```
> # The library contains the function to compute the GREG estimates
with their MSEs
> library("JoSAE")
> # The library contains the function to compute the direct
estimates with their MSEs, the EBLUP unit level estimates and area
level estimates with their MSEs
> library("sae")
> # The function compute the eblup unit level mean estimates
together with their corresponding analytic MSEs
> source("eblup.sae.R")
> # The function is a correction of the eblup.mse.f.wrap funciton of
the library JoSAE to be used if you get an error from the
eblup.mse.f.wrap function (the error is due to a incompatibility
with the latest versions of R)
> source("GREG.sae.R")
> load("~/Desktop/Assignment/SurveyA.RData")
> load("~/Desktop/Assignment/Census.RData")
> load("~/Desktop/Assignment/Data_Area_Level.RData")
> ##########################################
> #Small area estimation with unit level data#
> ##########################################
>
> #-------------------------#
> #Horvitz-Thompson estimates#
> #-------------------------#
> N <- nrow(surveyA)
> H.T. <- direct(Income, LLS, Weights,
data.frame(AREA=unique(surveyA$LLS), N), surveyA)
> head(H.T.)
  Domain SampSize     Direct        SD       CV
1    191        5  29242.909 13098.886 44.79338
2    210        5   3712.683  1648.351 44.39783
3    214        5  11489.206  5189.331 45.16701
4    234       13 174399.753 50038.643 28.69192
5    235       26 344637.793 68747.500 19.94775
6    236       30 398703.990 74718.100 18.74024
> barplot(H.T.$CV, xlab = "LLS", ylab = "CV",main = "Horvitz-
Thompson estimate")
> hist(H.T.$Direct, breaks = 100, xlim = c(0,2000000), main = "H-T
estimate of income", xlab = "estimated income")
> #-------------#
> #GREG estimates#
> #-------------#
> #GREG assisted by linear fixed-effect model
> # linear fixed effect estimate
> #-------------#
> #GREG estimates#
> #-------------#
> #GREG assisted by linear fixed-effect model
> # linear fixed effect estimate
> lm1 = lm(Income ~ SM_House + Work_Status + Gender + Year_Education
+ Single, data= surveyA)
> summary(lm1)
```

```
Call:
lm(formula = Income ~ SM_House + Work_Status + Gender + 
Year_Education + 
    Single, data = surveyA)

Residuals:
    Min      1Q  Median      3Q     Max 
-2568.6  -509.2    12.2   502.7  2522.9 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)    1976.689    197.401   10.01   <2e-16 ***
SM_House         65.618      1.939   33.84   <2e-16 ***
Work_Status    4989.482     63.217   78.93   <2e-16 ***
Gender         2093.739     48.091   43.54   <2e-16 ***
Year_Education  617.068      4.261  144.81   <2e-16 ***
Single        -1620.998     58.465  -27.73   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 731.3 on 1420 degrees of freedom
Multiple R-squared:  0.9646,     Adjusted R-squared:  0.9645 
F-statistic:  7748 on 5 and 1420 DF,  p-value: < 2.2e-16

> MGREG assisted by linear mixed-effect model
Error: unexpected symbol in "MGREG assisted"
> #Means for explanatory variables at domain level
> Mean_Single = tapply(census$Single, census$LLS, mean)
> Mean_Education = tapply(census$Gender, census$LLS, mean)
> Mean_Gender = tapply(census$Gender, census$LLS, mean)
> Mean_work = tapply(census$Work_Status, census$LLS, mean)
> Mean_House = tapply(census$SM_House, census$LLS, mean)
> Means = cbind.data.frame(LLS=unique(surveyC$LLS),
"SM_House"=Mean_House, "Work_Status"=Mean_work,
