Instrumental Variable (II)

Shengqiao Lin (shengqiaolin@utexas.edu)

2023-02-26

Goals for Today

- Assumptions
- Tutorial
- Replication #1: Lerman and McCabe JOP 2012 (RDD as IV)
- Replication #2: López-Cariboni CPS 2022 (panel data with IV)

Assumptions for IV Strategy

- Independence
- Exclusion
- First-stage
- Monotonicity

Independence

- Instrument Zi is independent of potential outcomes and potential treatments
  - Potential outcomes: Y_{i0},Y_{i1}
  - Potential treatments: D_{i0}, D_{i1}
- i.e. IV is as-if randomly assigned
- Violation: assignment based on potential outcomes/treatments
  - E.g., Private tutoring -> GPA -> Salary
  - Assignment of tutoring opportunities is not based on the potential improvement in GPA or salary.
- **Justification + suggestive evidence (T. Test)**

Exclusion

- Instrument only affects outcomes through the treatment
- Violation: Not backdoor paths
- Not really testable (infinite paths in reality)
  - E.g., Private tutoring -> GPA -> Salary
  - Tutored students make higher salary because they gain skills of communication
- Justification needed (why alternative paths do not exist/work)

First-Stage

- Instrument predicts the treatments
- Violations: weak instrument

  - E.g., Private tutoring -> GPA -> Salary
  - private tutoring does not improve GPA (significantly)
- Statistically testable; F value> 10 (Lee et al., F>109)
- Z -> D ✓ , what if D->Y insignificant?

Monotonicity

- No defiers
  - Z=1, D=0; Z=0, D=1
  - E.g., Private tutoring -> GPA -> Salary: tutoring reduces GPA
- It is usually not realistic and not testable
- Justification needed if being challenged

Tutorial: IV Regression

Acknowledgement: This example is adopted from Dr. Stephen Jessee's lecture on causal inference.

Simulated Data

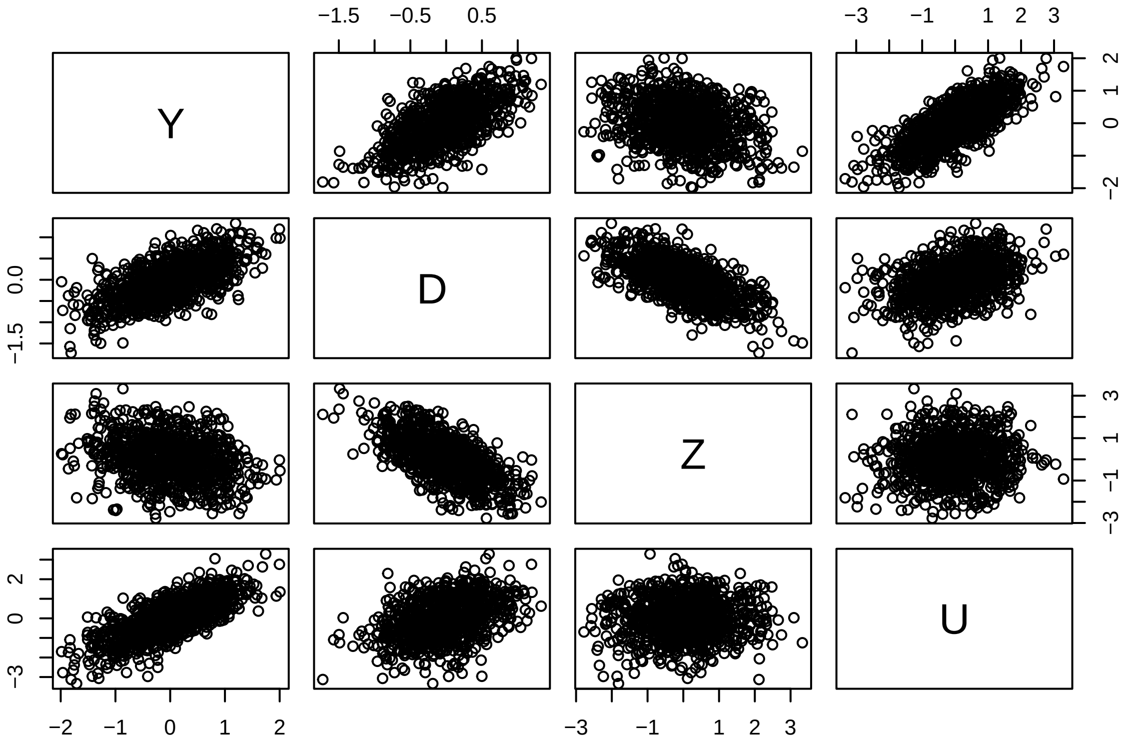
Let's start with generating some simulated data where Y is the outcome of interest, D is the treatment, Z is the instrumental variable, and U is a set of unobserved confounders (sometimes observable).

```r
set.seed(12345)
# setting sample size
N <- 1000

Z <- rnorm(N) # instrumental variable
U <- rnorm(N) # unobserved confounders

D <- .2*U - .3*Z + rnorm(N, sd=.3) # (confounded) treatment variable
Y <- .4*U + .6*D + rnorm(N, sd=.4) # dependent variable

pairs(cbind(Y, D, Z, U))
```



Let's do a naive estimation first. The estimated effect is nearly 1 (when real effect is .6).

```
reg.Y.D <- lm(Y ~ D)
summary(reg.Y.D)
```

```
##
## Call:
## lm(formula = Y ~ D)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.94296 -0.35151 -0.01812  0.33887  1.69304
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.004351   0.016809   0.259    0.796
## D           0.962743   0.036676  26.250   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5306 on 998 degrees of freedom
## Multiple R-squared:  0.4084, Adjusted R-squared:  0.4078
## F-statistic:   689 on 1 and 998 DF,  p-value: < 2.2e-16
```

