Code for Figure 2

August 2, 2016

1 Preliminaries

Load the necessary libraries, source the file with the R functions:

```
library(ggplot2)

source("functions.R")
```

Will use the same theme throughout, so just declare this variable:

```
themeUsed <- theme_bw(base_size = 20) +
  theme(axis.line = element_line(colour = "black"),
        plot.title = element_text(size = 15, hjust=0.5),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank(),
        legend.background = element_rect(fill="transparent"),
        legend.key = element_blank(),
        legend.text.align = 0,
        legend.position = c(0.15,0.28),
        axis.line.x = element_line(color="black", size = 0.5), ##this is to show axes - bug in this vers
        axis.line.y = element_line(color="black", size = 0.5))
```

2 Dataset for panel a)

These results are for the $I = 20, p \geq 2$ exchangeable case with equal within-study and between-study variances, with $S_i^2 \equiv S^2$ and $\rho_i = \frac{\rho(i-1)}{I}$ (within study). The parameters we vary are $\sigma^2_{\overline{S^2}}$ and p, as we take $\rho^{BS} = 0.5$ We save the relative efficiencies for only one of coefficients, as they are all equal.

```
##number of studies
I <- 20

##save results in data frame
##het is sigma^2/S^2
bigMat <- expand.grid(p = 2:20,
                      rhoBS=0.5,
                      het = c(0, 1/5, 1, 5),
                      RelEff = NA)
for(rr in 1:nrow(bigMat))
```

```
{
    p <- bigMat[rr, 1]
    rhoBS <- bigMat[rr, 2]
    het <- bigMat[rr, 3]

    ##index over the studies
    i <- 1:I
    ##within-study correlations
    rho.iW <- (i-1)/I
    ##overall correlations
    rho.i <- (rho.iW + rhoBS*het)/(1+het)

    ##calculate the two sums
    ##Sum1 is over 1/(1-rho_i)
    ##Sum2 is over 1/(1+(p-1)*rho_i)
    Sum1 <- sum(1/(1-rho.i))
    Sum2 <- sum(1/(1+(p-1)*rho.i))

    bigMat$RelEff[rr] <- I/p * (Sum1 + (p-1)*Sum2)/(Sum1*Sum2)
}


##make het a factor (for ggplot)
bigMat$het <- as.factor(bigMat$het)
```

3 Create plot for panel a)

4 Get the combined dataset for panel b)

Load the files representing the summary for each scenario with 0 heterogeneity and save the results in a single dataframe, Vars0:

```
allFiles <- list.files("simResultsComb")
##remove all the files with "Suppl" in title
allFiles <- allFiles[-grep("Suppl", allFiles)]

##save the three Vars for the different combinations
VarsAll <- expand.grid(I = c(5, 10, 15, 20),
                       p = 2:10,
                       corrBtw = c(0, 0.5),
                       varBtw = c(0, 1))

##only keep combinations of corrBtw=0 & varBtw=0, corrBtw=0.5 & varBtw=1
VarsAll <- VarsAll[(VarsAll$corrBtw == 0 & VarsAll$varBtw == 0) |
                   (VarsAll$corrBtw == 0.5 & VarsAll$varBtw == 1) ,]

VarsAll$known <- VarsAll$unknown <- VarsAll$univKnown <- VarsAll$univ <- NA

for(file in 1:length(allFiles))
{
    load(paste("simResultsComb",
               allFiles[file], sep="/"))
```

```
    rowNr <- gsub("combine_cost_of_estimation_", "", allFiles[file])
    rowNr <- gsub(".RData", "", rowNr)
    rowNr <- as.numeric(rowNr)

    VarsAll[rowNr, colnames(VarsAll)] <- Vars[rowNr, colnames(VarsAll)]
}
VarsAll$Ratio <- VarsAll$known/VarsAll$unknown
VarsAll$RelEff <- VarsAll$unknown/VarsAll$univ
VarsAll$RelEffT <- VarsAll$known/VarsAll$univKnown
range(VarsAll$Ratio)

## [1] 0.7774397 0.9941824

##only keep the ones with varBtw = 0 and I=20
Vars0 <- VarsAll[VarsAll$varBtw == 0 & VarsAll$I == 20 ,]
```