"Gender"=Mean_Gender, "Year_Education"=Mean_Education,
"Single"=Mean_Single)
Error in unique(surveyC$LLS) : object 'surveyC' not found
> Means = cbind.data.frame(LLS=unique(surveyC$LLS),
"SM_House"=Mean_House, "Work_Status"=Mean_work,
"Gender"=Mean_Gender, "Year_Education"=Mean_Education,
"Single"=Mean_Single)
Error in unique(surveyC$LLS) : object 'surveyC' not found
> Means = cbind.data.frame(LLS=unique(surveyA$LLS),
"SM_House"=Mean_House, "Work_Status"=Mean_work,
"Gender"=Mean_Gender, "Year_Education"=Mean_Education,
"Single"=Mean_Single)
> #linear mixed effect estimates
> lme_Unit = lme(Income ~
SM_House+Work_Status+Gender+Year_Education+Single, data=surveyC,
random=~1|LLS)
Error in is.data.frame(data) : object 'surveyC' not found
> #linear mixed effect estimates
> lme_Unit = lme(Income ~
SM_House+Work_Status+Gender+Year_Education+Single, data=surveyA,
```

```
random=~1|LLS)
> lme_Unit$sigma
[1] 599.1484
> #Producing GREG estimates
> GREG = GREG_function(Means, lme.obj = lme_Unit)
> GREG.Results = cbind.data.frame(GREG$domain.ID, GREG$GREG,
GREG.RMSE = GREG$GREG.se)
> GREG.Results$CV = GREG.Results$GREG.RMSE/
GREG.Results$`GREG$GREG`*100
> barplot(GREG.Results$CV, xlab = "LLS", ylab = "CV",main =
"indirect MGREG estimate")
> plot(density(GREG.Results$`GREG$GREG`), main = "mean household
equivalised income by MGREG", xlab = "domain income", ylab =
"density",xlim=c(11000,22000))
> #---------------#
> #EBLUE estimates#
> #---------------#
> Eblup.Unit = data.frame(eblup_function(y=surveyC$Income,
x=as.matrix(surveyA[,c(2,3,4,5,6)]), regioncode.s=surveyA[,1], m=57,
p=6, x.outs=as.matrix(census[,c(2,3,4,5,6)]), census$LLS))
 Show Traceback

 Rerun with Debug
 Error in cbind(y, x, regioncode.s) : object 'surveyC' not found
> #---------------#
> #EBLUE estimates#
> #---------------#
> Eblup.Unit = data.frame(eblup_function(y=surveyA$Income,
x=as.matrix(surveyA[,c(2,3,4,5,6)]), regioncode.s=surveyA[,1], m=57,
p=6, x.outs=as.matrix(census[,c(2,3,4,5,6)]), census$LLS))
> # m: the number of small areas (m=57)
> # p: the number of covariates + 1 (p=6)
> Eblup.Unit$CV = sqrt(Eblup.Unit$mse)/Eblup.Unit$EBLUP*100
> # m: the number of small areas (m=57)
> # p: the number of covariates + 1 (p=6)
> Eblup.Unit$CV = sqrt(Eblup.Unit$mse)/Eblup.Unit$EBLUP*100
> # m: the number of small areas (m=57)
> # p: the number of covariates + 1 (p=6)
> Eblup.Unit$CV = sqrt(Eblup.Unit$mse)/Eblup.Unit$EBLUP*100
> barplot(Eblup.Unit$CV, xlab = "LLS", ylab = "CV",main = "EBLUP
estimate for unit level")
> plot(density(Eblup.Unit$EBLUP), main = "mean household equivalised
income by EBLUP", xlab = "domain income", ylab = "density",
xlim=c(11000,22000))
> ##########################################
> #Estimation of income and HCR at level data#
> ##########################################
>
> #EBLUP_FH for income at area level
> #EBLUP_FH for income at area level
> Lme_Area = lme(Mean_Income ~
SM_House+Work_Status+Gender+Year_Education+Single, data=data_al,
random=~1|LLS)
> Income_Area = mseFH(Lme_Area, vardir=Var_Mean_Income, method="FH",
```

```
MAXITER = 100000000, data=data_al)
> Income_Area$CV = sqrt(Income_Area$mse)/Income_Area$est$eblup
> plot(density(Income_Area$est$eblup), main = "mean hh equivalised
income at area level by EBLUP_FH", xlab = "domain income", ylab =
"density", xlim=c(11000,22000))