Ideally, to get the real effect, we want to control U. But it is unrealistic because U is unobserved/unobservable.

```
reg.Y.DU <- lm(Y ~ D + U)
summary(reg.Y.DU)
```

```
##
## Call:
## lm(formula = Y ~ D + U)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.34994 -0.25046 -0.01267  0.25335  1.02902
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.006515   0.012083   0.539     0.59
## D           0.599115   0.028922  20.715   <2e-16 ***
## U           0.401353   0.013128  30.572   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3814 on 997 degrees of freedom
## Multiple R-squared:  0.6947, Adjusted R-squared:  0.6941
## F-statistic:  1134 on 2 and 997 DF,  p-value: < 2.2e-16
```

Constant Effects Models

Assuming we have an instrumental variable Z which affects Y through D and is independent from U. We start with estimating the effect of Z on Y.

```r
reg.Y.Z <- lm(Y ~ Z)
summary(reg.Y.Z)
```

```
##
## Call:
## lm(formula = Y ~ Z)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0067 -0.4642  0.0171  0.4674  2.0041
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.01427    0.02115  -0.675      0.5
## Z           -0.17199    0.02116  -8.126  1.3e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6681 on 998 degrees of freedom
## Multiple R-squared:  0.06206,    Adjusted R-squared:  0.06112
## F-statistic: 66.03 on 1 and 998 DF,  p-value: 1.302e-15
```

The effect of Z on D is -.3 and the effect of D on Y is .6. So the real effect of Z on Y is -.3*.6=-.18, which is very closed to the estimated effect.

We are interested in the real effect of D on Y.

```r
reg.Y.Z <- lm(Y ~ Z) #Reduced form
reg.D.Z <- lm(D ~ Z) #First stage
stargazer::stargazer(reg.D.Z,reg.Y.Z,type='text',no.space = TRUE)
```

```
##
## ==============================================================
##                                    Dependent variable:
```

```
##                                 ----------------------------
##                                      D              Y
##                                     (1)            (2)
## ------------------------------------------------------------
## Z                                 -0.296***      -0.172***
##                                    (0.011)        (0.021)
## Constant                           -0.014         -0.014
##                                    (0.011)        (0.021)
## ------------------------------------------------------------
## Observations                        1,000          1,000
## R2                                  0.416          0.062
## Adjusted R2                         0.416          0.061
## Residual Std. Error (df = 998)      0.350          0.668
## F Statistic (df = 1; 998)         711.528***     66.035***
## ============================================================
## Note:                            *p<0.1; **p<0.05; ***p<0.01
```

Then we can calculate the real effect of D on Y:

```
##      Z
## 0.5817
```

Two-stage Least Squares (2SLS)

2SLS will be more practical when you have more than 1 IV.

```
reg.D.Z <- lm(D ~ Z)
reg.Y.D_fitted<- lm(Y ~ reg.D.Z$fitted.values)
summary(reg.Y.D_fitted)
```

```
##
## Call:
## lm(formula = Y ~ reg.D.Z$fitted.values)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0067 -0.4642  0.0171  0.4674  2.0041
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -0.006164   0.021220  -0.290    0.772
## reg.D.Z$fitted.values  0.581700   0.071584   8.126  1.3e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6681 on 998 degrees of freedom
## Multiple R-squared:  0.06206,    Adjusted R-squared:  0.06112
## F-statistic: 66.03 on 1 and 998 DF,  p-value: 1.302e-15
```

IV Regression

We can also do it in one step:

```
##
## Call:
```

```
## ivreg(formula = Y ~ D | Z)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -1.950362 -0.387842 -0.000555  0.382135  1.532142
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.006164   0.017740  -0.347    0.728
## D            0.581700   0.059845   9.720   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5586 on 998 degrees of freedom
## Multiple R-Squared: 0.3445,  Adjusted R-squared: 0.3438
## Wald test: 94.48 on 1 and 998 DF,  p-value: < 2.2e-16
```

Control IV to test Exclusion Restriction? No

Will Y~D+Z help to test the exclusion restriction assumption? NO

```
reg.Y.DZ <- lm(Y ~ D + Z)
summary(reg.Y.DZ)
```

```
##
## Call:
## lm(formula = Y ~ D + Z)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.97997 -0.34316 -0.02367  0.34395  1.78548
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.002932   0.016157   0.181    0.856
## D           1.234409   0.046137  26.755   <2e-16 ***
## Z           0.192983   0.021144   9.127   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.51 on 997 degrees of freedom
## Multiple R-squared:  0.4541, Adjusted R-squared:  0.453
## F-statistic: 414.6 on 2 and 997 DF,  p-value: < 2.2e-16
```

