Problem Set 6: Instrumental Variables

GOV 2003

Due at 11:59 pm (ET) on Oct 27, 2021

Instruction

Before you begin, please read the following instructions **carefully**:

- **No late submission is allowed** without prior approval from the instructors.
- **All answers should be typed up**. We recommend the use of `Rmarkdown`. A `Rmarkdown` template for this problem set is provided. Answers to analytical solutions should also be typed up.
- **A PDF copy of your answer** including your computer code should be uploaded to `Gradescope` before the deadline. **Do not submit the markdown file itself**.
- This problem set includes a bonus question for extra credit. No deduction in the total points will be made from this question. Note that the maximum points of this problem set is 15 points. That is, if the student receives 3 points from the bonus question and 14 points from the other questions, the total points will be 15 points.

Introduction

This problem set consists of two parts. In `Part A`, we will revisit Kalla and Broockman (2020) to study **randomized experiment with noncompliance** and analytically investigate the violations of exclusion restriction and monotonicity assumption. In `Part B`, we will replicate the analysis of Baccini and Weymouth (2021) using **instrumental variables**.

Part A: Randomized experiment with noncompliance

Setup 1

Let Y_i denote the support for inclusive policies for each voter i and Z_i denote the assignment of canvassing with two categories — assignment of *placebo* conversation unrelated to immigration ($Z_i = 0$) and that of *full intervention* including non-judgemental exchange of narratives ($Z_i = 1$). Here, we assume no interference between units (voters) and unit-level complete randomization of the canvassing (i.e., ignore the household-level random assignment from the original studies for simplicity). We only consider the individuals who open their doors and identify themselves before the intervention and placebo scripts diverge.

Note that not all individuals continue with the full intervention. To incorporate the presence of individuals who would not receive the entire intervention, we introduce a new variable $D_i \in \{0, 1\}$ to denote the non-judgemental exchange of narratives between the respondent and canvasser. We assume **one-sided noncompliance** throughout the question. Observe the following potential outcomes $D_i(z)$ and the two compliance types.

- $D_i(1) = 1$ if assigned to the full intervention, respondent i would complete the entire non-judgemental exchange of narratives (**compliers**).
- $D_i(1) = 0$ if assigned to the full intervention, respondent i would *not* complete the entire non-judgemental exchange of narratives (**noncompliers**).
- $D_i(0) = 0$ for any i due to the one-sided noncompliance, meaning that once assigned to placebo conversation there would be *no* non-judgemental exchange of narratives between the respondent and canvasser.

Name	Description
id	a unique ID of voters
type	canvassing with two categories (Placebo or Full Intervention)
narratives	= 1 if the individual completed the non-judgemental exchange of narratives
factor	a pooled outcome index that captures the support for inclusive policies

Question 1 (5 pts; 1 pt for each)

(a) Write down the three assumptions for identifying LATE using Z_i as an instrument variable and interpret them in the context of this study.

Hint: Review the lecture slides p.7 and p.11.

(b) Estimate the proportion of compliers using the data frame `canvass` in `IV.RData`.

Hint: Recall that $\text{ITT}_D = \pi_{\text{co}}$, the proportion of compliers, under one-sided noncompliance.

(c) Estimate τ_{LATE} and its standard error using Wald estimator and its asymptotic variance.

(d) Estimate τ_{LATE} and its standard error using two-stage least squares estimator and a robust variance estimator.

Hint: You may use the `AER::ivreg()` function for fitting the model, and `ivpack::robust.se()` function for estimating a robust standard error.

(e) Compare two results from (c) and (d). Briefly discuss the results.

Answer 1

(a)

- Randomization: $[\{Y_i(d,z), \forall d,z\}, D_i(1), D_i(0)] \perp\!\!\!\perp Z_i$. The assignment of canvassing is randomized and thus independent of the potential outcomes (the support for inclusive policies) and compliance types.
- First-stage: $\text{ITT}_D = \pi_{\text{co}} \neq 0$. At least one respondent complete the entire nonjudgemental exchange of narratives.
- Exclusion restriction: Z_i only affects Y_i through D_i. The assignment of canvassing only affects the support for inclusive policies through the nonjudgemental exchange of narratives.

