R Code

Yuhao Liu

07/08/2021

```
library(readxl)
```

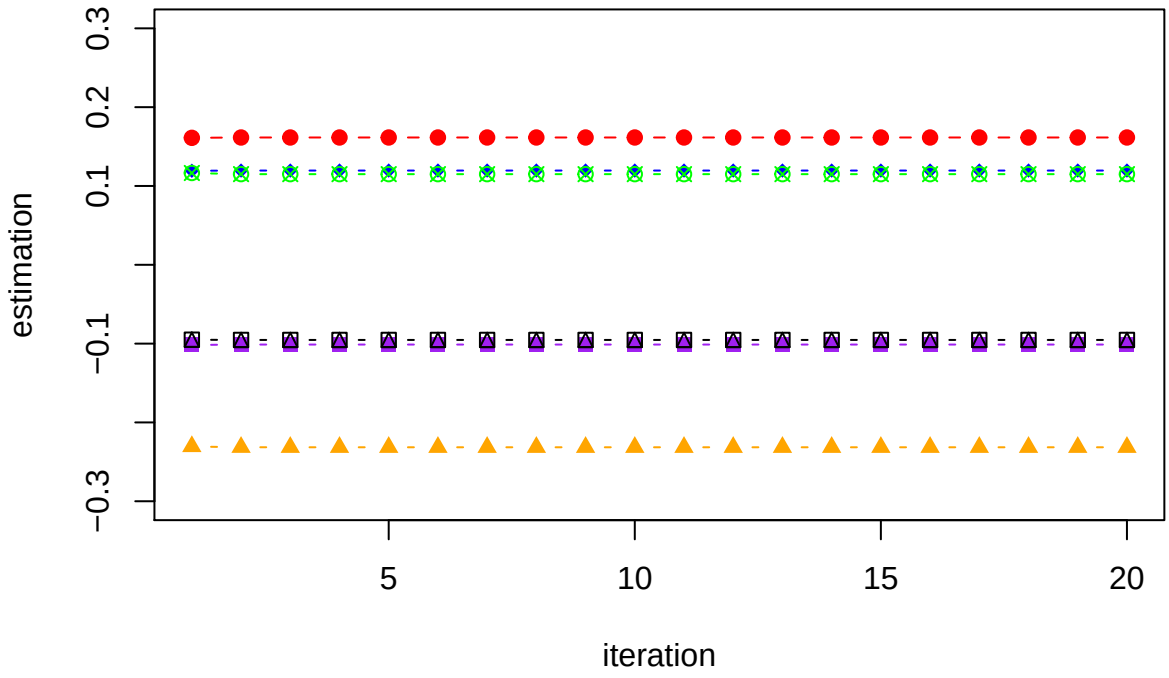
```
## Warning: package 'readxl' was built under R version 3.6.2
```

```
dat <- read_excel("CCC.xlsx")


#data
logwage<-dat$lwage76
edu<-dat$ed76
exper<-dat$age76-dat$ed76-6
expersq100<-exper^2/100
black<-dat$black
south<-dat$reg76r
urban<-dat$smsa76r

pub<-dat$nearc4a
priv<-dat$nearc4b
college4<-dat$nearc4
college2<-dat$nearc2

#12.25 a
#reduced form regression
y<-as.matrix(edu)
x<-as.matrix(cbind(exper,expersq100,black,south,urban,pub,priv,matrix(1,nrow(y),1)))
betaa<-solve(t(x)%*%x,t(x)%*%y)

#standard error
n <- nrow(y)
k <- ncol(x)
invx <- solve(t(x)%*%x)
a <- n/(n-k)
e <- y-x%*%betaa
leverage <- rowSums(x*(x%*%invx))
sig2 <- (t(e) %*% e)/(n-k)
u1 <- x*(e%*%matrix(1,1,k))
u2 <- x*((e/sqrt(1-leverage))%*%matrix(1,1,k))
u3 <- x*((e/(1-leverage))%*%matrix(1,1,k))
v0 <- as.numeric(sig2)*invx
v1 <- invx %*% (t(u1)%*%u1) %*% invx
v1a <- a * invx %*% (t(u1)%*%u1) %*% invx
v2 <- invx %*% (t(u2)%*%u2) %*% invx
v3 <- invx %*% (t(u3)%*%u3) %*% invx
```

```
s0 <- sqrt(diag(v0)) # Homoskedastic formula
s1 <- sqrt(diag(v1)) # HC0
s1a <- sqrt(diag(v1a)) # HC1
s2 <- sqrt(diag(v2)) # HC2
s3 <- sqrt(diag(v3)) # HC3



#2sls
y<-as.matrix(logwage)
x<-as.matrix(cbind(edu,exper,expersq100,black,south,urban,matrix(1,nrow(y),1)))

z<-as.matrix(cbind(pub,priv,exper,expersq100,black,south,urban,matrix(1,nrow(y),1)))
zz<-t(z)%*%z
xz<-t(x)%*%z
zy<-t(z)%*%y

beta2sls<-solve(xz%*%solve(zz,t(xz)),xz%*%solve(zz,zy))
#standard error
n <- nrow(y)
e<-y-x%*%beta2sls
sig2<-(t(e) %*% e)/n
ze<-z*matrix(rep(e,dim(z)[2]),ncol = dim(z)[2],nrow = n)
Qzx=t(z)%*%x/n
Qzz=t(z)%*%z/n
Om=t(ze)%*%ze/n

v0<-solve(t(Qzx)%*%solve(Qzz,Qzx))*as.numeric(sig2)#homo
v1<-solve(t(Qzx)%*%solve(Qzz,Qzx),t(Qzx))%*%solve(Qzz,Om)%*%solve(Qzz,Qzx)%*%
  +solve(t(Qzx)%*%solve(Qzz,Qzx))#HC0
v1a<-(n/(n-dim(z)[2]))*v1#HC1
s0<-sqrt(diag(v0/n))#homo
s1<-sqrt(diag(v1/n))#HC0
s1a <- sqrt(diag(v1a/n))#HC1

#12.25 b
#new reduced form regression with 2 year college
y<-as.matrix(edu)
x<-as.matrix(cbind(exper,expersq100,black,south,urban,pub,priv,college2,matrix(1,nrow(y),1)))
betab<-solve(t(x)%*%x,t(x)%*%y)
#standard error
n <- nrow(y)
k <- ncol(x)
invx <- solve(t(x)%*%x)
a <- n/(n-k)
e <- y-x%*%betab
leverage <- rowSums(x*(x%*%invx))
sig2 <- (t(e) %*% e)/(n-k)
u1 <- x*(e%*%matrix(1,1,k))
u2 <- x*((e/sqrt(1-leverage))%*%matrix(1,1,k))
u3 <- x*((e/(1-leverage))%*%matrix(1,1,k))