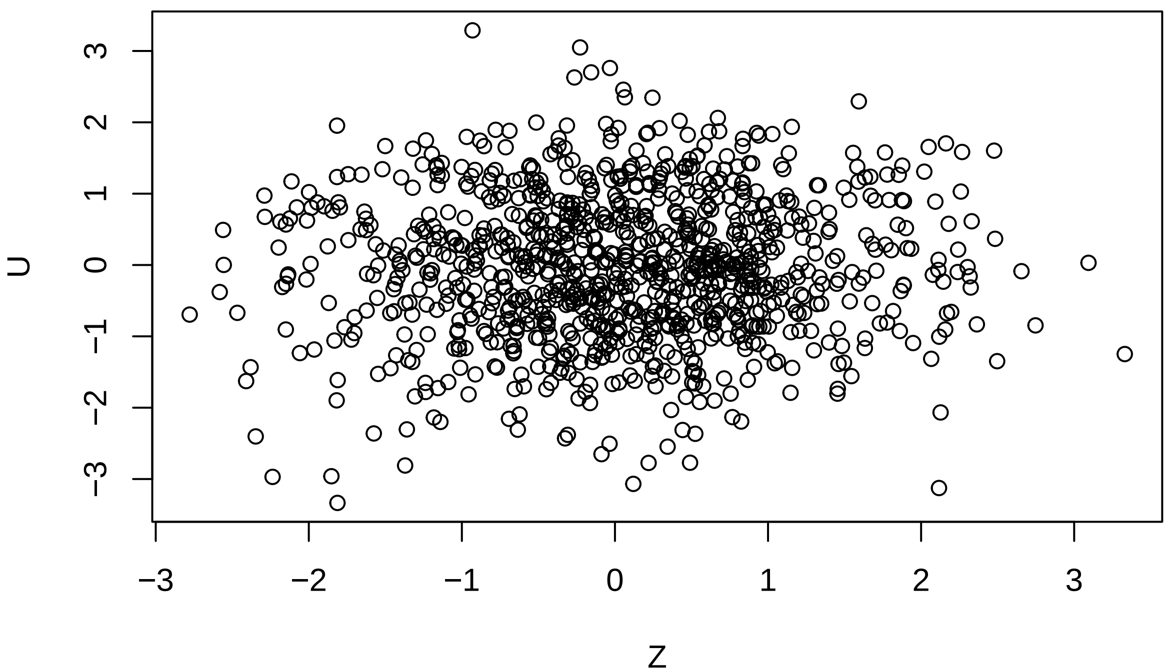
the coefficient on Z is significant even controlling for D! But we created these data such that Z only affects Y through D. Let's see what happened.

```
reg.U.Z <- lm(U ~ Z)
summary(reg.U.Z)
```

```
##
## Call:
## lm(formula = U ~ Z)
##
## Residuals:
```

```
##      Min       1Q  Median      3Q     Max
## -3.2315 -0.6975  0.0011  0.7156  3.3577
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.03220    0.03191  -1.009    0.313
## Z            0.03915    0.03193   1.226    0.221
##
## Residual standard error: 1.008 on 998 degrees of freedom
## Multiple R-squared:  0.001504,   Adjusted R-squared:  0.000503
## F-statistic: 1.503 on 1 and 998 DF,  p-value: 0.2205
```

```
plot(Z, U)
```

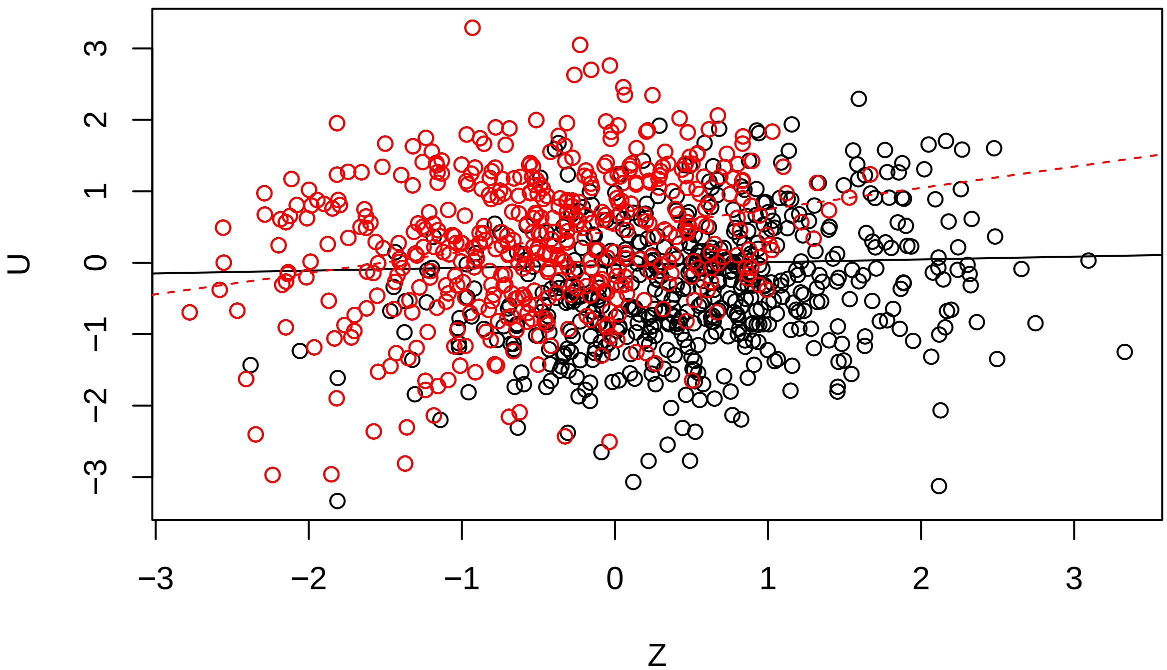



```
plot(Z, U)
abline(reg.U.Z)
points(Z[D>0], U[D>0], col="red")
reg.U.Z.posD <- lm(U[D>0] ~ Z[D>0])
summary(reg.U.Z.posD)
```

```
##
## Call:
## lm(formula = U[D > 0] ~ Z[D > 0])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.94792 -0.66253  0.03713  0.69522  3.11393
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.45232    0.05069   8.923  < 2e-16 ***
## Z[D > 0]     0.29789    0.05200   5.728  1.8e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.9554 on 476 degrees of freedom
## Multiple R-squared:  0.06449,    Adjusted R-squared:  0.06253
## F-statistic: 32.81 on 1 and 476 DF,  p-value: 1.8e-08
```

