HW Answers: Instrumental Variables Simulation

Tytus Wilam

```
set.seed(1234)

### Answer 1(a)
# sidenote: I choose to generate the dataset as a Secular Supreme Being

c <- c(rep("never",600), rep("complier", 250), rep("always",150))
# values -- {"never": never taker, "complier": complier, "always": always taker}
d0 <- ifelse(c=="never" | c=="complier", 0, 1) #c(rep(0, 850), rep(1,150))
d1 <- ifelse(c=="never", 0, 1) #c(rep(0,600), rep(1,400))


### Answer 1(b)
# We removed the group that makes choices inverse to
# the assignment, the defiers, because we assume
# monotonicity: D(1) > D(0)

### Answer 1(c)
# (i) The exclusion restriction says that if your treatment
# wouldn't be different even if your instrument assignment
# was different, then your outcome (or, more generally, the
# distribution of your outcome) also won't be different

y0 <- c(rnorm(600,0,1), rnorm(250, 3, 1), rnorm(150,6,1))
y1 <- c(rnorm(600,0,1), rnorm(250, 7, 1), rnorm(150,6,1))
df <- tbl_df(data.frame(c, d0, d1, y0, y1)) %>% group_by(c) %>% mutate(y=y1-y0)

### Answer 1(d)
SATE_complier <- summarize(df, mean(y))
```

```
### Answer 1(e)
# CACE = Complier Average Causal Effect

### Answer 1(f)
ITT <- mean(df$y1) - mean(df$y0)
ITT
```

```
## [1] 1.041105
```

```
### Answer 1(g)
dat.full <- select(df, c, d0, d1, y0, y1)

### Answer 2
z <- rbinom(1000, 1, 0.5)

### Answer 3
dat.obs <- ungroup(dat.full) %>% mutate(z) %>% mutate(d=ifelse(z==1, d1, d0)) %>% mutate(y=ifelse(d==1,
dat.obs
```

```
## # A tibble: 1,000 x 3
##        z     d      y
##    <int> <dbl>  <dbl>
##  1     0     0 -1.21
##  2     1     0  0.277
##  3     0     0  1.08
##  4     1     0 -2.35
##  5     0     0  0.429
##  6     1     0  0.506
##  7     1     0 -0.575
##  8     1     0 -0.547
##  9     0     0 -0.564
## 10     1     0 -0.890
## # ... with 990 more rows
```

```r
### Answer 4 (a)
## What I did:
# I calculate the percentage of subjects who underwent treatment
# in the treatment and no treatment group. The percentage of no treatment
# in the assigned group is my estimate of the percentage of never-takers in
# population and the percentage of treatet in the no-assignment group is my
# estimate of always-takers. I arrive at the percentage of compliers by subtracting
# the always- and never- takers from 100%.

## Code:
dat.obs %>% group_by(z) %>% summarize(mean(d))
```

```
## # A tibble: 2 x 2
##       z `mean(d)`
##   <int>     <dbl>
## 1     0     0.148
## 2     1     0.398
```

```r
## Estimates:
# always-takers: 14.8 %
# never-takers: 60.3 %
# compliers 24.9 %
```

```r
### Answer 4 (b) and (c)
# The naive regression is equivalent to average treatment effect for the population
# without distinguishing between always-takers, compliers, and never-takers

fit<-lm(y ~ d, dat.obs)
naive_TE <- fit$coefficients[2]
naive_TE
```

```
##        d
## 5.984244
```

```r
fit<-lm(y ~ z, dat.obs)
ITT_est <- fit$coefficients[2]
ITT_est
```

```
##         z
## 0.920938
```

```r
ITT_est.means <- mean(filter(dat.obs, z==1)$y) - mean(filter(dat.obs, z==0)$y)
ITT_est.means
```

```
## [1] 0.920938
```

```r
### Answer 4 (d)

CACE.est <- ITT_est/0.249
CACE.est
```

```
##         z
## 3.698546
```

```r
# CACE.est = 3.6985
```

```r
### Answer 4 (e)
fit <- lm(d ~ z, dat.obs)
d_predict <- fit$fitted.values
fit <- lm(y ~ d_predict, dat.obs)
CACE.2SLS <- fit$coefficients[2]
CACE.2SLS
```

```
## d_predict
##  3.694632
```

```r
# CACE.@SLS = 3.6946
```

```r
### Answer 4 (f)

fit <- ivreg(y ~ d, ~z, dat.obs)
CACE.ivreg.error <- fit$sigma
CACE.ivreg.error
```

```
## [1] 1.736125
```

```r
CACE.ivreg <- fit$coefficient[2]
CACE.ivreg
```

```
##         d
## 3.694632
```

```r
# CACE.ivreg = 3.6946
# residual standard error = 1.7361
```

```r
estimators <- tbl_df(data.frame(ivreg=numeric(0), ivreg.error=numeric(0)))
DGP <- function(dat.full){
z <- rbinom(1000, 1, 0.5)
dat.obs <- ungroup(dat.full) %>% mutate(z) %>% mutate(d=ifelse(z==1, d1, d0)) %>% mutate(y=ifelse(d==1,
fit <- ivreg(y ~ d, ~z, dat.obs)
CACE.ivreg.error <- fit$sigma
CACE.ivreg <- fit$coefficient[2]
c(CACE.ivreg, CACE.ivreg.error)
#estimators <- rbind(estimators, c(CACE.ivreg, CACE.ivreg.error)) #%>% add_row(ivreg=CACE.ivreg, ivreg.