v0 <- as.numeric(sig2)*invx
v1 <- invx %*% (t(u1)%*%u1) %*% invx
```

```
v1a <- a * invx %*% (t(u1)%*%u1) %*% invx
v2 <- invx %*% (t(u2)%*%u2) %*% invx
v3 <- invx %*% (t(u3)%*%u3) %*% invx
s0 <- sqrt(diag(v0)) # Homoskedastic formula
s1 <- sqrt(diag(v1)) # HC0
s1a <- sqrt(diag(v1a)) # HC1
s2 <- sqrt(diag(v2)) # HC2
s3 <- sqrt(diag(v3)) # HC3
tstatistic<-betab[8]/s1a[8]
pval<-dnorm(tstatistic)

#12.25 c
intact1<-pub*dat$age76
intact2<-pub*(dat$age76)^2/100
#new reduced form regression with 2 new interaction
y<-as.matrix(edu)
x<-as.matrix(cbind(exper,expersq100,black,south,urban,pub,priv,intact1,intact2,matrix(1,nrow(y),1)))
betac<-solve(t(x)%*%x,t(x)%*%y)
#standard error
n <- nrow(y)
k <- ncol(x)
invx <- solve(t(x)%*%x)
a <- n/(n-k)
e <- y-x%*%betac
leverage <- rowSums(x*(x%*%invx))
sig2 <- (t(e) %*% e)/(n-k)
u1 <- x*(e%*%matrix(1,1,k))
u2 <- x*((e/sqrt(1-leverage))%*%matrix(1,1,k))
u3 <- x*((e/(1-leverage))%*%matrix(1,1,k))
v0 <- as.numeric(sig2)*invx
v1 <- invx %*% (t(u1)%*%u1) %*% invx
v1a <- a * invx %*% (t(u1)%*%u1) %*% invx
v2 <- invx %*% (t(u2)%*%u2) %*% invx
v3 <- invx %*% (t(u3)%*%u3) %*% invx
s0 <- sqrt(diag(v0)) # Homoskedastic formula
s1 <- sqrt(diag(v1)) # HC0
s1a <- sqrt(diag(v1a)) # HC1
s2 <- sqrt(diag(v2)) # HC2
s3 <- sqrt(diag(v3)) # HC3
tstatistic1<-betac[8]/s1a[8]
pval1<-dnorm(tstatistic1)
tstatistic2<-betac[9]/s1a[9]
pval2<-dnorm(tstatistic1)

#12.25 d 2sls with expanded instrument
intact1<-pub*dat$age76
intact2<-pub*(dat$age76)^2/100
y<-as.matrix(logwage)
x<-as.matrix(cbind(edu,exper,expersq100,black,south,urban,matrix(1,nrow(y),1)))

z<-as.matrix(cbind(pub,priv,intact1,intact2,exper,expersq100,black,south,urban,matrix(1,nrow(y),1)))
zz<-t(z)%*%z
xz<-t(x)%*%z
```

```r
zy<-t(z)%*%y

beta2slsnew<-solve(xz%*%solve(zz,t(xz)),xz%*%solve(zz,zy))
#standard error
n <- nrow(y)
e<-y-x%*%beta2slsnew
sig2<-(t(e) %*% e)/n
ze<-z*matrix(rep(e,dim(z)[2]),ncol = dim(z)[2],nrow = n)
Qzx=t(z)%*%x/n
Qzz=t(z)%*%z/n
Om=t(ze)%*%ze/n

v0<-solve(t(Qzx)%*%solve(Qzz,Qzx))*as.numeric(sig2)#homo
v1<-solve(t(Qzx)%*%solve(Qzz,Qzx),t(Qzx))%*%solve(Qzz,Om)%*%solve(Qzz,Qzx)%*%
  +solve(t(Qzx)%*%solve(Qzz,Qzx))#HC0
v1a<-(n/(n-dim(z)[2]))*v1#HC1
s0<-sqrt(diag(v0/n))#homo
s1<-sqrt(diag(v1/n))#HC0
s1a <- sqrt(diag(v1a/n))#HC1
tstat<-beta2slsnew/s1
p_val<-dnorm(tstat)




#13.28
#a
y<-as.matrix(logwage)
x<-as.matrix(cbind(edu,exper,expersq100,black,south,urban,matrix(1,nrow(y),1)))
