econ4274\_problem\_set\_3

Man Adrian Ho Hin

2023-04-20

## Question 1a

rm(list=ls())  
DATA = read.csv("cigarette.csv")  
  
p0\_2000=mean(DATA$C2000[which(DATA$TAX==0)])  
p1\_2000=mean((DATA$C2000[which(DATA$TAX==1)]))  
p0\_2006=mean(DATA$C2006[which(DATA$TAX==0)])  
p1\_2006=mean((DATA$C2006[which(DATA$TAX==1)]))  
p0\_p1\_diff\_2000=p1\_2000-p0\_2000  
p0\_p1\_diff\_2006=p1\_2006-p0\_2006  
  
d0=p0\_2006-p0\_2000  
d1=p1\_2006-p1\_2000  
DID=p0\_p1\_diff\_2006-p0\_p1\_diff\_2000

|  |  |  |  |
| --- | --- | --- | --- |
|  | Average cigarette consumption | | |
|  | Before | After | After − Before |
| Control | 22.629 | 20.200 | -2.429 |
| Treatment | 23.410 | 20.250 | -3.160 |
| Treatment − Control | 0.781 | 0.050 | -0.731 |

## Question 1b

library(stargazer)

##   
## Please cite as:

## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

temp1 = cbind(DATA$TAX, DATA$C2000,0)  
temp2 = cbind(DATA$TAX, DATA$C2006,1)  
temp\_data = rbind(temp1,temp2)  
colnames(temp\_data) = c("TAX", "consumption", "y2006")  
  
m1=lm(temp\_data[,2]~temp\_data[,3]+temp\_data[,1]+temp\_data[,3]\*temp\_data[,1])  
stargazer(m1, type="html", out="Q1b.html")

|  |  |
| --- | --- |
| Regression Table | |
|  | *Dependent variable:* |
|  | consumption |
|  | (1) |
| y2006 | -2.429\*\*\* |
|  | (0.737) |
| TAX | 0.781 |
|  | (1.105) |
| y2006:TAX | -0.731 |
|  | (1.562) |
| Observations | 90 |
| R2 | 0.160 |
| Adjusted R2 | 0.131 |
| Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01 | |

## Question 2a

rm(list=ls())  
library(foreign)  
data = read.dta("colony.dta")  
  
base = subset(data, baseco == 1)  
head(base)

## shortnam africa lat\_abst rich4 avexpr logpgp95 logem4 asia loghjypl  
## 2 AGO 1 0.1366667 0 5.363636 7.770645 5.634789 0 -3.4112477  
## 4 ARG 0 0.3777778 0 6.386364 9.133459 4.232656 0 -0.8722738  
## 6 AUS 0 0.3000000 1 9.318182 9.897972 2.145931 0 -0.1707883  
## 12 BFA 1 0.1444445 0 4.454545 6.845880 5.634789 0 -3.5404594  
## 13 BGD 0 0.2666667 0 5.136364 6.877296 4.268438 1 -2.0635681  
## 16 BHS 0 0.2683333 0 7.500000 9.285448 4.442651 0 NA  
## baseco  
## 2 1  
## 4 1  
## 6 1  
## 12 1  
## 13 1  
## 16 1

base\_without = subset(base, rich4 == 0)  
head(base\_without)

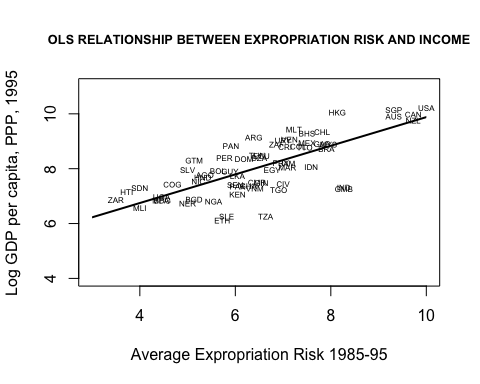
## shortnam africa lat\_abst rich4 avexpr logpgp95 logem4 asia loghjypl  
## 2 AGO 1 0.1366667 0 5.363636 7.770645 5.634789 0 -3.4112477  
## 4 ARG 0 0.3777778 0 6.386364 9.133459 4.232656 0 -0.8722738  
## 12 BFA 1 0.1444445 0 4.454545 6.845880 5.634789 0 -3.5404594  
## 13 BGD 0 0.2666667 0 5.136364 6.877296 4.268438 1 -2.0635681  
## 16 BHS 0 0.2683333 0 7.500000 9.285448 4.442651 0 NA  
## 20 BOL 0 0.1888889 0 5.636364 7.926602 4.262680 0 -1.9661129  
## baseco  
## 2 1  
## 4 1  
## 12 1  
## 13 1  
## 16 1  
## 20 1

