Mid-term

Ming-Hsuan Lin

March 8, 2015

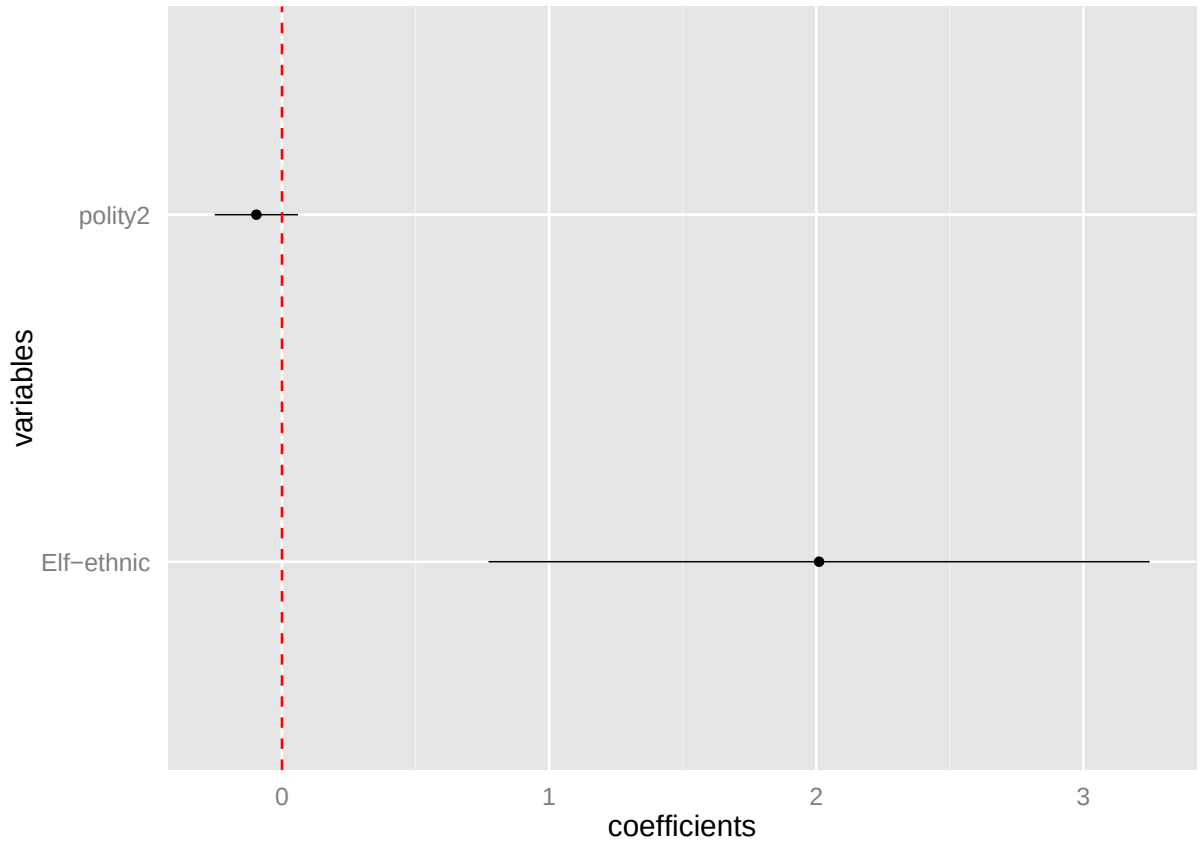

Figure 1: Coefficients Plotting of Model

Table 1: coefficients of model

	Estimate	Std.Error	T.Statistic	P.Value	Lower.95	Upper.95
intercept	-0.22	0.85	-0.26	1.00	-1.89	1.45
Elf-ethnic	2.01	0.63	3.18	0.00	0.77	3.25
polity2	-0.10	0.08	-1.20	0.00	-0.25	0.06