(b)

```
load("IV.RData")
# Proportion of compliers (using ITT_D)
canvass$Z = ifelse(canvass$type == "Full Intervention", 1, 0)
canvass$D = canvass$narratives
pi_co <- mean(canvass$D[canvass$Z == 1]) - mean(canvass$D[canvass$Z == 0])
pi_co
```

```
## [1] 0.6800643
```

(c)

```
canvass$Y = canvass$factor
# Compute ITT's
ITT_Y <- mean(canvass$Y[canvass$Z == 1]) - mean(canvass$Y[canvass$Z == 0])
ITT_D <- mean(canvass$D[canvass$Z == 1]) - mean(canvass$D[canvass$Z == 0])
# (ITT_D = pi_co)

## Wald estimator
Wald_est <- ITT_Y/ITT_D
Wald_est
```

```
## [1] 0.1350285
```

```
## Variance
n1 = sum(canvass$Z)
n0 = sum(1-canvass$Z)
# Compute variance terms using neyman estimator
Var_ITT_Y <- var(canvass$Y[canvass$Z == 0])/n0 + var(canvass$Y[canvass$Z == 1])/n1
Var_ITT_D <- var(canvass$D[canvass$Z == 0])/n0 + var(canvass$D[canvass$Z == 1])/n1

# Compute covariance term
# demean
demeaned_y <- canvass$Y[canvass$Z == 1] - mean(canvass$Y[canvass$Z == 1])
demeaned_d <- canvass$D[canvass$Z == 1] - mean(canvass$D[canvass$Z == 1])
```

```r
# denominator
denom <- sum(canvass$Z)*(sum(canvass$Z) - 1)
Covar_ITT <- (demeaned_y %*% demeaned_d)/denom

Var_Wald_est <- Var_ITT_Y/ITT_D^2 + (ITT_Y^2*Var_ITT_D)/ITT_D^4 - (2*ITT_Y/ITT_D^3)*Covar_ITT
Var_Wald_est
```

```
##             [,1]
## [1,] 0.007023553
```

```r
sqrt(Var_Wald_est)
```

```
##            [,1]
## [1,] 0.08380664
```

(d)

```r
## TSLS estimator
library(AER)
```

```
## Loading required package: car

## Loading required package: carData

## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

## Loading required package: survival
```

```r
library(ivpack)
ivmodel <- ivreg(Y ~ D | Z, data = canvass)
ivmodel$coefficients["D"]
```

```
##         D
## 0.1350285
```

```r
## Variance
se <- robust.se(ivmodel)
```

```
## [1] "Robust Standard Errors"
```

```r
se[,"Std. Error"]["D"]
```

```
##          D
## 0.08373892
```

(e) Two estimators yield the same estimates since the treatment (D_i) and instrument (Z_i) are all binary.

Question 2 (2 pts; 1 pt for each)

Now, in this question, we analytically investigate **two-sided noncompliance** in a randomized experiment. Suppose Z_i is an encouragement of some treatment D_i. Observe the following four compliance types (or principal strata):

- Complier: $D_i(1) = 1$ and $D_i(0) = 0$
- Noncomliers:
  - Always-taker: $D_i(1) = D_i(0) = 1$
  - Never-taker: $D_i(1) = D_i(0) = 0$
  - Defier: $D_i(1) = 0$ and $D_i(0) = 1$

Let π_{co}, π_{at}, π_{nt}, and π_{de} be the proportions of each type.

(a) Write down four canonical IV assumptions.

(b) Let $\tau_{\text{iv}} = \frac{\text{ITT}_Y}{\text{ITT}_D}$ be IV estimand and $\text{ITT}_{Y,\text{co}} = \frac{1}{n_{\text{co}}} \sum_{i:C_i=\text{co}} Y_i(1, D_i(1)) - Y_i(0, D_i(0))$ be LATE. Suppose that the exclusion restriction is violated. Show that we can write the difference between the two estimands as

$$\tau_{\text{iv}} - \text{ITT}_{Y,\text{co}} = \text{ITT}_{Y,\text{at}} \frac{\pi_{\text{at}}}{\pi_{\text{co}}} + \text{ITT}_{Y,\text{nt}} \frac{\pi_{\text{nt}}}{\pi_{\text{co}}}.$$

When would using IV estimator be particularly problematic?