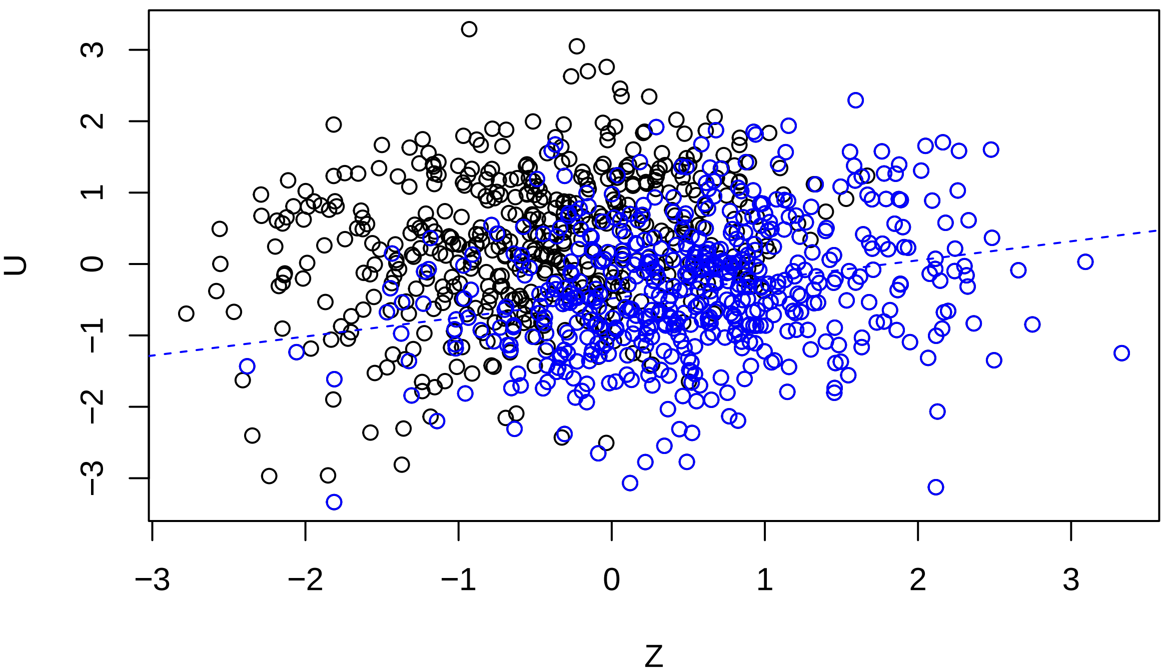
```
abline(reg.U.Z.posD, col="red", lty=2)
```




```
plot(Z, U)
points(Z[D<0], U[D<0], col="blue")
reg.U.Z.negD <- lm(U[D<0] ~ Z[D<0])
summary(reg.U.Z.negD)
```

```
##
## Call:
## lm(formula = U[D < 0] ~ Z[D < 0])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2089 -0.5604 -0.0299  0.5862  2.3508
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.48192    0.04640 -10.387  < 2e-16 ***
## Z[D < 0]     0.26672    0.04543   5.871 7.73e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8991 on 520 degrees of freedom
## Multiple R-squared:  0.06216,    Adjusted R-squared:  0.06036
## F-statistic: 34.47 on 1 and 520 DF,  p-value: 7.734e-09
```

```
abline(reg.U.Z.negD, col="blue", lty=2)
```



Conditioning on D (in this case high/low values) creates a non-causal association between Z and U it thus also creates a non-causal association b/t Z and Y, which operates through U. So this specification is just misguiding.

Replication

Lerman and McCabe, Personal Experience and Public Opinion: A Theory and Test of Conditional Policy Feedback, JOP 2017. Replication data

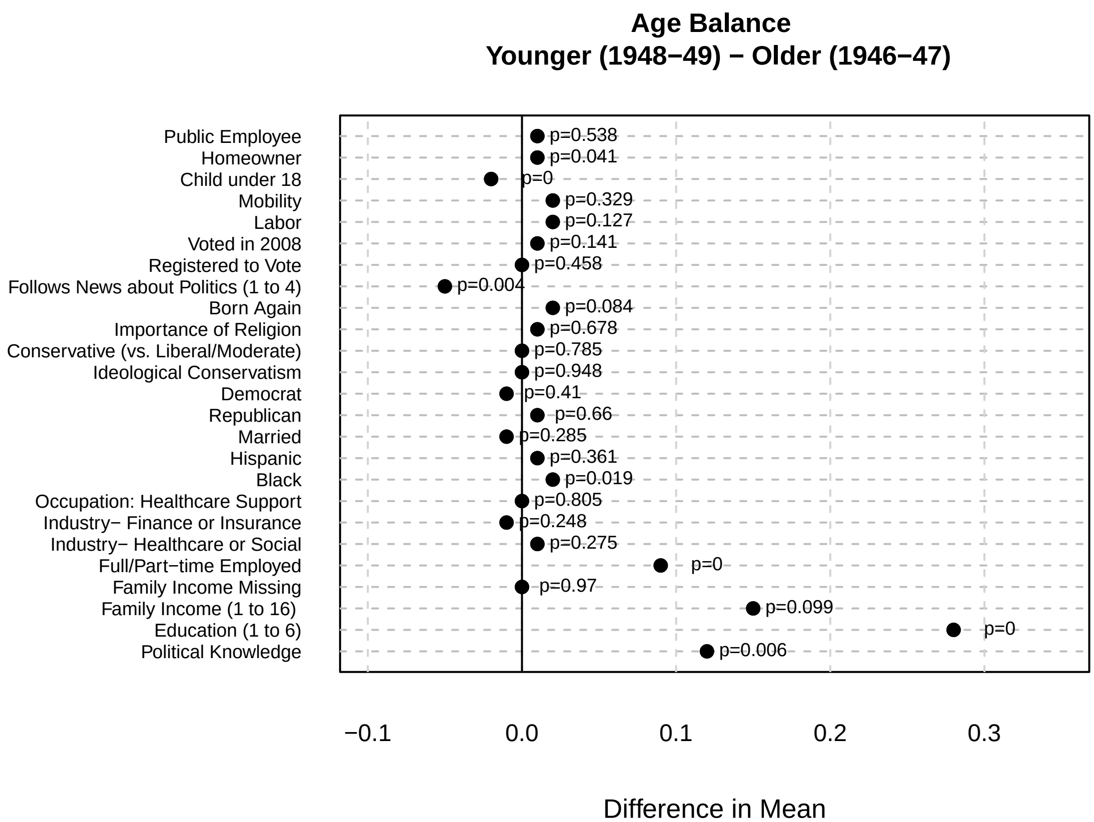
- Research Question: Personal Experience -> Public Opinion
- Empirical evidence: public health insurance -> support Medicare and Affordable Care Act
- Heterogeneity:
  - Political Knowledge (lower+)
  - Health (poor+)
  - Partisanship (Rep+)
- Design:
  - IV: Age>=65, eligible for ACA, i.e. born in or before 1947
  - RDD as IV, 63&64 v.s. 65&66