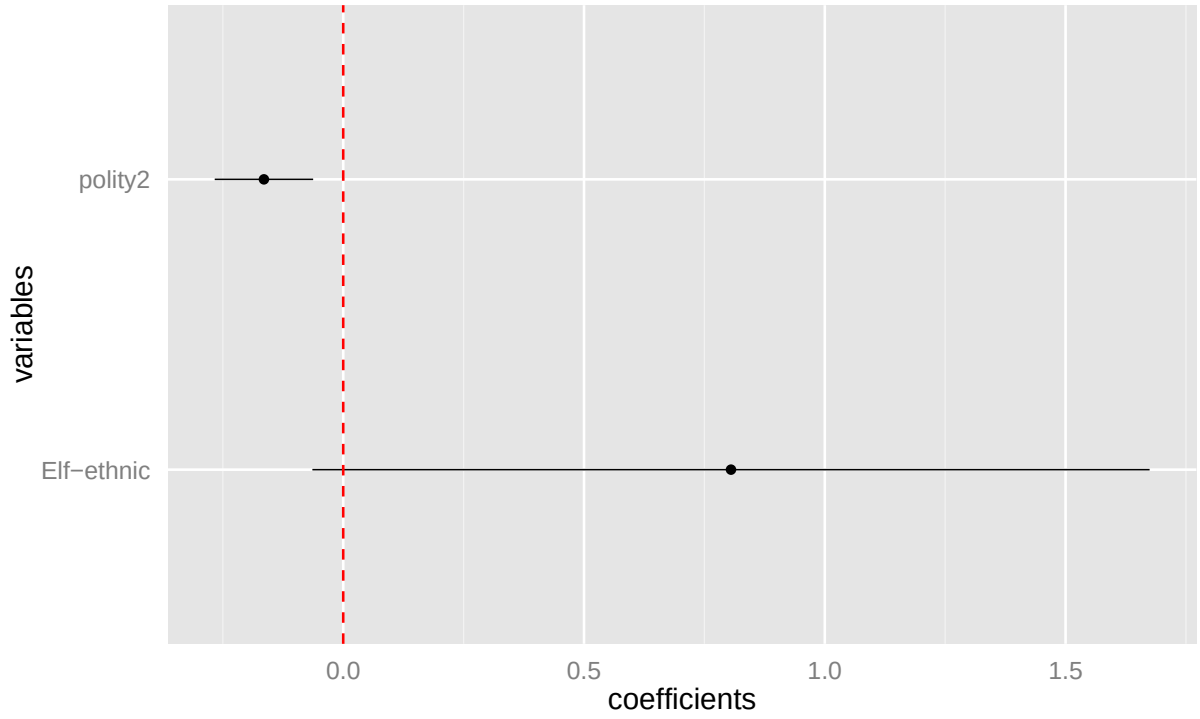

Figure 2: Coefficients Plotting of ModelListDel

Table 2: coefficients of modelListDel

	Estimate	Std.Error	T.Statistic	P.Value	Lower.95	Upper.95
intercept	0.67	0.57	1.18	0.00	-0.45	1.79
Elf-ethnic	0.81	0.44	1.82	0.00	-0.06	1.67
polity2	-0.16	0.05	-3.16	0.00	-0.27	-0.06