> Income_Area$est$fit$goodness
   loglike        AIC        BIC        KIC       AICc      AICb1
AICb2       KICc      KICb1      KICb2
 -539.3888  1092.7776  1107.0789  1099.7776         NA         NA
NA         NA         NA         NA
nBootstrap
    0.0000
> #EBLUP_FH for H.C.R at area level
> Lme_Area_HCR = lme(HCR ~
1+SM_House+Work_Status+Gender+Year_Education+Single, data=data_al,
random=~1|LLS)
> HCR_Area = mseFH(Lme_Area_HCR, vardir = Var_HCR, method = "FH",
MAXITER = 100000000, data=data_al)
> HCR_Area$CV = sqrt(HCR_Area$mse)/HCR_Area$est$eblup
> plot(density(HCR_Area$est$eblup), main = "head count ratio at area
level by EBLUP_FH", xlab = "head count ratio", ylab = "density",
xlim=c(0,1))
> HCR_Area$est$fit$goodness
   loglike        AIC        BIC        KIC       AICc      AICb1
AICb2       KICc      KICb1      KICb2
10.4327691 -6.8655383  7.4358206  0.1344617         NA         NA
NA         NA         NA         NA
nBootstrap
 0.0000000
> library(stargazer)
> # The library contains the function to compute the GREG estimates
with their MSEs
> library("JoSAE")
> # The library contains the function to compute the direct
estimates with their MSEs, the EBLUP unit level estimates and area
level estimates with their MSEs
> library("sae")
> # The function compute the eblup unit level mean estimates
together with their corresponding analytic MSEs
> source("eblup.sae.R")
> # The function is a correction of the eblup.mse.f.wrap funciton of
the library JoSAE to be used if you get an error from the
eblup.mse.f.wrap function (the error is due to a incompatibility
with the latest versions of R)
> source("GREG.sae.R")
> #-------------------------#
> #Horvitz-Thompson estimates#
> #-------------------------#
> N <- nrow(surveyA)
> H.T. <- direct(Income, LLS, Weights,
data.frame(AREA=unique(surveyA$LLS), N), surveyA)
> head(H.T.)
  Domain SampSize     Direct        SD       CV
1    191        5  29242.909 13098.886 44.79338
```

```
2    210        5   3712.683  1648.351 44.39783
3    214        5  11489.206  5189.331 45.16701
4    234       13 174399.753 50038.643 28.69192
5    235       26 344637.793 68747.500 19.94775
6    236       30 398703.990 74718.100 18.74024
> barplot(H.T.$CV, xlab = "LLS", ylab = "CV",main = "Horvitz-
Thompson estimate")
> hist(H.T.$Direct, breaks = 100, xlim = c(0,2000000), main = "H-T
estimate of income", xlab = "estimated income")
> #--------------#
> #GREG estimates#
> #--------------#
> #GREG assisted by linear fixed-effect model
> # linear fixed effect estimate
> lm1 = lm(Income ~ SM_House + Work_Status + Gender + Year_Education
+ Single, data= surveyA)
> summary(lm1)

Call:
lm(formula = Income ~ SM_House + Work_Status + Gender +
Year_Education +
    Single, data = surveyA)

Residuals:
    Min      1Q  Median      3Q     Max
-2568.6  -509.2    12.2   502.7  2522.9

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    1976.689    197.401   10.01   <2e-16 ***
SM_House         65.618      1.939   33.84   <2e-16 ***
Work_Status    4989.482     63.217   78.93   <2e-16 ***
Gender         2093.739     48.091   43.54   <2e-16 ***
Year_Education  617.068      4.261  144.81   <2e-16 ***
Single        -1620.998     58.465  -27.73   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 731.3 on 1420 degrees of freedom
Multiple R-squared:  0.9646,      Adjusted R-squared:  0.9645