Independence

- Is IV independent from potential outcomes/treatments?
- Is IV assigned as-if random?
- When deciding the cutoff, did policy makers select 65 instead of 64 because:
  - Potential outcomes: 65 will support ACA significantly more than 64
  - Potential treatment: 65 will comply significantly more than 64
- [63,66] -> not a big concern, if not…

Exclusion

- Are there any other paths through which Z will affect Y?
- Will 65 support ACA more than 64 for other reasons that enrolled in public health insurance?
  - People change as they age -> [63,66] ✓
  - Other confounders change as they age ✓



- Picky/Unreasonable reviewers:
  - 65 is also a threshold of many other policies/life changes
  - SUTVA? What if I support because my friends are eligible?
  - Self-reported ages

First stage

- Does IV predict treatment (public health insurance)?

```
prop.table(table(byr4243,privpubins3r),1)
```

```
##         privpubins3r
## byr4243         0         1
##       0 0.1966449 0.8033551
##       1 0.1724931 0.8275069
```

```
prop.table(table(byr4445,privpubins3r),1)
```

```
##         privpubins3r
## byr4445         0         1
##       0 0.2726586 0.7273414
##       1 0.1966449 0.8033551
```

```r
prop.table(table(byr4647,privpubins3r),1)
```

```
##        privpubins3r
## byr4647         0         1
##       0 0.7548199 0.2451801
##       1 0.2726586 0.7273414
```

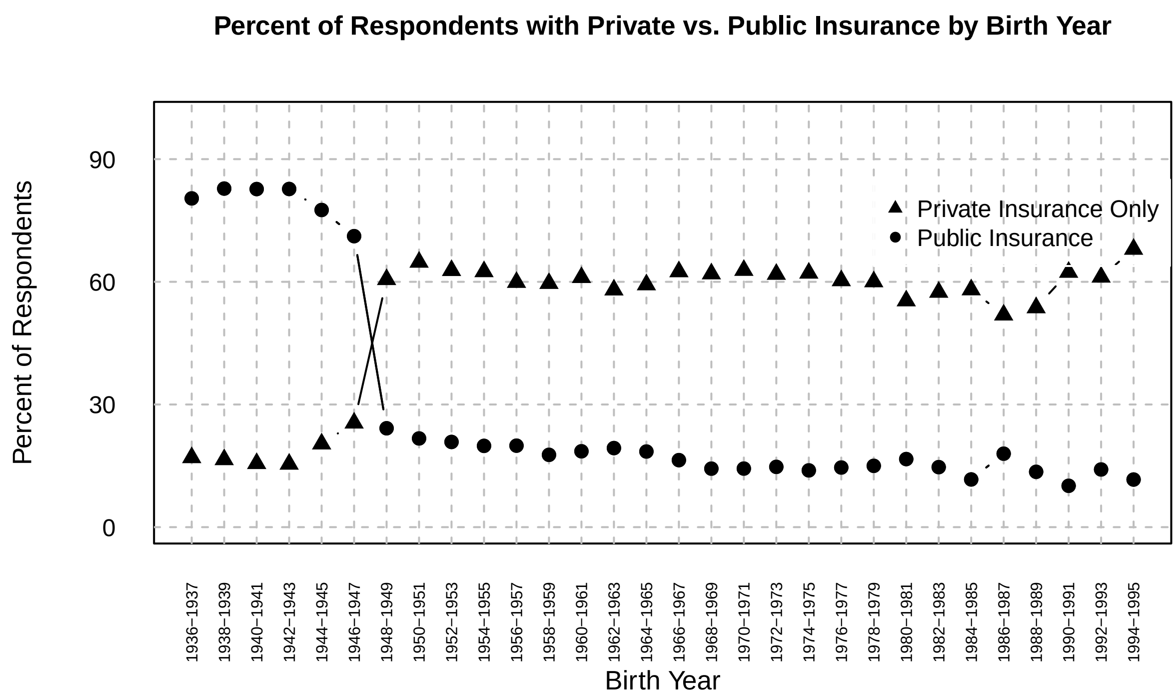
```r
prop.table(table(byr4849,privpubins3r),1)
```

```
##        privpubins3r
## byr4849         0         1
##       0 0.7708333 0.2291667
##       1 0.7548199 0.2451801
```

```r
prop.table(table(byr5051,privpubins3r),1)
```

```
##        privpubins3r
## byr5051         0         1
##       0 0.7769560 0.2230440
##       1 0.7708333 0.2291667
```

```
##
##     0     1
## 25787 28748
```



Monotonicity

- Will the eligibility discourage people to enroll public health insurance?
  - Not like but we don't have evidence
  - Picky reviewer: reduced income?