Load the files representing the summary for each scenario with non-0 heterogeneity and save the results in a single dataframe, VarsAll:

```
allFiles <- list.files("simResultsComb")
##keep only files with "Suppl" in title
allFiles <- allFiles[grep("Suppl", allFiles)]

##save the three Vars for the different combinations
VarsAll <- expand.grid(I = 20,
                       p = 2:10,
                       corrBtw = c(0, 0.5),
                       varBtw = c(1/5, 1, 5))

VarsAll$known <- VarsAll$unknown <- VarsAll$univKnown <- VarsAll$univ <- NA

for(file in 1:length(allFiles))
{
    load(paste("simResultsComb", allFiles[file], sep="/"))

    rowNr <- gsub("combine_cost_of_estimation_Suppl_", "", allFiles[file])
    rowNr <- gsub(".RData", "", rowNr)
    rowNr <- as.numeric(rowNr)

    VarsAll[rowNr, colnames(VarsAll)] <- Vars[rowNr, colnames(VarsAll)]
}
VarsAll$Ratio <- VarsAll$known/VarsAll$unknown
VarsAll$RelEff <- VarsAll$unknown/VarsAll$univ
VarsAll$RelEffT <- VarsAll$known/VarsAll$univKnown
range(VarsAll$Ratio)

## [1] 0.9300576 0.9993072
```

Separate out what is needed for Panel b):

```
Vars <- VarsAll[VarsAll$corrBtw == 0.5,]

##add in the 0 heterogeneity case
Vars <- rbind(Vars, Vars0)
```

```
RelEff.b <-
  rbind(cbind(as.matrix(Vars[,c("I", "p", "corrBtw", "varBtw", "RelEff"),]),"RelEff"),
        cbind(as.matrix(Vars[,c("I", "p", "corrBtw", "varBtw", "RelEffT"),]),"RelEffT"))
colnames(RelEff.b)[6] <- "Estimate"
RelEff.b <- as.data.frame(RelEff.b)
sapply(RelEff.b, class)

##         I         p   corrBtw    varBtw    RelEff Estimate
## "factor" "factor" "factor" "factor" "factor" "factor"

RelEff.b$p <- as.numeric(as.character(RelEff.b$p))
RelEff.b$corrBtw <- as.numeric(as.character(RelEff.b$corrBtw))
RelEff.b$varBtw <- as.numeric(as.character(RelEff.b$varBtw))
RelEff.b$RelEff <- as.numeric(as.character(RelEff.b$RelEff))
```

```
panelB <- ggplot(RelEff.b, aes(x=p, y=RelEff))+
  geom_point(size=3.0, aes(color=as.factor(varBtw), shape=as.factor(varBtw)))+
  geom_line(aes(linetype=Estimate,color=as.factor(varBtw), shape=as.factor(varBtw)))+
  themeUsed +
  ylab(expression(RelEff)) +
  ylim(limits=c(min(bigMat$RelEff)*0.8, 1)) +
  scale_color_discrete(name = "",
                       labels =
                         c(expression(paste(sigma^2, "/", S^2, "=",
                                            0)),
                           expression(paste(sigma^2, "/", S^2, phantom() %~~% phantom(),
                                            1/5)),
                           expression(paste(sigma^2, "/", S^2, phantom() %~~% phantom(),
                                            1)),
                           expression(paste(sigma^2, "/", S^2, phantom() %~~% phantom(),
                                            5)))) +
  scale_shape_discrete(name = "",
                       labels =
                         c(expression(paste(sigma^2, "/", S^2, "=",
                                            0)),
                           expression(paste(sigma^2, "/", S^2, phantom() %~~% phantom(),
                                            1/5)),
                           expression(paste(sigma^2, "/", S^2, phantom() %~~% phantom(),
                                            1)),
                           expression(paste(sigma^2, "/", S^2, phantom() %~~% phantom(),
                                            5)))) +
  labs(color="", shape="",
       title=expression(atop("(b)",paste("Random effects: ",
                                         S[i]^2,  phantom() %~~% phantom() , 1, ", ",
                                         rho[i], phantom() %~~% phantom(), (i-1)/I, ", ",
                                         rho^BS, " = ", 0.5, ", ",
                                         ##"\n",
                                         I, " = ", 20)))) +
  scale_linetype_discrete(name = "",
                          labels =
                            c(expression(paste(RelEff)),
                              expression(paste(RelEff^T))))
```

