

# **mensik\_assignment\_4**

## **Table of contents**

<b>1</b>	<b>1. Sesame Street Experiment</b>	<b>1</b>
1.1	1. Four complier types . . . . .	1
1.2	2. Monotonicity . . . . .	2
1.3	3. Estimate of LATE of $Y \leftarrow D$ instrumented by $Z$ , & Discussion . . . . .	2
<b>2</b>	<b>2. Violated exclusion restrictions</b>	<b>4</b>
2.1	strong & valid exclusion . . . . .	5
2.2	weak & valid exclusion . . . . .	5
2.3	strong & invalid exclusion . . . . .	6
2.4	weak & invalid exclusion . . . . .	6

```
data <- read.dta13("/Users/kristinamensik/Documents/GitHub/regression-study/sesame-1.dta")
```

## **1 1. Sesame Street Experiment**

### **1.1 1. Four complier types**

The four complier types describe how individuals respond to the treatment assignment in an experiment: compliers, defiers, always takers, and never takers.

In this experiment, even though parents parents are randomized, the treatment itself is watching Sesame Street. Thus, compliers are kids who watch Sesame Street when their parents encourage them to do so. Defiers are kids who do the opposite - they don't watch the treatment when they're encouraged to, and instead watch only when they are not encouraged to (this is easy to imagine). Always takers are kids who watch Sesame Street regardless of whether their parents encourage them, and never takers are kids who do not watch Sesame Street when their parents encourage them to and also when their parents don't (probably readers - nerds).

Under the independence assumption and monotonicity assumption (ie if there were no defiers), the proportion of never and always takers can be estimated. Always takers can be estimated as the proportion of kids who watched Sesame Street in the control group (since they would watch it regardless of encouragement). Never takers can be estimated as the proportion of kids who did not watch Sesame Street in the treatment group (since they would not watch it regardless of encouragement). Defiers are a problem.

## 1.2 2. Monotonicity

Monotonicity requires that the instrument leads all individual units to respond in the same direction regarding treatment uptake. In this context, it means that no child would watch Sesame Street when not encouraged but refuse to watch it when encouraged.

I can imagine that in this experimental context especially (rebellious children), this assumption is challenged by the presence of defiers, who always violate monotonicity. They're problematic in any context because they're not readily detectable and can skew the results by effectively reversing the treatment effect, so IV estimator would not for sure estimate a weighted average of the underlying causal effects.

Here, it is very easy to imagine children doing the opposite of what their parents encourage them to do, especially since the treatment is ongoing rather than a one-time thing. By this I mean I can imagine some kids responding to the pattern of their parents encouraging them to watch sesame street, but wanting to watch so only doing so when they're parents don't encourage them. On the other hand, I could imagine reactionary kids being more likely to be never-takers instead: getting encouraged to watch once is enough to dispell their interest in the show.

## 1.3 3. Estimate of LATE of Y <- D instrumented by Z, & Discussion

```
summary(lm(watched ~ encouraged, data = data))
```

Call:

```
lm(formula = watched ~ encouraged, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.90789	0.09211	0.09211	0.09211	0.45455

Coefficients:

Estimate	Std. Error	t value	Pr(> t )
----------	------------	---------	----------

```

(Intercept) 0.54545    0.04060   13.434 < 2e-16 ***
encouraged  0.36244    0.05102   7.104  1.4e-11 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3809 on 238 degrees of freedom
Multiple R-squared: 0.1749, Adjusted R-squared: 0.1715
F-statistic: 50.46 on 1 and 238 DF, p-value: 1.397e-11

```

```

m1 <- ivreg(letters ~ watched | encouraged, data = data)

summary(m1, diagnostics = TRUE)

```

Call:  
`ivreg(formula = letters ~ watched | encouraged, data = data)`

Residuals:

Min	1Q	Median	3Q	Max
-20.593	-9.593	-4.527	10.723	34.473

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	20.593	3.659	5.628	5.11e-08 ***
watched	7.934	4.606	1.723	0.0863 .

Diagnostic tests:

	df1	df2	statistic	p-value
Weak instruments	1	238	50.464	1.4e-11 ***
Wu-Hausman	1	237	1.328	0.25
Sargan	0	NA	NA	NA

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.46 on 238 degrees of freedom  
Multiple R-Squared: 0.1355, Adjusted R-squared: 0.1318  
Wald test: 2.967 on 1 and 238 DF, p-value: 0.08626

```

coeftest(m1, vcov = vcovHC(m1, type = "HC1"))

```

```

t test of coefficients:

            Estimate Std. Error t value Pr(>|t|)
(Intercept) 20.5928     3.6965  5.5709 6.835e-08 ***
watched      7.9340     4.6332  1.7124   0.08812 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

M1 is a just-identified IV model, so the coefficient on watched is ‘LATE:’ the local average treatment effect of watching Sesame Street on the literacy test scores for compliers (kids who watch Sesame Street when encouraged by their parents) under the monotonicity assumption. We see that the predicted literacy test score for kids who did not watch sesame street is 20.59, and that watching sesame street increases this predicted score by 7.9 points on average for compliers. The coefficient and significant tests are about the same with and without robust standard errors, and the effect is not statistically significant ( $p = 0.088$ ).

