Assignment 5 - Grant Jackson

Grant Jackson

2024-11-04

Theory Questions

Question 1:

i: The exclusion restriction might fail because being near a community college could affect wages directly rather than just through post-secondary education. Being near a community college may bring more employment opportunities, affecting wages directly. Areas with a community college could have better economic conditions or industries that affect wages directly.

ii: The exclusion restriction might fail because implementing a school funding tax could signal other pre-existing factors that directly impact house prices. Communities that implement a school funding tax probably have wealthier residents who can afford the tax, and wealthier residents could impact higher house prices directly. These communities could also have greater community investment priority, making the area more appealing and directly raising house prices.

iii: The exclusion restriction might fail because a parent's wealth likely drives other factors that directly impact a person's level of happiness. A parents wealth while raising a child could directly impact their level of happiness for their life. It could also affect the level of education, healthcare, and opportunities in childhood, which impacts their happiness.

Question 2:

Weak instruments mean that 'Zi' is weakly correlated with the endogenous 'Xi'. In the first stage, this leads to inaccurate Xi predictions. In the second stage, the inaccurate Xi's lead to non-normal sampling distribution of the 2SLS estimator. This should also bias the estimator towards the first stage OLS regression. All of these effects make it near impossible to make reliable statistical conclusions, even with large sample sizes.

Coding Exercise

Question 1:

```r
model <- feols(
  pt_id_with_lean ~ treat,
  data = data,
  vcov = 'HC1'
)
etable(model, fitstat = c("n", "f"))
```

```
##                               model
## Dependent Var.:     pt_id_with_lean
##
## Constant         0.1814*** (0.0270)
## treat              0.0795. (0.0408)
```

```
## ---------------- ------------------
## S.E. type       Heteroskedas.-rob.
## Observations                   411
## F-test                      3.7830
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation: The F-statistic of 3.780 is concerning because it is below the rule of thumb of F>10 for a strong instrument. This indicates we have a weak instrument.

Question 2:

```r
feols(
  pt_voteevalalignindex ~ treat,
  data = data,
  vcov = 'HC1'
)
```

```
## OLS estimation, Dep. Var.: pt_voteevalalignindex
## Observations: 411
## Standard-errors: Heteroskedasticity-robust
##             Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) 2.348039   0.118168 19.87027 < 2.2e-16 ***
## treat       0.279980   0.168995  1.65674  0.098339 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 1.70918   Adj. R2: 0.004235
```

2SLS Calculation:

```r
s1_te <- 0.0795

s2_te <- 0.279980

estimator = s2_te / s1_te

estimator
```

```
## [1] 3.521761
```

Question 3:

```r
feols(
  pt_voteevalalignindex ~ 1| pt_id_with_lean~treat,
  data = data,
  vcov = 'HC1'
)
```

```
## TSLS estimation - Dep. Var.: pt_voteevalalignindex
##                  Endo.    : pt_id_with_lean
##                  Instr.   : treat
## Second stage: Dep. Var.: pt_voteevalalignindex
## Observations: 411
## Standard-errors: Heteroskedasticity-robust
##                   Estimate Std. Error t value  Pr(>|t|)
## (Intercept)        1.70926   0.567150 3.01378 0.0027407 **
```

```
## fit_pt_id_with_lean  3.52189   2.501995 1.40763 0.1599988
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 2.01056   Adj. R2: -0.377896
## F-test (1st stage), pt_id_with_lean: stat = 3.78297, p = 0.052462, on 1 and 409 DoF.
##                         Wu-Hausman: stat = 1.65162, p = 0.199468, on 1 and 408 DoF.
```

```r
summary(data$pt_voteevalalignindex)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -3.000   1.000   3.000   2.489   4.000   6.000
```

Interpretation: The coefficient exactly matches our hand calculation. With the p-value>0.10, this estimate is not statistically significant. This was expected given the weak results in the first stage.

Question 4:

```r
# Getting population means for comparison
pop_means <- data.frame(
  variable = c("age", "voted2006", "pt_church"),
  pop_mean = c(
    mean(data$age, na.rm = TRUE),
    mean(data$voted2006, na.rm = TRUE),
    mean(data$pt_church, na.rm = TRUE)
  )
)

# Getting complier means for each characteristic
# For age
age_compliers <- feols(
  I(pt_id_with_lean * age) ~ 1 | pt_id_with_lean ~ treat,
  data = data,
  vcov = 'HC1'
)

# For 2006 voting
vote_compliers <- feols(
  I(pt_id_with_lean * voted2006) ~ 1 | pt_id_with_lean ~ treat,
  data = data,
  vcov = 'HC1'
)

# For church attendance
church_compliers <- feols(
  I(pt_id_with_lean * pt_church) ~ 1 | pt_id_with_lean ~ treat,
  data = data,
  vcov = 'HC1'
)
```

```
## NOTE: 12 observations removed because of NA values (LHS: 12).
```

```r
# Comparing complier means to population means
results <- data.frame(
  variable = c("age", "voted2006", "pt_church"),
  complier_mean = c(
    coef(age_compliers)[1],
```

```r
    coef(vote_compliers)[1],
    coef(church_compliers)[1]
  )
)

# Merging with population means
comparison <- merge(pop_means, results, by = "variable")
print(comparison)
```

```
##    variable   pop_mean complier_mean
## 1       age 38.9148418  -1.728328865
## 2 pt_church  0.4088346   0.009896468
## 3 voted2006  0.6228710  -0.063151623
```

Interpretation: The compiler means indicate that those who respond to the treatment by registering are systematically different from the average voter by being younger, less religious, and less politically engaged in the past. IT is safe to assume the effectiveness of the treatment might vary substantially across different demographic groups. We should not generalize our LATE estimate to the broader population.