#estimators
}

for(i in 1:1000){
  temp <- DGP(dat.full)
  estimators <- rbind(estimators, temp)
}

mean(estimators[,1])
```

```
## [1] 4.012671
```

```r
sd(estimators[,1])
```

```
## [1] 0.4323231
```

```r
# The estimator is unbiased. The mean is 3.9953 which is very close to 4.
# The standard deviation of the sampling distribution is 0.4323, which is
# smaller than the residual standard error of the original dataset, which
# was 1.7361
```

```r
### Answer 6

## Assumption 1: SUTVA
# Stable Unit Treatment Value Assumption
# Potential outcomes of each person are unaffected
# by the treatment status and outcomes of other persons.
# This assumption is likely violated in the proposed setting:
# (1) students might get more out of the math bootcamp if their
# friends also are assinged.
# (2) assuming that there is a fixed amount of money available for
# the bootcamp, it will yield different effects depending on the
# total proportion of compliers and always-takers in the population,
# i.e. the numebr of children who end up in the camp.

## Assumption 2: Random Assignment
# Instrument is randomly assigned. (i.e. ignrability of the instrument)
# This assumption is likely to be violated in the proposed setting
# because the group assigned the instrument (i.e. those who got the
# encouragement letter) are liekly to persuade their friends to join
# the camp with them. This will systematically increase the likelihood
# of treatment of those who have many friends.
```

```
## Assumpiton 3: Exclusion Restriction.
# Instrument affects the outcome only throuugh the treatment, i.e.
# if instrument were different but outcome stayed the same the outcome
# would also stay the same.
# This assumption might be violated althought it is difficult to
# assess the strength of the violation. For example, students who receive
# the encouragement might think that they have more aptitude than the
# students who were not encouraged (this is possible if students know about
# the emails other students receive). This improved self-perception could affect
# math study habits and performance regardless of the bootcamp.

## Assumption 4: Monotonicity.
# We expect no defiers but they are possible in principle.

## Assumption 5: Nonzero average causal effect of
## instrument on treatment.
# This assumption seems plausible in light of research on nudging.
# It seems plausible that even a single encouraging email would make
# a difference in the propensity of a student to enroll in the camp.
# If this assumption is violated we will see an identical d distribuiton
# in the encouraged and not-encouraged groups.

### Answer 6 (b)
# This violates ignorability. From the standpoint of the researcher
# it will seem like the exclusion restriction is also violated because it seems like Z
# influences the outcome independently of its influence on treatment.

### Answer 6(c)
# This would violate the exclusion restriction. Never takers do not
# take treatment, so any change between Y(Z=0, D=0) and Y(Z=1, D=0) is the causal
# influence of the instrument rather than the treatment.

### Answer 6(d)
# c <- c(rep("defy",600), rep("complier", 250), rep("always",150))
# d0 <- ifelse(c=="complier", 0, 1) #c(rep(0, 850), rep(1,150))
# d1 <- ifelse(c=="defy" | c=="never", 0, 1) #c(rep(0,600), rep(1,400))

### Answer 6(e)
# Alway takers are students who would participate in the math bootcamp
# regardless of receiving encouragement. We could exclude them from the study in one
# of two ways:
# (1) by lookin gat their characteristics (based on previous years) and not sending
# the encouragement letter to them. This might result, for example, in excluding the
# rich kids.
# (2) By administering the instrument after the initial deadline for participating in
# the bootcamp. Only those who received the letter after the deadline would be able to sign
# up. Presumably the always takers would have signed up by the deadline. This would only work
# if the study was kept a secret.
```