Looking at the diagnostic tests, the weak instruments test gives an F stat (50.46) that’s above the rule of thumb cutoff of 10 (although this cutoff is debated) so if we buy that cutoff rule, we can be reasonably sure that the instrument is not weak. The Wu-Hausman test gives a p-value of 0.25, which means we fail to reject the null hypothesis – although it does not prove watched is exogenous, it means we can’t detect endogeneity with the given sample size and noise. We get NA on the Sargan test because we have a just-identified (one instrument, 1 regressor) model.

It might be worth noting that it’s still possible that parents who do encourage their kid to watch sesame street when told to do other things that are unobserved (maybe, some parents want to compensate for encouraging their kid to watch TV by reading to them more at night), which would violate the exclusion restriction. But random assignment of encouragement should give reasonable assurance about independence/no confounders.

## 2 2. Violated exclusion restrictions

DGP

```

set.seed(12345)

mkdata <- function(d.to.y, z.to.d, u.to.y, u.to.d, z.to.y, N){
  Z <- rnorm(N) ## instrument
  U <- rnorm(N) ## unobs. confounder
  D <- Z * z.to.d + U * u.to.d + rnorm(N, sd=1)
  Y <- D * d.to.y + U * u.to.y + Z * z.to.y + rnorm(N, sd=1)
  data.frame(D = D, Z = Z, U = U, Y = Y)
}

```

```

}

# True causal effect

```

Dimensions: -strong: large z.to.d -weak: small z.to.d -valid exclusion: z.to.y = 0 -invalid exclusion: z.to.y != 0

Monte Carlo study 1.tell the DGP 2.for given scenario (a) simulate dataset (b) estimage IV model, (c) store coefs 3. repeat some number of times 4. look at AVE bias across simulations and spread

## 2.1 strong & valid exclusion

```

nsim <- 1000
N <- 1000

sims <- rep(NA, nsim)
for(i in 1:nsim){
  d_sim <- mkdata(d.to.y = 1, z.to.d = 0.4, u.to.y = 0.5,
    u.to.d = 0.5, z.to.y = 0, N = N)
  m <- ivreg(Y ~ D | Z, data = d_sim)
  m <- coeftest(m, vcov = vcovHC(m, type = "HC1"))
  sims[i] <- m["D",1]
}
mean(sims) - 1

```

[1] -0.001972431

-0.001972431

## 2.2 weak & valid exclusion

```

sims <- rep(NA, nsim)
for(i in 1:nsim){
  d_sim <- mkdata(d.to.y = 1, z.to.d = 0.04, u.to.y = 0.5,
    u.to.d = 0.5, z.to.y = 0, N = N)
  m <- ivreg(Y ~ D | Z, data = d_sim)
  m <- coeftest(m, vcov = vcovHC(m, type = "HC1"))

```

```

  sims[i] <- m["D",1]
}
mean(sims) - 1

```

[1] -0.0896712

## 2.3 strong & invalid exclusion

```

sims <- rep(NA, nsim)
for(i in 1:nsim){
  d_sim <- mkdata(d.to.y = 1, z.to.d = 0.4, u.to.y = 0.5,
    u.to.d = 0.5, z.to.y = 0.3, N = N)
  m <- ivreg(Y ~ D | Z, data = d_sim)
  m <- coeftest(m, vcov = vcovHC(m, type = "HC1"))
  sims[i] <- m["D",1]
}
mean(sims) - 1

```

[1] 0.7550596

## 2.4 weak & invalid exclusion

```

nsim <- 1000
N <- 1000

sims <- rep(NA, nsim)
for(i in 1:nsim){
  d_sim <- mkdata(d.to.y = 1, z.to.d = 0.04, u.to.y = 0.5,
    u.to.d = 0.5, z.to.y = 0.3, N = N)
  m <- ivreg(Y ~ D | Z, data = d_sim)
  m <- coeftest(m, vcov = vcovHC(m, type = "HC1"))
  sims[i] <- m["D",1]
}
mean(sims) - 1

```

[1] 4.677162

In this simulation, a strong instrument with a valid exclusion restriction performs well: the IV estimator recovers the true effect with essentially no bias ( $\sim 0.002$ ). When the ER is valid but the instrument is weak, the estimator becomes noisier and produces a modest bias ( $\sim 0.09$ ). When the ER is violated, the estimator changes much more: even with a strong instrument, we see large systematic bias ( $\sim +0.76$ ). The worst case happens with a violated assumption and weak instrument, where bias is 4.67. This shows that violating the exclusion restriction can have a much larger impact on the IV estimator than instrument strength, but both are important.