F-statistic:  7748 on 5 and 1420 DF,  p-value: < 2.2e-16

> #MGREG assisted by linear mixed-effect model
> #Means for explanatory variables at domain level
> Mean_Single = tapply(census$Single, census$LLS, mean)
> Mean_Education = tapply(census$Gender, census$LLS, mean)
> Mean_Gender = tapply(census$Gender, census$LLS, mean)
> Mean_work = tapply(census$Work_Status, census$LLS, mean)
> Mean_House = tapply(census$SM_House, census$LLS, mean)
> Means = cbind.data.frame(LLS=unique(surveyA$LLS),
"SM_House"=Mean_House, "Work_Status"=Mean_work,
"Gender"=Mean_Gender, "Year_Education"=Mean_Education,
"Single"=Mean_Single)
> #linear mixed effect estimates
```

```
> lme_Unit = lme(Income ~
SM_House+Work_Status+Gender+Year_Education+Single, data=surveyA,
random=~1|LLS)
> lme_Unit$sigma
[1] 599.1484
> #Producing GREG estimates
> GREG = GREG_function(Means, lme.obj = lme_Unit)
> GREG.Results = cbind.data.frame(GREG$domain.ID, GREG$GREG,
GREG.RMSE = GREG$GREG.se)
> GREG.Results$CV = GREG.Results$GREG.RMSE/
GREG.Results$`GREG$GREG`*100
> barplot(GREG.Results$CV, xlab = "LLS", ylab = "CV",main =
"indirect MGREG estimate")
> plot(density(GREG.Results$`GREG$GREG`), main = "mean household
equivalised income by MGREG", xlab = "domain income", ylab =
"density",xlim=c(11000,22000))
> #---------------#
> #EBLUE estimates#
> #---------------#
> Eblup.Unit = data.frame(eblup_function(y=surveyA$Income,
x=as.matrix(surveyA[,c(2,3,4,5,6)]), regioncode.s=surveyA[,1], m=57,
p=6, x.outs=as.matrix(census[,c(2,3,4,5,6)]), census$LLS))
> # m: the number of small areas (m=57)
> # p: the number of covariates + 1 (p=6)
> Eblup.Unit$CV = sqrt(Eblup.Unit$mse)/Eblup.Unit$EBLUP*100
> barplot(Eblup.Unit$CV, xlab = "LLS", ylab = "CV",main = "EBLUP
estimate for unit level")
> plot(density(Eblup.Unit$EBLUP), main = "mean household equivalised
income by EBLUP", xlab = "domain income", ylab = "density",
xlim=c(11000,22000))
> #EBLUP_FH for income at area level
> Lme_Area = lme(Mean_Income ~
SM_House+Work_Status+Gender+Year_Education+Single, data=data_al,
random=~1|LLS)
> Income_Area = mseFH(Lme_Area, vardir=Var_Mean_Income, method="FH",
MAXITER = 100000000, data=data_al)
> Income_Area$CV = sqrt(Income_Area$mse)/Income_Area$est$eblup
> plot(density(Income_Area$est$eblup), main = "mean hh equivalised
income at area level by EBLUP_FH", xlab = "domain income", ylab =
"density", xlim=c(11000,22000))
> Income_Area$est$fit$goodness
   loglike        AIC        BIC        KIC       AICc      AICb1
AICb2       KICc      KICb1      KICb2
 -539.3888  1092.7776  1107.0789  1099.7776         NA         NA
NA         NA         NA         NA
nBootstrap
    0.0000
> #EBLUP_FH for H.C.R at area level
> Lme_Area_HCR = lme(HCR ~
1+SM_House+Work_Status+Gender+Year_Education+Single, data=data_al,
random=~1|LLS)
> HCR_Area = mseFH(Lme_Area_HCR, vardir = Var_HCR, method = "FH",
MAXITER = 100000000, data=data_al)
> HCR_Area$CV = sqrt(HCR_Area$mse)/HCR_Area$est$eblup
```

```
> plot(density(HCR_Area$est$eblup), main = "head count ratio at area
level by EBLUP_FH", xlab = "head count ratio", ylab = "density",
xlim=c(0,1))
> HCR_Area$est$fit$goodness
   loglike        AIC        BIC        KIC       AICc      AICb1
AICb2       KICc      KICb1      KICb2
10.4327691 -6.8655383  7.4358206  0.1344617         NA         NA
NA         NA         NA         NA
nBootstrap
 0.0000000
```