for tau
# First one = DerSimonian and Laird
B_DL <- EstB(X = X, W = W, Y = Y)
# controle: rma(yi = Y, vi = vi, method = 'DL') --> klopt
# One using Hedges
B_HE <- c(fitHE$beta)
# One using REML
B_REML <- c(fitREML$beta)

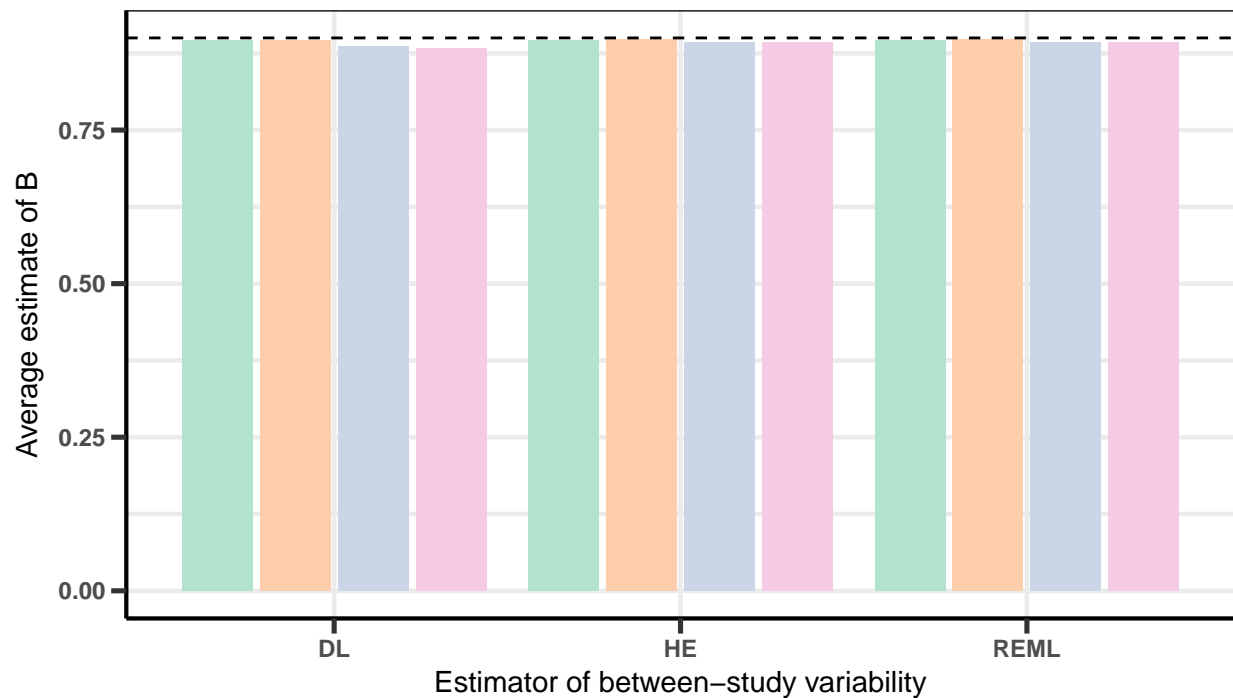
# Save in data frame
results <- data.frame('B' = c(B_DL, B_HE, B_REML),
  'TrueB' = Tes,
  'tau2' = c(tau2_DL, tau2_HE, tau2_REML),
  'TrueTau2' = sigm2U[j],
  'estimatorBS' = c('DL', 'HE', 'REML')
) %>% as_tibble() %>%
  bind_rows(results, .)
}
}

```

Results

Estimator for population effect

We will first plot the average (over all simulations) of the estimates for β . We do this for each value of between-study variability and each of the three estimators for the latter. The dashed line corresponds to the true value of β .



True between–study variance 0 1 4 9

Averaged over 1000 simulations. Dashed line = true value.

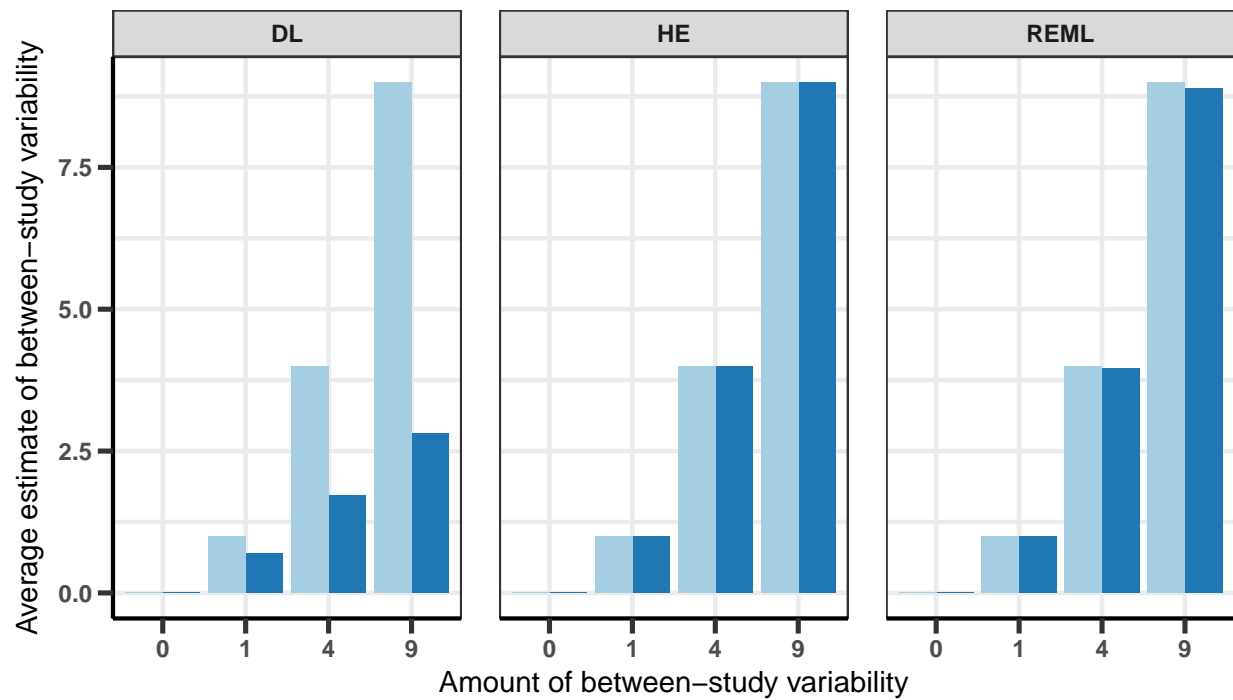
It seems to be off using DL!