Hints:

- Observe that under monotonicity assumption,

$$\begin{aligned}\text{ITT}_D &= \underbrace{\text{ITT}_{D,\text{co}}}_{=1} \pi_{\text{co}} + \underbrace{\text{ITT}_{D,\text{at}}}_{=0} \pi_{\text{at}} + \underbrace{\text{ITT}_{D,\text{nt}}}_{=0} \pi_{\text{nt}} + \text{ITT}_{D,\text{de}} \underbrace{\pi_{\text{de}}}_{=0 \text{ (mono)}} \\ &= \pi_{\text{co}}\end{aligned}$$

- Use the following decomposition:

$$\text{ITT}_Y = \text{ITT}_{Y,\text{co}}\pi_{\text{co}} + \text{ITT}_{Y,\text{at}}\pi_{\text{at}} + \text{ITT}_{Y,\text{nt}}\pi_{\text{nt}} + \text{ITT}_{Y,\text{de}}\pi_{\text{de}}$$

(c) [Bonus: 1pt] Suppose now that monotonicity is violated instead of the exclusion restriction. Show that we can write the difference between the estimands as

$$\tau_{\text{iv}} - \text{ITT}_{Y,\text{co}} = (\text{ITT}_{Y,\text{co}} + \text{ITT}_{Y,\text{de}}) \frac{\pi_{\text{de}}}{\pi_{\text{co}} - \pi_{\text{de}}}.$$

When do you expect this to be small?

Hints:

- Note that now $\text{ITT}_D \neq \pi_{\text{co}}$ anymore. Derive the correct expression of ITT_D.
- Again, use the following decomposition:

$$\text{ITT}_Y = \text{ITT}_{Y,\text{co}}\pi_{\text{co}} + \text{ITT}_{Y,\text{at}}\pi_{\text{at}} + \text{ITT}_{Y,\text{nt}}\pi_{\text{nt}} + \text{ITT}_{Y,\text{de}}\pi_{\text{de}}$$

Answer 2

(a)

- Randomization of Z_i
- Presence of some compliers $\pi_{\text{co}} \neq 0$ (first-stage)
- Exclusion restriction $Y_i(z,d) = Y_i(z',d)$
- Monotonicity: $D_i(1) \geq D_i(0)$ for all i (no defiers)

(b)

By the monotonicity assumption (no defiers; $\pi_{\text{de}} = 0$):

$$\begin{aligned}\text{ITT}_Y &= \text{ITT}_{Y,\text{co}}\pi_{\text{co}} + \text{ITT}_{Y,\text{at}}\pi_{\text{at}} + \text{ITT}_{Y,\text{nt}}\pi_{\text{nt}} + \text{ITT}_{Y,\text{de}}\pi_{\text{de}} \\ &= \text{ITT}_{Y,\text{co}}\pi_{\text{co}} + \text{ITT}_{Y,\text{at}}\pi_{\text{at}} + \text{ITT}_{Y,\text{nt}}\pi_{\text{nt}}\end{aligned}$$

Divide both sides of the equation by ITT_D:

$$\begin{aligned}\frac{\text{ITT}_Y}{\text{ITT}_D} &= \text{ITT}_{Y,\text{co}}\frac{\pi_{\text{co}}}{\text{ITT}_D} + \text{ITT}_{Y,\text{at}}\frac{\pi_{\text{at}}}{\text{ITT}_D} + \text{ITT}_{Y,\text{nt}}\frac{\pi_{\text{nt}}}{\text{ITT}_D} \\ &= \text{ITT}_{Y,\text{co}}\frac{\pi_{\text{co}}}{\pi_{\text{co}}} + \text{ITT}_{Y,\text{at}}\frac{\pi_{\text{at}}}{\pi_{\text{co}}} + \text{ITT}_{Y,\text{nt}}\frac{\pi_{\text{nt}}}{\pi_{\text{co}}} && [\text{by the first hint } (\text{ITT}_D = \pi_{\text{co}})] \\ &= \text{ITT}_{Y,\text{co}} + \text{ITT}_{Y,\text{at}}\frac{\pi_{\text{at}}}{\pi_{\text{co}}} + \text{ITT}_{Y,\text{nt}}\frac{\pi_{\text{nt}}}{\pi_{\text{co}}}\end{aligned}$$

Thus we have

$$\frac{\text{ITT}_Y}{\text{ITT}_D} - \text{ITT}_{Y,\text{co}} = \text{ITT}_{Y,\text{at}}\frac{\pi_{\text{at}}}{\pi_{\text{co}}} + \text{ITT}_{Y,\text{nt}}\frac{\pi_{\text{nt}}}{\pi_{\text{co}}} \quad \square$$

This quantity would be large if we have a small proportion of compliers. In other words, using this IV estimator would be particularly problematic in case of weak encouragement.