Results

- Outcomes: (1) support ACA, (2) support "do not cut medicare''
- Treatment: Public v.s. private/no insurance
- 2SLS

```r
m_iv_s1<- feols(privpubins3r~byr4647+rep+ind+con+mod+ideostrength+hcsocial+
                fininsur+healthcaresupport+child18+male+married+labor+mobility+
                homeowner+religimp+employed+votereg+vote08+black+hispanic2+military+
                educ+fincome+newsint+publicemp+bornagain,se='hc1',data=data)
etable(m_iv_s1,keep = 'byr4647')
```

```
##                             m_iv_s1
## Dependent Var.:        privpubins3r
##
## byr4647          0.4414*** (0.0127)
## ________________ __________________
## S.E. type        Heteroskedas.-rob.
## Observations                  4,436
## R2                          0.35529
## Adj. R2                     0.35134
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

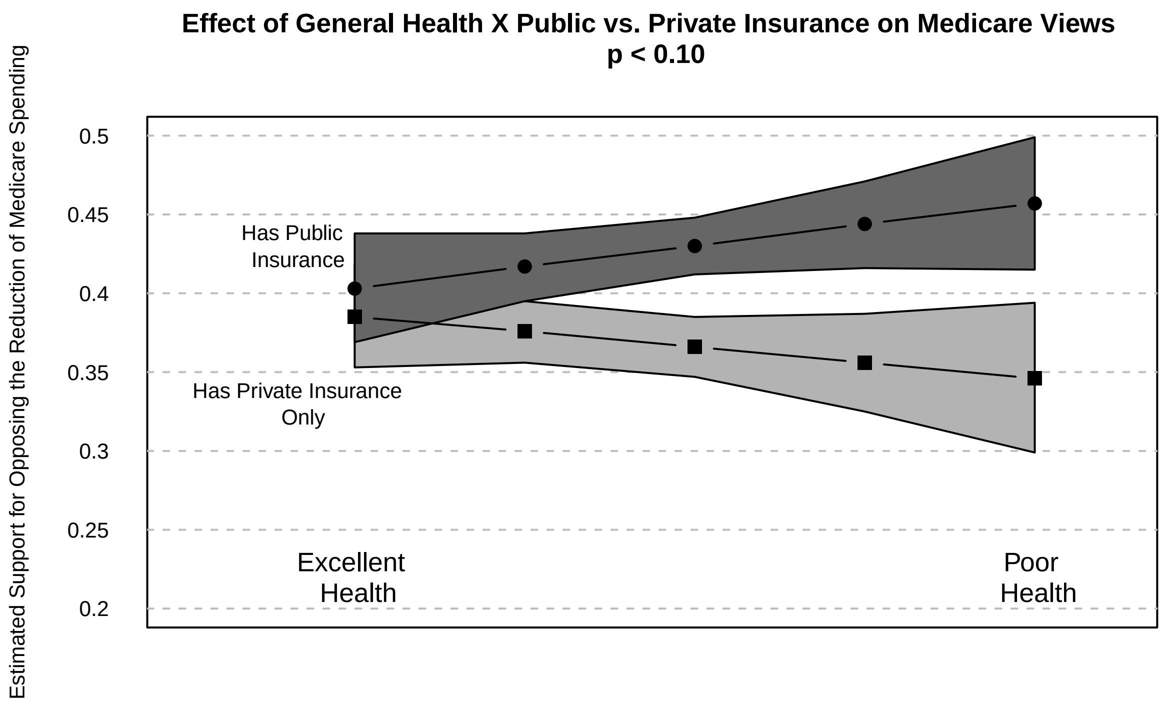
```r
m_iv1<- feols(suppafford~rep+ind+con+mod+ideostrength+hcsocial+
                fininsur+healthcaresupport+child18+male+married+labor+mobility+
                homeowner+religimp+employed+votereg+vote08+black+hispanic2+military+
                educ+fincome+newsint+publicemp+bornagain|privpubins3r~byr4647,se='hc1',data=data)
m_iv2<- feols(dontcutmedicare~rep+ind+con+mod+ideostrength+hcsocial+
                fininsur+healthcaresupport+child18+male+married+labor+mobility+
                homeowner+religimp+employed+votereg+vote08+black+hispanic2+military+
                educ+fincome+newsint+publicemp+bornagain|privpubins3r~byr4647,se='hc1',data=data)
etable(m_iv1,m_iv2,keep = 'privpubins3r')
```

```
##                            m_iv1             m_iv2
## Dependent Var.:       suppafford   dontcutmedicare
##
## privpubins3r    0.0459* (0.0229) 0.0836** (0.0270)
## ________________ ________________ __________________
## S.E. type       Heterosked.-rob. Heteroskeda.-rob.
## Observations               4,389             4,347
## R2                       0.56589           0.37139
## Adj. R2                  0.56320           0.36746
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