z<-as.matrix(cbind(pub,priv,exper,expersq100,black,south,urban,matrix(1,nrow(y),1)))
zz<-t(z)%*%z
xz<-t(x)%*%z
zy<-t(z)%*%y
n <- nrow(y)
betaa=matrix(solve(xz%*%solve(zz,t(xz)),xz%*%solve(zz,zy)),ncol = 1)

for (m in 2:20) {

  e<-y-x%*%betaa[,m-1]
  ze<-z*matrix(rep(e,dim(z)[2]),ncol = dim(z)[2],nrow = n)
  W=solve(t(ze)%*%ze/n)
  betaa=cbind(betaa,solve(xz%*%W%*%t(xz),xz%*%W%*%zy))
}

#plot
plot(1:20, betaa[1,], type = "b", frame = TRUE, pch = 19,
     col = "red", xlab = "iteration", ylab = "estimation",ylim = c(-0.3,0.3))
lines(1:20, betaa[2,], pch = 18, col = "blue", type = "b", lty = 2)
lines(1:20, betaa[3,], pch = 17, col = "orange", type = "b", lty = 2)
lines(1:20, betaa[4,], pch = 15, col = "purple", type = "b", lty = 2)
lines(1:20, betaa[5,], pch = 14, col = "black", type = "b", lty = 2)
lines(1:20, betaa[6,], pch = 13, col = "green", type = "b", lty = 2)
```




```
convcheckera<-c()
for (b in 1:19) {
  convcheckera[b]=norm(betaa[,b]-betaa[,b+1],type = "2")
}
#b
y<-as.matrix(logwage)
x<-as.matrix(cbind(edu,exper,expersq100,black,south,urban,matrix(1,nrow(y),1)))

z<-as.matrix(cbind(pub,priv,intact1,intact2,exper,expersq100,black,south,urban,matrix(1,nrow(y),1)))
zz<-t(z)%*%z
xz<-t(x)%*%z
zy<-t(z)%*%y
n <- nrow(y)
betab=matrix(solve(xz%*%solve(zz,t(xz)),xz%*%solve(zz,zy)),ncol = 1)

for (m in 2:20) {

  e<-y-x%*%betab[,m-1]
  ze<-z*matrix(rep(e,dim(z)[2]),ncol = dim(z)[2],nrow = n)
  W<-solve(t(ze)%*%ze/n)
  betab=cbind(betab,solve(xz%*%W%*%t(xz),xz%*%W%*%zy))
}
convcheckerb<-c()
for (b in 1:19) {
  convcheckerb[b]=norm(betab[,b]-betab[,b+1],type = "2")
}
```

```r
#c
#for a
y<-as.matrix(logwage)
x<-as.matrix(cbind(edu,exper,expersq100,black,south,urban,matrix(1,nrow(y),1)))
z<-as.matrix(cbind(pub,priv,exper,expersq100,black,south,urban,matrix(1,nrow(y),1)))
zz<-t(z)%*%z
xz<-t(x)%*%z
zy<-t(z)%*%y
n <- nrow(y)

ba<-betaa[,20]
e<-y-x%*%ba
ze<-z*matrix(rep(e,dim(z)[2]),ncol = dim(z)[2],nrow = n)
W<-solve(t(ze)%*%ze/n)
Ja<-n*t(t(z)%*%(e)/n)%*%W%*%(t(z)%*%(e)/n)

#for b
y<-as.matrix(logwage)
x<-as.matrix(cbind(edu,exper,expersq100,black,south,urban,matrix(1,nrow(y),1)))

z<-as.matrix(cbind(pub,priv,intact1,intact2,exper,expersq100,black,south,urban,matrix(1,nrow(y),1)))
zz<-t(z)%*%z
xz<-t(x)%*%z
zy<-t(z)%*%y
n <- nrow(y)
bb<-betab[,20]

e<-y-x%*%bb
ze<-z*matrix(rep(e,dim(z)[2]),ncol = dim(z)[2],nrow = n)
W<-solve(t(ze)%*%ze/n)
Jb<-n*t(t(z)%*%(e)/n)%*%W%*%(t(z)%*%(e)/n)
```