(c)

Observe that

$$\begin{aligned}\text{ITT}_D &= \mathbb{E}[D_i(1) - D_i(0)] \\ &= \mathbb{P}(D_i = 1 \mid Z_i = 1) - \mathbb{P}(D_i = 1 \mid Z_i = 0) \\ &= (\pi_{\text{co}} + \pi_{\text{at}}) - (\pi_{\text{at}} + \pi_{\text{de}}) \\ &= \pi_{\text{co}} - \pi_{\text{de}}.\end{aligned}$$

Meanwhile, under exclusion restriction, $\text{ITT}_{Y,\text{at}} = 0$ and $\text{ITT}_{Y,\text{nt}} = 0$. Thus,

$$\begin{aligned}\text{ITT}_Y &= \text{ITT}_{Y,\text{co}}\pi_{\text{co}} + \text{ITT}_{Y,\text{at}}\pi_{\text{at}} + \text{ITT}_{Y,\text{nt}}\pi_{\text{nt}} + \text{ITT}_{Y,\text{de}}\pi_{\text{de}} \\ &= \text{ITT}_{Y,\text{co}}\pi_{\text{co}} + \text{ITT}_{Y,\text{de}}\pi_{\text{de}}\end{aligned}$$

Divide both sides of the equation by ITT_D:

$$
\begin{aligned}
\frac{\mathrm{ITT}_Y}{\mathrm{ITT}_D} &= \mathrm{ITT}_{Y,\mathrm{co}} \frac{\pi_{\mathrm{co}}}{\pi_{\mathrm{co}} - \pi_{\mathrm{de}}} + \mathrm{ITT}_{Y,\mathrm{de}} \frac{\pi_{\mathrm{de}}}{\pi_{\mathrm{co}} - \pi_{\mathrm{de}}} \\
&= \mathrm{ITT}_{Y,\mathrm{co}} \left(\frac{\pi_{\mathrm{de}}}{\pi_{\mathrm{co}} - \pi_{\mathrm{de}}} + 1 \right) + \mathrm{ITT}_{Y,\mathrm{de}} \frac{\pi_{\mathrm{de}}}{\pi_{\mathrm{co}} - \pi_{\mathrm{de}}} \\
&= (\mathrm{ITT}_{Y,\mathrm{co}} + \mathrm{ITT}_{Y,\mathrm{de}}) \frac{\pi_{\mathrm{de}}}{\pi_{\mathrm{co}} - \pi_{\mathrm{de}}} + \mathrm{ITT}_{Y,\mathrm{co}}
\end{aligned}
$$

Thus we have

$$
\tau_{\mathrm{iv}} - \mathrm{ITT}_{Y,\mathrm{co}} = (\mathrm{ITT}_{Y,\mathrm{co}} + \mathrm{ITT}_{Y,\mathrm{de}}) \frac{\pi_{\mathrm{de}}}{\pi_{\mathrm{co}} - \pi_{\mathrm{de}}} \quad \square
$$

This quantity would be large if we have a small proportion of defiers.

Part B: Instrumental variables in observational study

Setup B

In this part, we will replicate the analysis of Baccini and Weymouth (2021). The goal of this paper is to study how different groups in society may construe manufacturing job losses in contrasting ways. Specifically, the authors examine the heterogeneous effect of manufacturing layoffs on voting behavior among white voters and voters of color. Since layoffs are not randomly assigned, the authors construct a Bartik instrument that relies on the sectoral composition of each county and industry-specific national trends in layoffs.[1] Essentially, this instrument is the predicted number of layoffs for a given group applying the rate of layoffs in all other counties to baseline employment In Question 3 and 4, we will replicate their county- and individual-level analysis respectively. Please read the paper p.554-559 for more details.