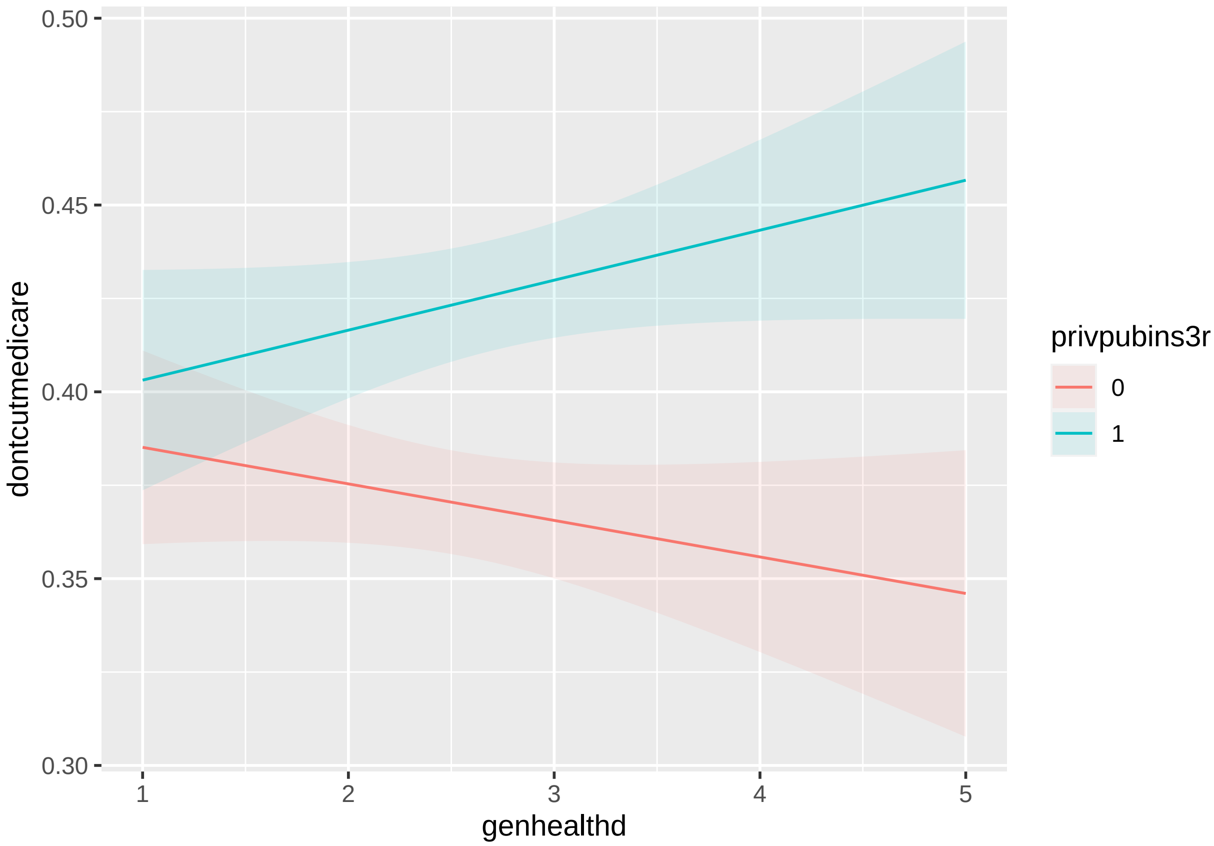
Heterogeneity

Health (1-6)

- Specification: add treatment × Health in the model
- Marginal effect (predicted mean)
- Finding: Effects are stronger for people in poor heath



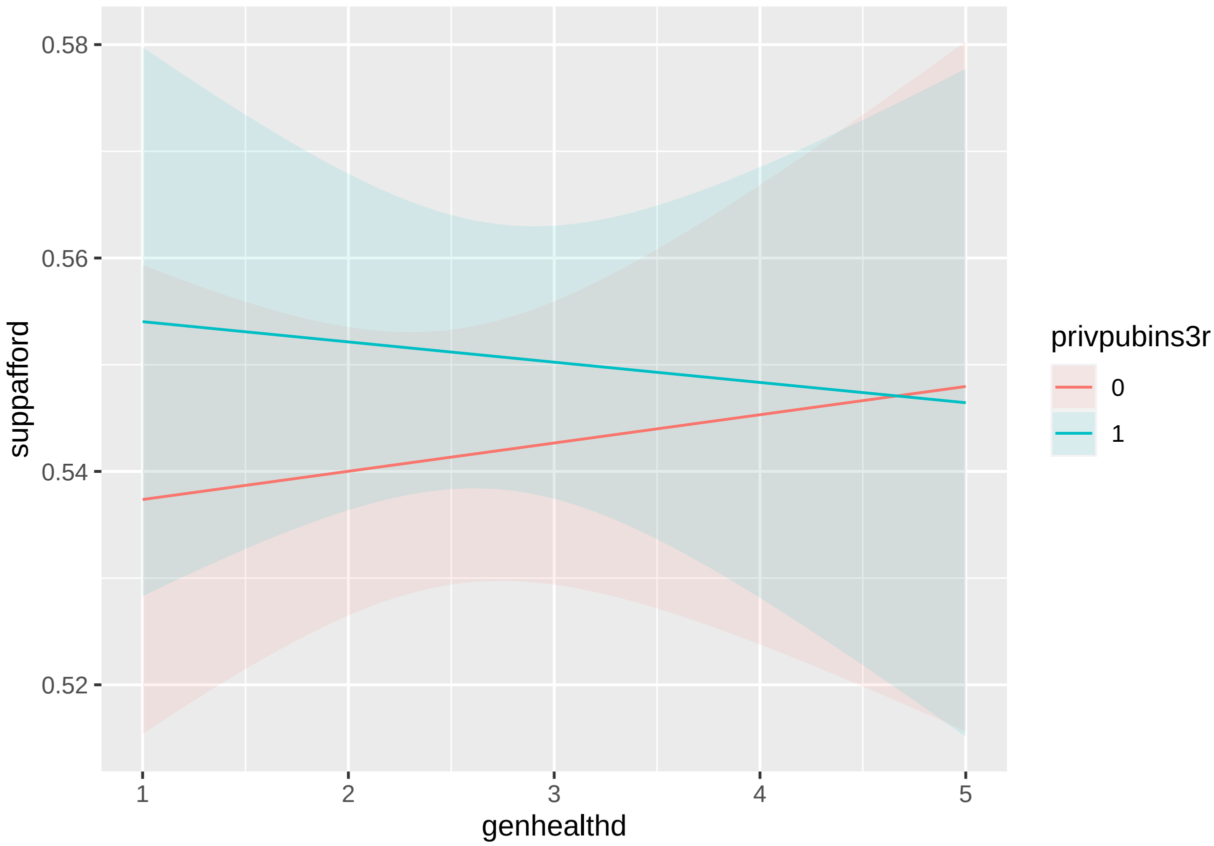

```r
m_health1<- feols(dontcutmedicare~privpubins3r*genhealthd+rep+ind+con+mod+ideostrength+hcsocial+
                fininsur+healthcaresupport+child18+male+married+labor+mobility+
                homeowner+religimp+employed+votereg+vote08+black+hispanic2+military+
                educ+fincome+newsint+publicemp+bornagain,se='hc1',
                data=subset(data,is.na(byr4647)==0&is.na(privpubins3r)==0))
plot_predictions(m_health1, condition = c('genhealthd','privpubins3r'),conf_level = 0.90)
```