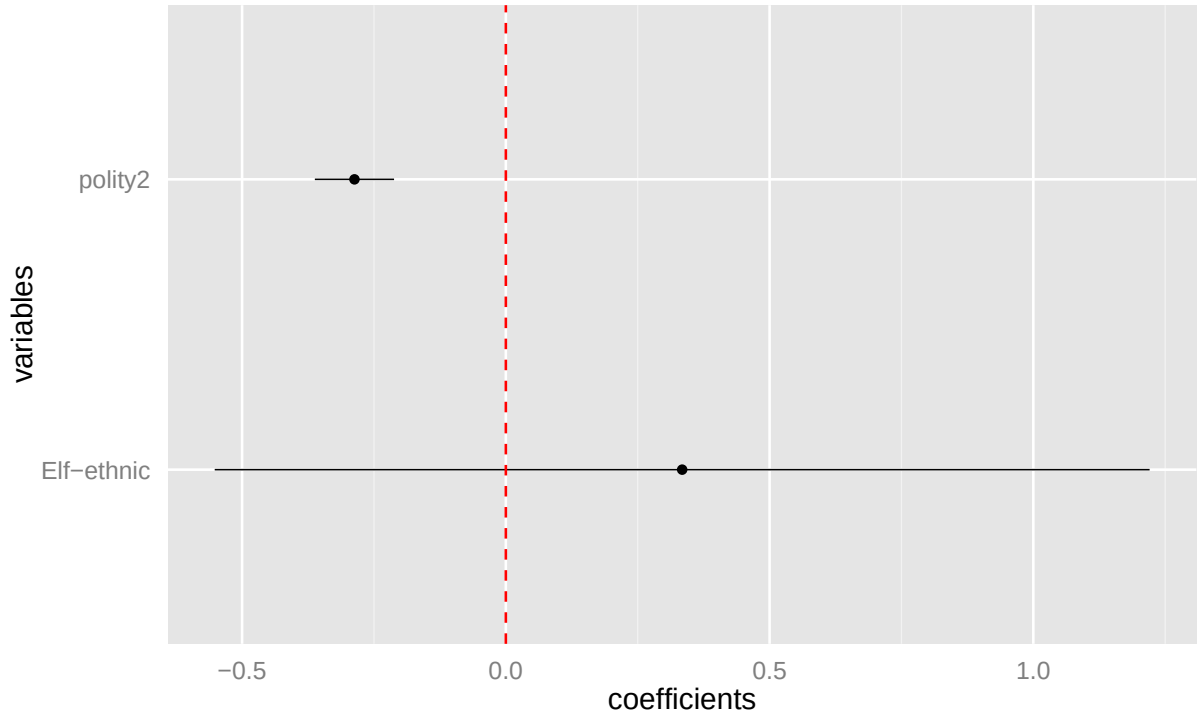

Figure 3: Coefficients Plotting of ModelAmelia

Table 3: coefficients of modelAmelia

	Estimate	Std.Error	T.Statistic	P.Value	Lower.95	Upper.95
intercept	1.94	0.46	4.22	0.00	1.04	2.84
Elf-ethnic	0.33	0.45	0.74	0.00	-0.55	1.22
polity2	-0.29	0.04	-7.49	0.00	-0.36	-0.21

Interpretation

According to the estimate of these three models, compare modelListDel and modelAmelia, the beta of Elf-ethnic, the original model presents 2.01035081, modelListDel presents 0.8052330 and modelAmelia is 0.3342616, which refers that the modelAmelia is more bias than modelListDel. In addition, coefficient of polity2 shows the same problem, the original model beta is -0.10, -0.16 in modelListDel and -0.29 in modelAmelia. Therefore, I ploted the variable of polity2 to gini net std. The plot shows that the outliers in original data were deleted in dataMiss, and data been dealt with amelia function cause more strong bias, because the deleting of missing value did not work randomly.