Question 3: County-level (6 pts; 1 pt for (a) and (c); 2 pts for (b) and (d))

In this question, use data frame `county2016` in `IV.RData` for analysis.

Name	Description
ddem_votes_pct1	change of democratic vote share
LAU_unemp_rate_4y	unemployment
pers_m_total_share_4y	share of male
pers_coll_share_4y	share of college educated people
white_counties_4y	white population share
msl_service_pc4y	service layoffs
id_state	state id
msl_pc4y2	manufacturing layoffs
msl_nw_pc4y2	non-white manufacturing layoffs
msl_w_pc4y2	white manufacturing layoffs
bartik_leo5	Bartik instrument (total)
bartik_leo5_w2	Bartik instrument (white)

(a) See equations (1) - (4) from the paper. Does this setup fit into the usual IV setup that we discussed in class? If not, how does it differ?

(b) Replicate county-level model (3) of Table 2 in Baccini and Weymouth (2021) using `ivreg()` function from the `AER` package. Here, note that the instrumental variable is bartik_leo5 (Bartik instrument (total)), the treatment variable is msl_pc4y2 (manufacturing layoffs), and the outcome variable is ddem_votes_pct1 (change of democratic vote share). Briefly discuss the results. Suppose you had dichotomized the instrument and the treatment with some cutoff values. What additional assumption(s) do you have to make in order to interpret the result as an average treatment effect (as opposed to a local average treatment effect)?

Hints:

[1] From the paper, a Bartik instrument is "formed by interacting initial values of some local industry feature (such as employment) with national industry growth rates."

- *Exogenous* covariates included in the county-level model (3) are as follows: LAU_unemp_rate_4y, pers_m_total_share_4y, pers_coll_share_4y, white_counties_4y, msl_service_pc4y, and state fixed effects using `factor(id_state)`
- Use `ivpack::robust.se()` function for estimating a robust standard error.

(c) Conduct F-test to check whether the instrument used in (b) is weak. Specify the null hypothesis and briefly explain the result.

Hint: See the sample codes below.

```
# TODO 1: Run first-stage
fs = lm(first-stage-formula, data = your-data)
# TODO 2: Run first-stage under null (i.e. exclude IV):
fn = lm(first-stage-formula-without-iv, data = your-data)
# Option 1. simple F-test
# TODO 3-1: Check F-statistic
lmtest::waldtest(fs, fn)
# Option 2. F-test robust to heteroskedasticity
# TODO 3-2: Check F-statistic
lmtest::waldtest(fs, fn, vcov = sandwich::vcovHC(fs, type="HC2"))
```

(d) Replicate county level model (6) of Table 2 in Baccini and Weymouth (2021) using `ivreg()` function from the `AER` package. Here, note that the instrument variable is bartik_leo5_w2 (Bartik instrument (white)), the treatment variable is msl_w_pc4y2 (white manufacturing layoffs), and the outcome variable is again ddem_votes_pct1 (change of democratic vote share). They include msl_nw_pc4y2 as an exogenous covariate (non-white manufacturing layoffs) as well. Can we interpret both of the estimated coefficients of msl_w_pc4y2 and msl_nw_pc4y2 as causal effects in this model? Why or why not? Do you agree or disagree with the following statement?

"In all models White Manufacturing Layoffs is negative and significant, whereas Non-White Manufacturing Layoffs is positive and significant. Taken together, the results suggest that manufacturing job losses may lead to different voting behavior across demographic lines (557)."

Hints:

- *Exogenous* covariates included in the county-level model (6) are as follows: msl_nw_pc4y2 and all the exogenous covariates specified in (b).
- Use `ivpack::robust.se()` function for estimating a robust standard error.

Answer 3

(a) Here, the treatment and the instrument variables are not binary. To claim our estimand to be a LATE, we may need to assume a linear effect in this model.