## Question 2b

# figure 1  
plot(base$logem4, base$logpgp95,   
 main = "REDUCED-FORM RELATIONSHIP BETWEEN INCOME AND SETTLER MORTALITY",  
 cex.main=0.8,  
 xlab = "Log of Settler Mortality",  
 ylab = "Log GDP per capita, PPP, 1995",  
 xaxt='n',  
 yaxt='n',  
 xlim = c(2,8),   
 ylim = c(4,10),   
 type="n")  
text(base$logem4, base$logpgp95, labels = base$shortnam, cex = 0.5)  
axis(side=1, at=seq(2, 8, by=2))  
axis(side=2, at=seq(4, 10, by=2))  
clip(x1 = 2,  
 x2 = 8,  
 y1 = 4,  
 y2 = 10)  
abline(lm(logem4~logpgp95, data = base), lwd=2)



# figure 2  
plot(base$avexpr, base$logpgp95,   
 main = "OLS RELATIONSHIP BETWEEN EXPROPRIATION RISK AND INCOME",  
 cex.main=0.8,  
 xlab = "Average Expropriation Risk 1985-95",  
 ylab = "Log GDP per capita, PPP, 1995",  
 xaxt='n',  
 yaxt='n',  
 xlim = c(3,10),   
 ylim = c(4,11),   
 type="n")  
text(base$avexpr, base$logpgp95, labels = base$shortnam, cex = 0.5)  
axis(side=1, at=seq(4, 10, by=2))  
axis(side=2, at=seq(4, 10, by=2))  
clip(x1 = 3,  
 x2 = 10,  
 y1 = 4,  
 y2 = 10)  
abline(lm(logpgp95~avexpr, data = base), lwd=2)



## Question 2c

# Panel C: Ordinary Least Squares  
ols1=lm(logpgp95~avexpr,data=base)  
ols2=lm(logpgp95~avexpr+lat\_abst,data=base)  
ols3=lm(logpgp95~avexpr,data=base\_without)  
ols4=lm(logpgp95~avexpr+lat\_abst,data=base\_without)  
stargazer(ols1, ols2, ols3, ols4, type="html", out="Q2cC.html")

# Panel B: First Stage for Average Protection Against Expropriation Risk in 1985–1995  
f1=lm(avexpr~logem4,data=base)   
f2=lm(avexpr~logem4+lat\_abst,data=base)  
f3=lm(avexpr~logem4,data=base\_without)  
f4=lm(avexpr~logem4+lat\_abst,data=base\_without)  
stargazer(f1,f2,f3,f4, type = "html", out="Q2cB.html")

# Panel A: Two-Stage Least Squares  
library(AER)

## Loading required package: car

## Loading required package: carData

## Loading required package: lmtest

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

## Loading required package: sandwich

## Loading required package: survival

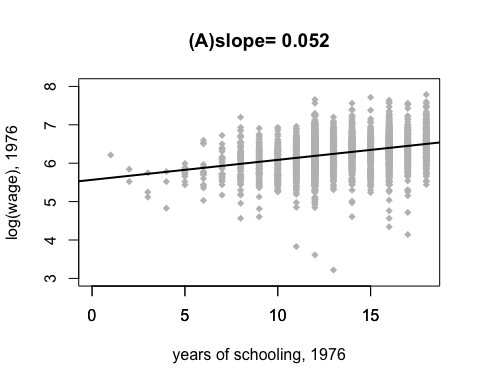
iv1=ivreg(logpgp95~avexpr |   
 logem4   
 ,data=base)  
iv2=ivreg(logpgp95~avexpr+lat\_abst |   
 logem4+lat\_abst  
 ,data=base)  
iv3=ivreg(logpgp95~avexpr |   
 logem4  
 ,data=base\_without)  
iv4=ivreg(logpgp95~avexpr+lat\_abst |   
 logem4+lat\_abst  
 ,data=base\_without)  
stargazer(iv1,iv2,iv3,iv4, type = "html", out="Q2cA.html")