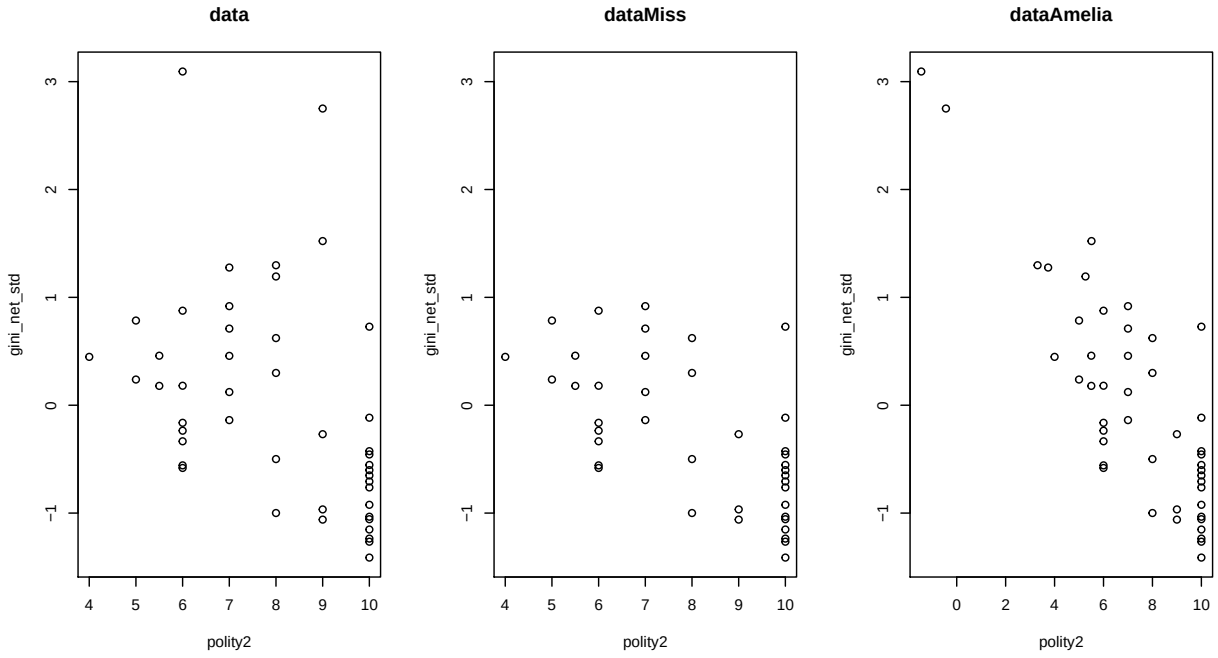

Figure 4: Data Plotting of Three Models

R Code

```
rm(list=ls())
setwd("C:\\Users\\OO\\Desktop\\lab")
load("midTermData.rda")

# Function to load packages
loadPkg=function(toLoad){
  for(lib in toLoad){
  if(! lib %in% installed.packages()[,1])
      { install.packages(lib, repos='http://cran.rstudio.com/') }
    suppressMessages( library(lib, character.only=TRUE) ) }
}

# Load libraries
packs=c('foreign', 'lmtest', 'sandwich', 'Amelia', 'sbgcop')
loadPkg(packs)

#OLS

ols=function(formula,data,impute=FALSE){
  #missing if miss=TURE
  if(impute==T){
    set.seed(6886)
    data=amelia(x=data, m=1)$imp$imp1}
    data=data[complete.cases(data),]
  #(new)y,x,n,p,df
    dv = all.vars(formula)[1]
    ivs = all.vars(formula)[ 2:length(all.vars(formula)) ]
    y = data[,dv]
    x = data.matrix(cbind(1, data[,ivs]))
    n = nrow(x) # Number of observations
    p = length(ivs) # Number of parameters
    df = n-p-1 # degrees of freedom
    #coefficients
    xTx = t(x)%*%x
    xTy = t(x)%*%y
    beta = solve(xTx)%*%xTy
    #serrors
    yhat = x%*%beta
    e = y-yhat
    sigma2 = sum(e^2)/df
    varcov = sigma2*solve(xTx)
    serror = sqrt(diag(varcov))
    #t test, pvalue and CI
    tstate = beta/serror
    pvalue = round(2*pt(abs(tstate),df=n-p-1, lower.tail=F))
    up95=beta+qnorm(0.975)*serror
    lo95=beta-qnorm(0.975)*serror
    #R2
    ssReg = sum((yhat-mean(y))^2)
    ssTot = sum((y-mean(y))^2)
    R2 = ssReg/ssTot
```

```
    #Fstat and pvalue
    msReg = sum((yhat-mean(y))^2)/p
    msRes = sum(e^2)/df
    Fstat = round(msReg/msRes,3)
    Fstatp = round(pf(Fstat,p,df,lower.tail=F),3)
    xx=paste("F-statistic:",Fstat,"on",p,"and",df,"DF,","p-value:",Fstatp,sep=' ')
    #creat matrix and names
    coese=cbind(beta, serror, tstate,pvalue,lo95,up95)
    colnames(coese)=c("Estimate","Std.Error","T-Statistic","P-Value",
                      "Lower 95% CI","Upper 95% CI")
    rownames(coese)=c("intercept", "Elf-ethnic","polity2")
    colnames(varcov)=c("intercept","Elf-ethnic","polity2")
    rownames(varcov)=c("intercept","Elf-ethnic","polity2")
    #list
    list1=list("coefficients"=coese, "varcov"=varcov, "Rsq"=R2, "Fstat"=xx)
    return(list1)
}

#run midterm data
set.seed(6886)