(b)

```
load("IV.RData")
library(AER)
tab2mod3 <- ivreg(ddem_votes_pct1 ~ # change of democratic vote share
```

```r
                    LAU_unemp_rate_4y + # unemployment
                    pers_m_total_share_4y + # share of male
                    pers_coll_share_4y + # share of college educated people
                    white_counties_4y + # white population share
                    msl_service_pc4y + # service layoffs
                    factor(id_state) + # state id
                    msl_pc4y2 |# manufacturing layoffs
                    . - msl_pc4y2 + bartik_leo5, #Bartik instrument (total)
                  data = county2016)
tab2mod3[["coefficients"]]["msl_pc4y2"]
```

```
##  msl_pc4y2
## -0.0432703
```

```r
se_tab2mod3 <- ivpack::robust.se(tab2mod3)
```

```
## [1] "Robust Standard Errors"
```

```r
se_tab2mod3["msl_pc4y2",]
```

```
##    Estimate  Std. Error     t value    Pr(>|t|)
## -0.04327030  0.01922165 -2.25112275  0.02444963
```

To interpret the result as an average treatment effect, we should assume a constant effect.

(c)

```r
# 1: Run first-stage
fs = lm(msl_pc4y2 ~ LAU_unemp_rate_4y + # unemployment
                    pers_m_total_share_4y + # share of male
                    pers_coll_share_4y + # share of college educated people
                    white_counties_4y + # white population share
                    msl_service_pc4y + # service layoffs
                    factor(id_state) + # state id
                    bartik_leo5, data = county2016)
# 2: Run first-stage under null (i.e. exclude IV):
fn = lm(msl_pc4y2 ~ LAU_unemp_rate_4y + # unemployment
                    pers_m_total_share_4y + # share of male
                    pers_coll_share_4y + # share of college educated people
                    white_counties_4y + # white population share
                    msl_service_pc4y + # service layoffs
                    factor(id_state) # state id
        , data = county2016)
# Option 1. simple F-test
# 3-1: Check F-statistic
F1 <- lmtest::waldtest(fs, fn)
F1$F[2]
```

```
## [1] 1537.565
```

```r
# Option 2. F-test robust to heteroskedasticity
# 3-2: Check F-statistic
lmtest::waldtest(fs, fn, vcov = sandwich::vcovHC(fs, type="HC0"))
```

```
## Wald test
##
## Model 1: msl_pc4y2 ~ LAU_unemp_rate_4y + pers_m_total_share_4y + pers_coll_share_4y +
##     white_counties_4y + msl_service_pc4y + factor(id_state) +
##     bartik_leo5
## Model 2: msl_pc4y2 ~ LAU_unemp_rate_4y + pers_m_total_share_4y + pers_coll_share_4y +
##     white_counties_4y + msl_service_pc4y + factor(id_state)
##   Res.Df Df      F    Pr(>F)
## 1   3010
## 2   3011 -1 477.18 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
library(estimatr)
fs2 <- lm_robust(msl_pc4y2 ~ LAU_unemp_rate_4y + # unemployment
                    pers_m_total_share_4y + # share of male
                    pers_coll_share_4y + # share of college educated people
                    white_counties_4y + # white population share
                    msl_service_pc4y + # service layoffs
                    factor(id_state) + # state id
            bartik_leo5
          , data = county2016, se_type = "HC0")
F2 <- lmtest::waldtest(fs, fn, vcov = fs2$vcov)
F2$F[2]
```

```
## [1] 477.1808
```

(d)

```r
tab2mod6 <- ivreg(ddem_votes_pct1 ~ # change of democratic vote share
                    LAU_unemp_rate_4y + # unemployment
                    pers_m_total_share_4y + # share of male
                    pers_coll_share_4y + # share of college educated people
                    white_counties_4y + # white population share
                    msl_service_pc4y + # service layoffs
                    factor(id_state) + # state id
                    msl_nw_pc4y2 + # non-white manufacturing layoffs
                    msl_w_pc4y2 |# white manufacturing layoffs
                    . - msl_w_pc4y2 + bartik_leo5_w2, #Bartik instrument (white)
                  data = county2016)
tab2mod6[["coefficients"]]["msl_w_pc4y2"]
```

```
## msl_w_pc4y2
##  -0.1508751
```

```r
se_tab2mod6 <- ivpack::robust.se(tab2mod6)
```

```
## [1] "Robust Standard Errors"
```

```r
se_tab2mod6["msl_w_pc4y2",]
```

```
##      Estimate    Std. Error       t value      Pr(>|t|)
## -1.508751e-01  3.605169e-02 -4.184967e+00  2.942895e-05
```

Answers will depend. Since msl_nw_pc4y2 has been included as an exogenous variable despite the possibility of an unmeasured confounder, its coefficient may not be able to be interpreted as a causal effect.