The Table:

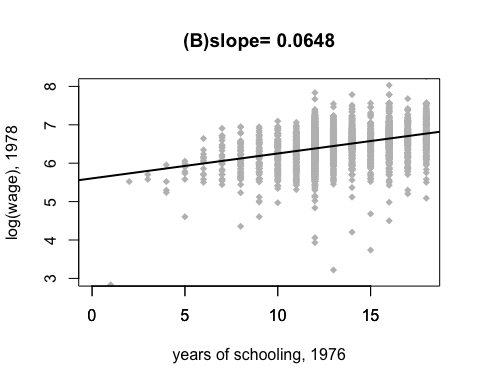


## Question 3a

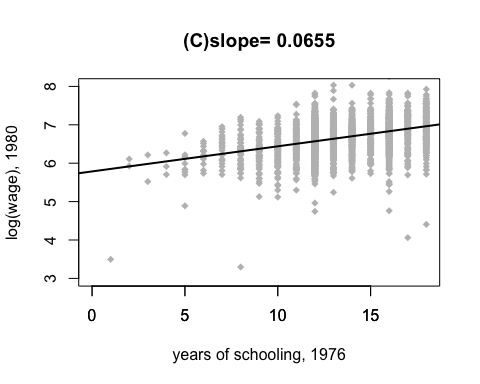
rm(list=ls())  
DATA = read.csv("card.csv")  
  
lm1=lm(lwage76~educ, data=DATA)  
lm2=lm(lwage78~educ, data=DATA)  
lm3=lm(lwage80~educ, data=DATA)  
  
plot(lwage76~educ,   
 data=DATA,  
 pch=18,  
 col = "grey",  
 xlab="years of schooling, 1976",  
 ylab="log(wage), 1976",  
 main=paste0("(A)slope= ",round(lm1$coefficients[2],4)),  
 ylim=c(3,8))  
axis(side=1, at=seq(0, 15, by=5))  
abline(lm1, lwd=2)



plot(lwage78~educ,   
 data=DATA,  
 pch=18,  
 col = "grey",  
 xlab="years of schooling, 1976",  
 ylab="log(wage), 1978",  
 main=paste0("(B)slope= ",round(lm2$coefficients[2],4)),  
 ylim=c(3,8))  
axis(side=1, at=seq(0, 15, by=5))  
abline(lm2, lwd=2)



plot(lwage80~educ,   
 data=DATA,  
 pch=18,  
 col = "grey",  
 xlab="years of schooling, 1976",  
 ylab="log(wage), 1980",  
 main=paste0("(C)slope= ",round(lm3$coefficients[2],4)),  
 ylim=c(3,8))  
axis(side=1, at=seq(0, 15, by=5))  
abline(lm3, lwd=2)



## Question 3b

ols1=lm(lwage76~educ+kww+iq+age+black+smsa+south+famed, data=DATA)  
ols2=lm(lwage78~educ+kww+iq+age+black+smsa+south+famed, data=DATA)  
ols3=lm(lwage80~educ+kww+iq+age+black+smsa+south+famed, data=DATA)  
  
library(stargazer)  
stargazer(ols1,ols2,ols3,type="html", out="Q3b.html")

## 

## Question 3c

f1=lm(educ~nearc2+kww+iq+age+black+smsa+south+famed,data=DATA)  
f2=lm(educ~nearc4+kww+iq+age+black+smsa+south+famed,data=DATA)  
f3=lm(educ~nearc4a+kww+iq+age+black+smsa+south+famed,data=DATA)  
f4=lm(educ~nearc4b+kww+iq+age+black+smsa+south+famed,data=DATA)  
  
cat("The t-value of nearc2 from f1 (model) = ", summary(f1)$coefficients[2, 3],"\n")

## The t-value of nearc2 from f1 (model) = 0.4447563

cat("The t-value of nearc4 from f2 (model) = ", summary(f2)$coefficients[2, 3],"\n")

## The t-value of nearc4 from f2 (model) = 2.095283

cat("The t-value of nearc4a from f3 (model) = ",summary(f3)$coefficients[2, 3],"\n")

## The t-value of nearc4a from f3 (model) = 2.062943

cat("The t-value of nearc4b from f4 (model) = ",summary(f4)$coefficients[2, 3],"\n")