# Set up the model formula

form = formula(gini_net_std ~ ELF_ethnic + polity2)

# Run the various models

model = ols(formula=form, data=data)
modelListDel = ols(formula=form, data=dataMiss)
modelAmelia = ols(formula = form, data=dataMiss, impute=TRUE)

#coefficients plotting
#creat dataframe for plotting

model1=data.frame(model$coefficients)
model2=data.frame(modelListDel$coefficients)
model3=data.frame(modelAmelia$coefficients)

#table coefficients
library(xtable)
xtable(model1)
xtable(model2)
xtable(model3)

#plotting model
library(ggplot2)
model1=cbind(c('intercept','Elf-ethnic','polity2'),model1)
colnames(model1)=c('row.names',colnames(model1)[-1])
tmp=ggplot(data=model1[-1,], aes(x=row.names))
tmp=tmp + geom_linerange(aes(ymin=Lower.95..CI, ymax=Upper.95..CI), size=.3)
tmp=tmp + geom_point(aes(y=Estimate))
tmp=tmp + geom_hline(yintercept=0, linetype=2, color = "red")
tmp=tmp + coord_flip() + xlab('variables') + ylab('coefficients')
```

```
tmp=tmp + theme(axis.ticks=element_blank())
tmp
```

```
#plotting modelListDel
model2=cbind(c('intercept','Elf-ethnic','polity2'),model2)
colnames(model2)=c('row.names',colnames(model2)[-1])
tmp=ggplot(data=model2[-1,], aes(x=row.names))
tmp=tmp + geom_linerange(aes(ymin=Lower.95..CI, ymax=Upper.95..CI), size=.3)
tmp=tmp + geom_point(aes(y=Estimate))
tmp=tmp + geom_hline(yintercept=0, linetype=2, color = "red")
tmp=tmp + coord_flip() + xlab('variables') + ylab('coefficients')
tmp=tmp + theme(axis.ticks=element_blank())
tmp
```

```
#plotting modelAmelia
model3=cbind(c('intercept','Elf-ethnic','polity2'),model3)
colnames(model3)=c('row.names',colnames(model3)[-1])
tmp=ggplot(data=model3[-1,], aes(x=row.names))
tmp=tmp + geom_linerange(aes(ymin=Lower.95..CI, ymax=Upper.95..CI), size=.3)
tmp=tmp + geom_point(aes(y=Estimate))
tmp=tmp + geom_hline(yintercept=0, linetype=2, color = "red")
tmp=tmp + coord_flip() + xlab('variables') + ylab('coefficients')
tmp=tmp + theme(axis.ticks=element_blank())
tmp
```

```
#compare three model by plot polity2 with Y
set.seed(6886)
ameliadata=amelia(dataMiss, m=1)$imp$imp1

par(mfrow=c(1,3))

plot(gini_net_std ~ polity2, data=data, main="data")
plot(gini_net_std ~ polity2, data=dataMiss, main="dataMiss")
plot(gini_net_std ~ polity2, data=ameliadata, main="dataAmelia")
```