Question 4: Individual-level (2 pts for (a))

In this question, we will use a subset of the individual-level data (`indv2016_subdat`) to discuss the identification strategy and its implication. Note that this is a non-random subset of the original data (which has much larger sample size and takes a sufficient amount of time to fit the model).

Name	Description
vote_dem	Pr(voting for Clinton = 1)
white	white
gender	gender
educ	education
age	age
approval_sen1	senator approval
unempl_white	white ∗ unemployment
gender_white	white ∗ gender
educ_white	white ∗ education
age_white	white ∗ age
sen_white	white ∗ senator approval
male_white	white ∗ share of male
coll_white	white ∗ share of college educated people
white_counties_white	white ∗ white population share
it_white_serv	white ∗ service layoffs
pan_id	group(state_fips county_fips)
it_white	white ∗ manufacturing layoffs
it_white_instr	white ∗ Bartik instrument (total)

(a) Replicate individual-level model (3) of Table 3 in Baccini and Weymouth (2021) using `ivreg()` function from the `AER` package. Here, note that the instrument variable is `it_white_instr` (white respondent ∗ Bartik instrument (total)), the treatment variable is `it_white` (white respondent ∗ manufacturing layoffs), and the outcome variable is `vote_dem` (Pr(voting for Clinton = 1)). Suppose we got the same estimates shown in Table 3. Based on this result, do you agree or disagree with the following statement?

"The coefficient of the interaction between layoffs and white respondents is always negative and significant. This indicates that whites were less likely than non-whites to vote for Clinton in counties that had experienced more manufacturing layoffs (558)."

Hints:

- *Exogenous* covariates included in the individual-level model (3) are as follows: white, gender, educ, age, approval_sen1, unempl_white, gender_white, educ_white, age_white, sen_white, male_white, coll_white, white_counties_white, it_white_serv, and county fixed effects using `factor`(pan_id)
- Use `ivpack::cluster.robust.se()` to estimate clustered robust standard error (clustered by pan_id).

(b) [Bonus: 2pts] Can you think of any additional analysis using this data (either county or individual level) to further test the authors' argument?

Answer 4

(a)

```r
tab3mod3 <- ivreg(vote_dem ~
                    white +
                    unempl_white +
                    gender +
                    gender_white +
                    educ +
                    educ_white +
                    age +
                    age_white +
                    approval_sen1 +
                    sen_white +
                    male_white +
                    coll_white +
                    white_counties_white +
                    it_white_serv +
                    factor(pan_id) +
                    it_white |
                    . - it_white + it_white_instr,
                  data = indv2016_subdat)
tab3mod3[["coefficients"]]["it_white"]
```

```
##   it_white
## -0.4480132
```

```r
res <- ivpack::cluster.robust.se(tab3mod3, tab3mod3$model$`factor(pan_id)`)
```

```
## [1] "Cluster Robust Standard Errors"
```

```r
res["it_white",]
```

```
##   Estimate Std. Error    t value   Pr(>|t|)
## -0.4480132  1.1616101 -0.3856830  0.6997568
```

Answers will depend, but since the authors are interested in the heterogeneous effect of manufacturing layoff depending on the respondent's race, the way treatment and instrument variables are specified in this model may not be proper to test the proposed argument.

(b) Answers will depend. In county-level analysis, ideally, we may want to run the separate model with manufacture layoff as the treatment and white and non-white vote share as the outcome and compare the causal effects. Since the county-level democratic vote share by race is not available, we can try to estimate it using available data (cf. racial polarized voting). In individual-level analysis, we may fit the model separately for white, and non-white subgroups and compare the estimated causal effects.

References

Baccini, L. and Weymouth, S. (2021). Gone for good: Deindustrialization, white voter backlash, and us presidential voting. *American Political Science Review*, 115(2):550–567.

Kalla, J. and Broockman, D. (2020). Reducing exclusionary attitudes through interpersonal conversation: Evidence from three field experiments. *American Political Science Review*, 114(2):410–425.