## The t-value of nearc4b from f4 (model) = -0.2345304

# Since the t-values of nearc4b < nearc2 < nearc4a < nearc4, therefore nearc4a and nearc4 are the 2 IVs that most correlated with educ

## Question 3d

ols1=lm(lwage76~educ+kww+iq+age+black+smsa+south+famed, data=DATA)  
ols2=lm(lwage78~educ+kww+iq+age+black+smsa+south+famed, data=DATA)  
ols3=lm(lwage80~educ+kww+iq+age+black+smsa+south+famed, data=DATA)  
  
iv1=ivreg(lwage76~educ+kww+iq+age+black+smsa+south+famed |   
 nearc4+nearc4a+kww+iq+age+black+smsa+south+famed  
 ,data=DATA)  
iv2=ivreg(lwage78~educ+kww+iq+age+black+smsa+south+famed |   
 nearc4+nearc4a+kww+iq+age+black+smsa+south+famed  
 ,data=DATA)  
iv3=ivreg(lwage80~educ+kww+iq+age+black+smsa+south+famed |   
 nearc4+nearc4a+kww+iq+age+black+smsa+south+famed  
 ,data=DATA)  
  
stargazer(ols1,iv1,ols2,iv2,ols3,iv3,type="html", out="Q3d.html")

## 

After applying the IVs, most of the estimation significant level drops, for example educ, kww and iq turns into statistically insignificant among 3 models. Also the standard error of each seems increased. Since the 2 IVs having a 5% significant level that correlated with educ, the OLS may suffer an endogeneity problem, so the estimators are biased and inconsistent.

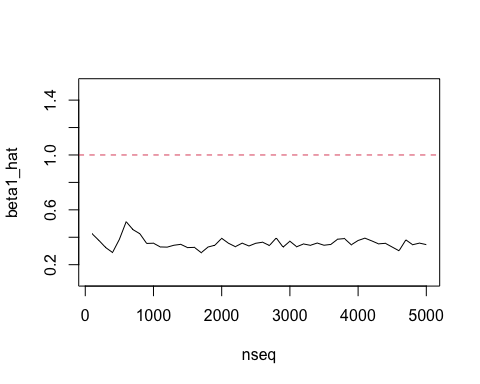
## Question 3e

DATA = read.csv("card.csv")  
  
DATA$work76 <- with(DATA, ifelse(is.na(DATA$wage76), 0, 1))  
DATA$work78 <- with(DATA, ifelse(is.na(DATA$wage78), 0, 1))  
DATA$work80 <- with(DATA, ifelse(is.na(DATA$wage80), 0, 1))  
  
logit1=glm(work76~educ+kww+iq+age+black+smsa+south+famed,data=DATA,family=binomial(link="logit"))  
logit2=glm(work78~educ+kww+iq+age+black+smsa+south+famed,data=DATA,family=binomial(link="logit"))  
logit3=glm(work80~educ+kww+iq+age+black+smsa+south+famed,data=DATA,family=binomial(link="logit"))  
probit1=glm(work76~educ+kww+iq+age+black+smsa+south+famed,data=DATA,family=binomial(link="probit"))  
probit2=glm(work78~educ+kww+iq+age+black+smsa+south+famed,data=DATA,family=binomial(link="probit"))  
probit3=glm(work80~educ+kww+iq+age+black+smsa+south+famed,data=DATA,family=binomial(link="probit"))  
  
stargazer(logit1,probit1,logit2,probit2,logit3,probit3,type="html",out="Q3e.html")

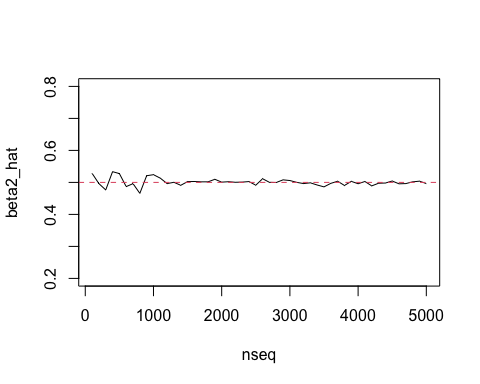
## 

## Question 4a

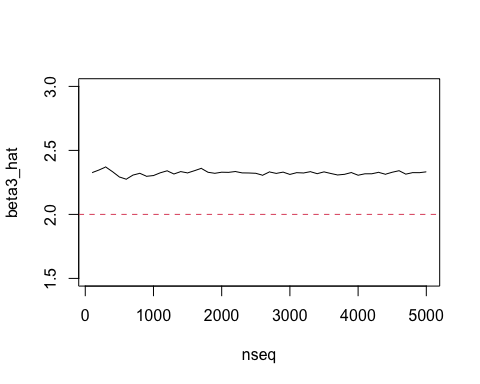
library(MASS)  
rm(list=ls())  
  