```
results %>%
  group_by(estimatorBS, TrueTau2) %>%
  summarise(AvgEstB = mean(B),
            TrueB = mean(TrueB))
```

```
# A tibble: 12 x 4
# Groups:   estimatorBS [?]
  estimatorBS TrueTau2 AvgEstB TrueB
  <fct>         <dbl>   <dbl> <dbl>
1 DL           0      0.895  0.9
2 DL           1      0.895  0.9
3 DL           4      0.887  0.9
4 DL           9      0.883  0.9
5 HE           0      0.895  0.9
6 HE           1      0.898  0.9
7 HE           4      0.893  0.9
8 HE           9      0.893  0.9
9 REML         0      0.895  0.9
10 REML        1      0.898  0.9
11 REML        4      0.893  0.9
12 REML        9      0.893  0.9
```

Estimator for between-study variance

We repeat the same for the estimates of between-study variability.



True between-study variance True value Estimated value

Averaged over 1000 simulations.

In a table, we have:

```
results %>%
  dplyr::select(-B, -TrueB) %>%
  group_by(TrueTau2, estimatorBS) %>%
  summarise(AvgEstTau2 = mean(tau2))
```

```
# A tibble: 12 x 3
# Groups:   TrueTau2 [?]
  TrueTau2 estimatorBS AvgEstTau2
    <dbl> <fct>          <dbl>
1         0 DL          0.000170
2         0 HE          0.000192
3         0 REML        0.000175
4         1 DL          0.689
5         1 HE          1.00
6         1 REML        0.990
7         4 DL          1.72
8         4 HE          4.00
9         4 REML        3.95
10        9 DL          2.81
11        9 HE          8.99
12        9 REML        8.89
```