5 Dataset for panel c)

These results are for the $p \geq 2$ exchangeable case with equal within-study and between-study variances, with $S_i^2 \equiv S^2, \rho_i = \frac{\rho(i-1)}{I}, \Sigma = 0$ (within study). The parameter we vary is the number of studies, I. We save the relative efficiencies for only one of coefficients, as they are all equal.

```
##save results in data frame
##het is sigma^2/S^2
bigMat <- expand.grid(p = 2:20,
                      I=c(5,10,15,20),
                      RelEff = NA)
for(rr in 1:nrow(bigMat))
{
    p <- bigMat[rr, 1]
    I <- bigMat[rr, 2]

    het <- 0

    ##index over the studies
    i <- 1:I
    ##within-study correlations
    rho.iW <- (i-1)/I
    ##overall correlations
    rho.i <- (rho.iW + rhoBS*het)/(1+het)

    ##calculate the two sums
    ##Sum1 is over 1/(1-rho_i)
    ##Sum2 is over 1/(1+(p-1)*rho_i)
    Sum1 <- sum(1/(1-rho.i))
    Sum2 <- sum(1/(1+(p-1)*rho.i))

    bigMat$RelEff[rr] <- I/p * (Sum1 + (p-1)*Sum2)/(Sum1*Sum2)
}

##make these changes for ggplot
bigMat$I <- as.factor(paste("I=", bigMat$I, sep=""))
bigMat$I <- relevel(bigMat$I, ref="I=5")
```

6 Create plot for panel c)

7 Get the combined dataset for panel d)

Load the files representing the summary for each scenario and save the results in a single dataframe, VarsAll:

```
allFiles <- list.files("simResultsComb")
##remove all the files with "Suppl" in title
allFiles <- allFiles[-grep("Suppl", allFiles)]

##save the three Vars for the different combinations
VarsAll <- expand.grid(I = c(5, 10, 15, 20),
                       p = 2:10,
```

```
                        corrBtw = c(0, 0.5),
                        varBtw = c(0, 1))

##only keep combinations of corrBtw=0 & varBtw=0, corrBtw=0.5 & varBtw=1
VarsAll <- VarsAll[(VarsAll$corrBtw == 0 & VarsAll$varBtw == 0) |
                     (VarsAll$corrBtw == 0.5 & VarsAll$varBtw == 1) ,]

VarsAll$known <- VarsAll$unknown <- VarsAll$univKnown <- VarsAll$univ <- NA

for(file in 1:length(allFiles))
{
    load(paste("simResultsComb",
               allFiles[file], sep="/"))

    rowNr <- gsub("combine_cost_of_estimation_", "", allFiles[file])
    rowNr <- gsub(".RData", "", rowNr)
    rowNr <- as.numeric(rowNr)

    VarsAll[rowNr, colnames(VarsAll)] <- Vars[rowNr, colnames(VarsAll)]
}
VarsAll$Ratio <- VarsAll$known/VarsAll$unknown
VarsAll$RelEff <- VarsAll$unknown/VarsAll$univ
VarsAll$RelEffT <- VarsAll$known/VarsAll$univKnown
range(VarsAll$Ratio)

## [1] 0.7774397 0.9941824

VarsAll$I <- as.factor(paste("I=", VarsAll$I, sep=""))
VarsAll$I <- relevel(VarsAll$I, ref="I=5")
```

Separate out what is needed for Panel d):

```
Vars.b <- VarsAll[VarsAll$corrBtw == 0 &
  VarsAll$varBtw == 0,]

##for panel c, have both RelEff and RelEffT
##probably easier to just create another object
RelEff.d <-
  rbind(cbind(as.matrix(Vars.b[,c("I", "p", "corrBtw", "varBtw", "RelEff"),]),"RelEff"),
        cbind(as.matrix(Vars.b[,c("I", "p", "corrBtw", "varBtw", "RelEffT"),]),"RelEffT"))
colnames(RelEff.d)[6] <- "Estimate"
RelEff.d <- as.data.frame(RelEff.d)
sapply(RelEff.d, class)

##        I        p  corrBtw   varBtw   RelEff Estimate
## "factor" "factor" "factor" "factor" "factor" "factor"

RelEff.d$p <- as.numeric(as.character(RelEff.d$p))
RelEff.d$corrBtw <- as.numeric(as.character(RelEff.d$corrBtw))
RelEff.d$varBtw <- as.numeric(as.character(RelEff.d$varBtw))
RelEff.d$RelEff <- as.numeric(as.character(RelEff.d$RelEff))
RelEff.d$I <- relevel(RelEff.d$I, ref="I=5")
```