set.seed(4274)  
  
beta1 = 1  
beta2 = 0.5  
beta3 = 2  
  
data\_gen = function(n){  
 temp = mvrnorm(n, c(0,0), matrix(c(1,0.8,0.8,1), 2,2))  
 e = temp[,1]  
 u = temp[,2]  
 z1 = rnorm(n,1,1)  
 z2 = rgamma(n,2)  
 z3 = rnorm(n,2,0.5)  
 x1 = rnorm(n,1,2)  
 x2 = z1 + 0.5\*z2 + 0.01\*z3 + u  
 y = beta1 +beta2\*x1 +beta3\*x2 +e  
 output = NULL  
 output = cbind(y, x1, x2, z1, z2, z3, e, u)  
}  
  
beta1\_hat = NULL  
beta2\_hat = NULL  
beta3\_hat = NULL  
nseq=seq(100,5000,by = 100)  
for(i in nseq){  
 DATA = data\_gen(i)  
 ols = lm(DATA[,'y']~DATA[,'x1']+DATA[,'x2'])  
 beta1\_hat = rbind(beta1\_hat, ols$coefficients[1])  
 beta2\_hat = rbind(beta2\_hat, ols$coefficients[2])  
 beta3\_hat = rbind(beta3\_hat, ols$coefficients[3])  
}  
  
plot(nseq, beta1\_hat, type="l", ylim=c(0.1,1.5))  
abline(h=beta1,col=2,lty=2)



plot(nseq, beta2\_hat, type="l", ylim=c(0.2,0.8))  
abline(h=beta2,col=2,lty=2)



plot(nseq, beta3\_hat, type="l", ylim=c(1.5,3))  
abline(h=beta3,col=2,lty=2)



## Question 4b

set.seed(4274)  
n = 100  
DATA = data\_gen(n)  
  
library(AER)  
  
iv1=ivreg(DATA[,'y']~DATA[,'x1']+DATA[,'x2'] | DATA[,'x1']+DATA[,'z1'])  
iv2=ivreg(DATA[,'y']~DATA[,'x1']+DATA[,'x2'] | DATA[,'x1']+DATA[,'z2'])  
iv3=ivreg(DATA[,'y']~DATA[,'x1']+DATA[,'x2'] | DATA[,'x1']+DATA[,'z3'])  
  
stargazer(iv1,iv2,iv3,type="html",out="Q4b.html")

## 

## Question 4c

set.seed(4274)  
n = 100  
DATA = data\_gen(n)  
  
ols=lm(DATA[,'y']~DATA[,'x1']+DATA[,'x2'])  
iv1=ivreg(DATA[,'y']~DATA[,'x1']+DATA[,'x2'] | DATA[,'x1']+DATA[,'z1'])  
iv2=ivreg(DATA[,'y']~DATA[,'x1']+DATA[,'x2'] | DATA[,'x1']+DATA[,'z2'])  
iv3=ivreg(DATA[,'y']~DATA[,'x1']+DATA[,'x2'] | DATA[,'x1']+DATA[,'z3'])  
iv4=ivreg(DATA[,'y']~DATA[,'x1']+DATA[,'x2'] | DATA[,'x1']+DATA[,'z1']+DATA[,'z2']+DATA[,'z3'])  
  
stargazer(ols,iv1,iv2,iv3,iv4,type="html",out="Q4c.html")

## 

It seems that TSLS\_z2 is the best. Because the actual beta1, beta2 and beta3 are 1, 0.5 and 2 respectively, TSLS\_z2 provided the closest beta estimations 0.97, 0.54, and 1.99. While the standard error of the beta estimations may be too large, so in some sense TSLS\_z1,z2,z3, which is the second closest estimation with smaller standard error, could be better than TSLS\_z2.

## Question 4d

z3 has a strong correlation with x2 which provide a close estimation for beta2 with a 0.05 standard error, while only for beta2 estimation. It also has an incorrect estimation and very high standard error among the other estimations, which implies that should be an improper IV.