

# Programming Assignment 5

Andrew Shaw

2025-11-06

**Problem 1:** Without estimating what is the effect of Z on D? Is this relationship causal? Why or why not?

The effect of Z on D is around 0.59 because 60% of the instrument are compliers and 1% are defiers and these are the only two categories that contribute to the change in treatment. Yes, the relationship is causal because Z doesn't affect y directly nor does it affect any of the error terms.

**Problem 2:** Now estimate the reduced form estimate of Z on y using lm() function. Is this relationship causal? Why or why not?

```
reduced_form <- lm(y ~ Z, data = mydata)
reduced_form
```

```
##
## Call:
## lm(formula = y ~ Z, data = mydata)
##
## Coefficients:
## (Intercept)          Z
##       0.16716     0.01785
```

Yes, this relationship is also causal but indirectly as Z has no direct path to y but affects y through D which was used in generating y satisfying the exclusion restriction.

**Problem 3:** Use the function ivreg to estimate the relationship between D and y. Is this relationship causal? Why or why not?

```
iv_model <- ivreg(y ~ D | Z, data = mydata)
iv_model
```

```
##
## Call:
## ivreg(formula = y ~ D | Z, data = mydata)
##
## Coefficients:
## (Intercept)          D
##       0.16078     0.03029
```

Yes, this relationship is causal as well because the relationship between D and y examined only comes from the variation from D that happened because of Z.

#### Problem 4: What happens to the estimate as the percent of defiers grows?

```

complier <- ifelse(Xsplit < 0.6, 1, 0)
always <- ifelse(Xsplit >= 0.6 & Xsplit < 0.8, 1, 0)
never <- ifelse(Xsplit >= 0.8 & Xsplit < 0.90, 1, 0)
defier <- ifelse(Xsplit >= 0.90, 1, 0)
D <- ifelse((Z == 1 & complier == 1) | (always == 1) |
(Z == 0 & defier == 1), 1, 0)
error2 <- rnorm(n, 0, 0.05)
y <- ifelse(complier == 1, 0.2 + 0.03 * D + 0.01 * Xsplit + error2,
0.1 + 0.01 * D + 0.01 * Xsplit + error2
)
mydata <- data.frame(y, D, Z, complier)

iv_reg_defiers <- ivreg(y ~ D | Z, data = mydata)
iv_reg_defiers

```

```

##
## Call:
## ivreg(formula = y ~ D | Z, data = mydata)
##
## Coefficients:
## (Intercept)          D
##      0.15786     0.03391

```

If the proportion of defiers were to increase, the coefficient for the treatment variable increases distorting the effect of treatment on y because now the first-stage effects average shrinks (e.g.  $0.60 - 0.10 = 0.50$ ) the denominator of the IV coefficient incorrectly inflates the coefficient.

**Problem 5:** We somehow have information about who is a complier. If we control for complier group and re-estimate the ivreg from problem 3, what happens? If we limit to only complier group what happens? Is no/one/or both causal? Explain.

```

iv_control <- ivreg(y ~ D + complier | Z + complier, data = mydata)
iv_control

```

```

##
## Call:
## ivreg(formula = y ~ D + complier | Z + complier, data = mydata)
##
## Coefficients:
## (Intercept)          D    complier
##      0.09765     0.03053     0.10505

```

Controlling for compliers, there is a change in the baseline intercept and small change to the IV coefficient. The complier coefficient is a description of the average change y as compared to other groups.

```
compliers_only <- subset(mydata, complier == 1)
iv_compliers <- ivreg(y ~ D | Z, data = compliers_only)
iv_compliers
```

```
##
## Call:
## ivreg(formula = y ~ D | Z, data = compliers_only)
##
## Coefficients:
## (Intercept)          D
##       0.20295     0.03002
```

If we subset the data to only those whose status is affirmative for compliers, incorrect assignments from always-takers and defiers is stripped out so the result is an even more precise estimate revealing the pure causal effect of Z on y.