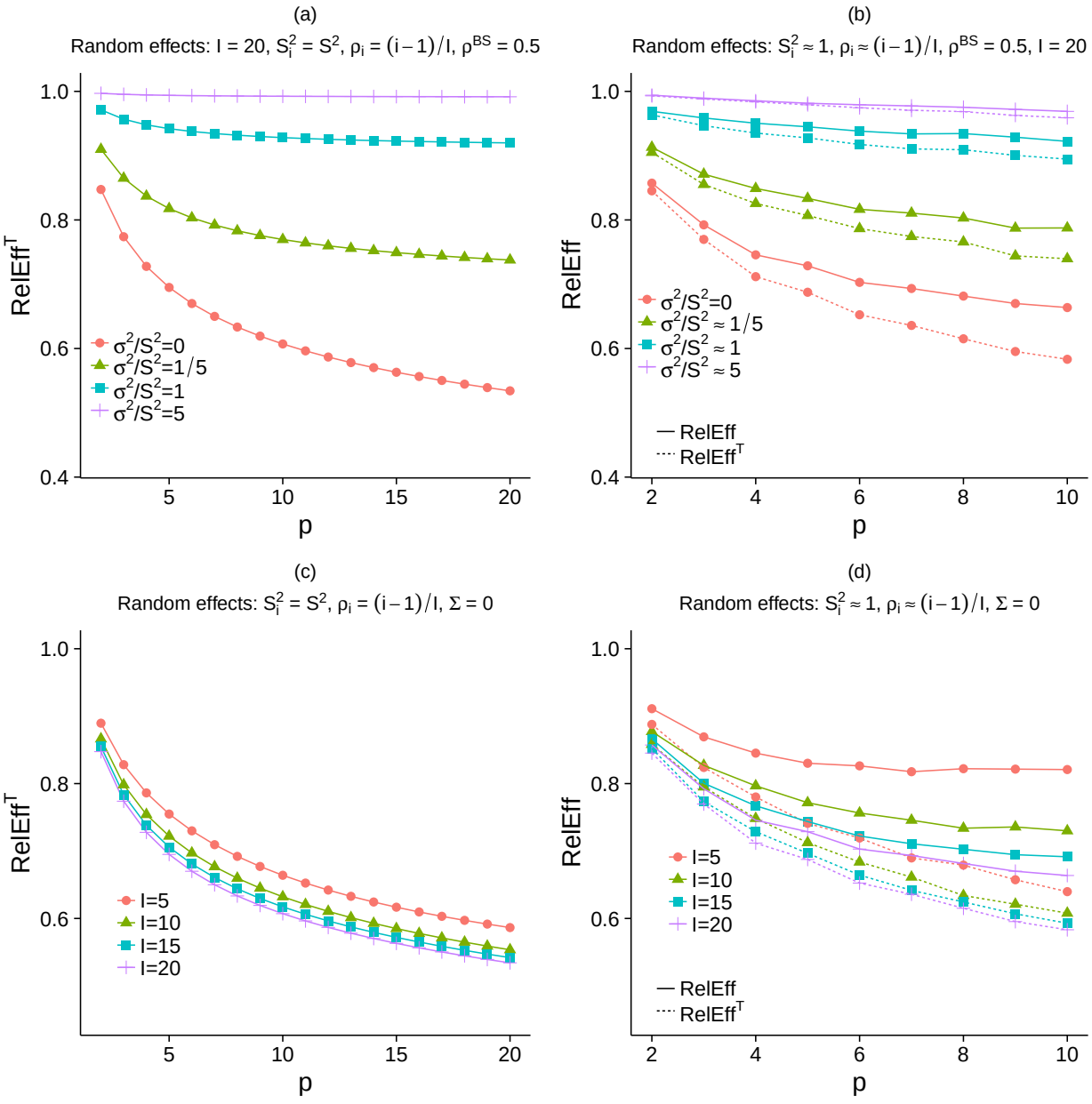
8 Create plot for panel d)

```
panelD <- ggplot(RelEff.d, aes(x=p, y=RelEff))+
  geom_point(size=3.0, aes(color=I, shape=I))+
  geom_line(aes(color=I, shape=I, linetype=Estimate))+
  themeUsed+
  ylab(expression(RelEff)) +
  ylim(min(bigMat$RelEff)*0.85, 1) +
  labs(color="", shape="",
       title=expression(atop("(d)", paste("Random effects: ",
                                          S[i]^2,  phantom() %~~% phantom() , 1, ", ",
                                          rho[i], phantom() %~~% phantom(), (i-1)/I, ", ",
                                          ##"\n",
                                          Sigma, " = ", 0)))) +
  scale_linetype_discrete(name = "",
                          labels =
                            c(expression(paste(RelEff)),
                              expression(paste(RelEff^T))))
```

9 Put all four panels together

```
multiplot(panelA, panelC, panelB, panelD, cols=2)
```

```
## Loading required package:  grid
```


(a)
Random effects: I = 20, $S_i^2 = S^2$, $\rho_i = (i-1)/I$, $\rho^{BS} = 0.5$
$RelEff^T$
p
σ^2/S^2=0
σ^2/S^2=1/5
σ^2/S^2=1
σ^2/S^2=5
(b)
Random effects: $S_i^2 \approx 1$, $\rho_i \approx (i-1)/I$, $\rho^{BS} = 0.5$, I = 20
RelEff
p
σ^2/S^2=0
$\sigma^2/S^2 \approx 1/5$
$\sigma^2/S^2 \approx 1$
$\sigma^2/S^2 \approx 5$
RelEff
$RelEff^T$
(c)
Random effects: $S_i^2 = S^2$, $\rho_i = (i-1)/I$, $\Sigma = 0$
$RelEff^T$
p
I=5
I=10
I=15
I=20
(d)
Random effects: $S_i^2 \approx 1$, $\rho_i \approx (i-1)/I$, $\Sigma = 0$
RelEff
p
I=5
I=10
I=15
I=20
RelEff
$RelEff^T$

```
panelA <- ggplot(bigMat,
                 aes(x=p, y=RelEff))+
  geom_line(aes(color=het, shape=het)) +
  geom_point(size=3.0, aes(color=het, shape=het)) +
  themeUsed +
  scale_y_continuous(name = expression(paste(RelEff^T)),
                     limits=c(min(bigMat$RelEff)*0.8, 1)) +
  scale_color_discrete(name = "",
                       labels =
                         c(expression(paste(sigma^2, "/", S^2, "=",
                                            0)),
                           expression(paste(sigma^2, "/", S^2, "=",
                                            1/5)),
                           ##expression(paste(sigma^2, "/", S^2, "=",
                           ##                 1/2)),
                           expression(paste(sigma^2, "/", S^2, "=",
                                            1)),
                           ##expression(paste(sigma^2, "/", S^2, "=",
                           ##                 2)),
                           expression(paste(sigma^2, "/", S^2, "=",
                                            5)))) +
  scale_shape_discrete(name = "",
                       labels =
                         c(expression(paste(sigma^2, "/", S^2, "=",
                                            0)),
                           expression(paste(sigma^2, "/", S^2, "=",
                                            1/5)),
                           ##expression(paste(sigma^2, "/", S^2, "=",
                           ##                 1/2)),
                           expression(paste(sigma^2, "/", S^2, "=",
                                            1)),
                           ##expression(paste(sigma^2, "/", S^2, "=",
                           ##                 2)),
                           expression(paste(sigma^2, "/", S^2, "=",
                                            5)))) +
  labs(title=expression(atop("(a)", paste("Random effects: I = 20, ",
                                          S[i]^2, " = ", S^2, ", ",
                                          rho[i], " = ", (i-1)/I, ", ",
                                          ##"\n",
                                          rho^BS, " = ", 0.5))))
```

```
panelC <- ggplot(bigMat,
                 aes(x=p, y=RelEff))+
  geom_line(aes(color=I, shape=I)) +
  geom_point(size=3.0, aes(color=I, shape=I)) +
  themeUsed +
  scale_y_continuous(name = expression(paste(RelEff^T)),
                     limits=c(min(bigMat$RelEff)*0.85, 1)) +
  scale_color_discrete(name="")+
  scale_shape_discrete(name="")+
  labs(title=expression(atop("(c)", paste("Random effects: ",
                                          S[i]^2, " = ", S^2, ", ",
                                          rho[i], " = ", (i-1)/I, ", ",
                                          ##"\n",
                                          Sigma, " = ", 0))))
```