- Question:
  - Why only one outcome (don't cut medicare) but the others (not support to ACA)?
  - Why it only present results without IV?

Extension

```
m_health2<- feols(suppafford~privpubins3r*genhealthd+rep+ind+con+mod+ideostrength+hcsocial+
                fininsur+healthcaresupport+child18+male+married+labor+mobility+
                homeowner+religimp+employed+votereg+vote08+black+hispanic2+military+
                educ+fincome+newsint+publicemp+bornagain,se='hc1',
                data=subset(data,is.na(byr4647)==0&is.na(privpubins3r)==0))
plot_predictions(m_health2, condition = c('genhealthd','privpubins3r'),conf_level = 0.90)
```

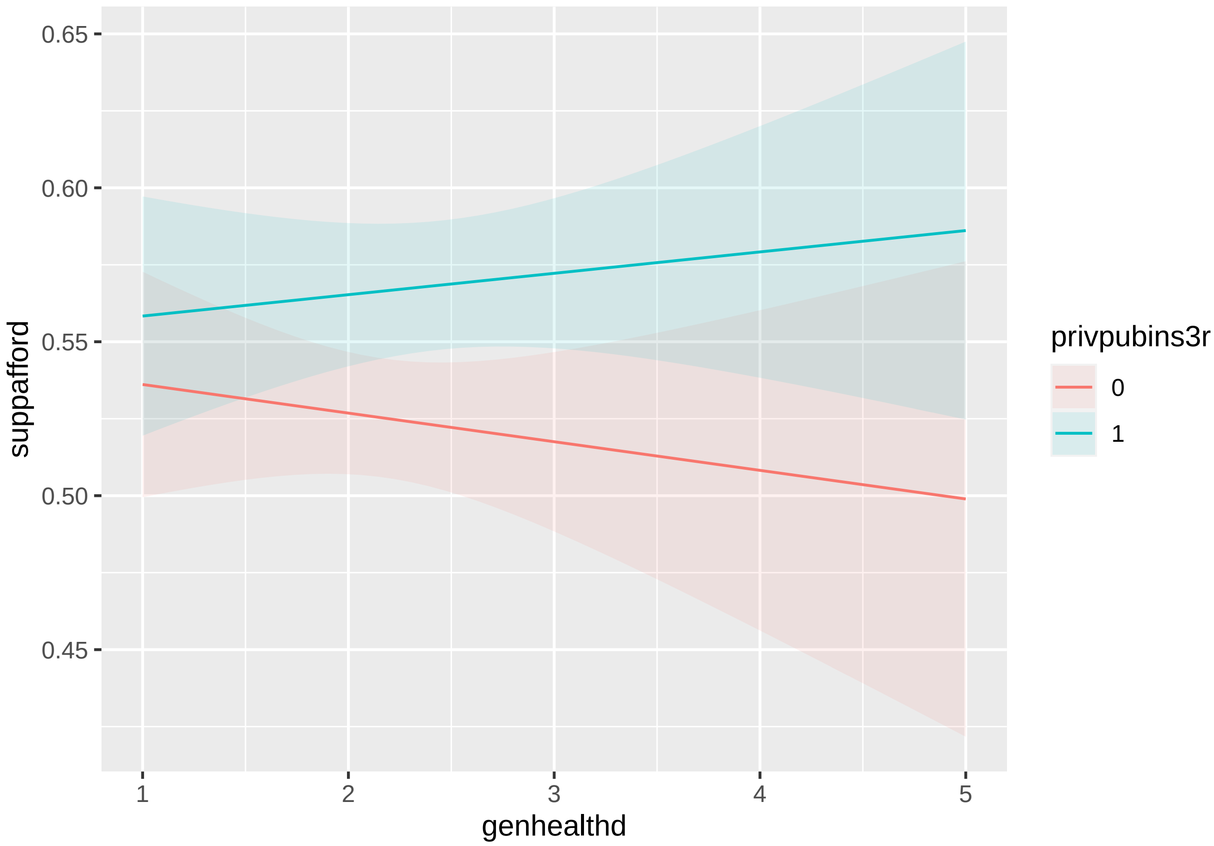



```
##                                    m_health2        m_health3            m_iv1
## Dependent Var.:                   suppafford       suppafford       suppafford
##
## privpubins3r                 0.0212 (0.0295) 0.0093 (0.0109) 0.0459* (0.0229)
## privpubins3r x genhealthd -0.0045 (0.0105)
## _________________________ ________________ _______________ ________________
## S.E. type                 Heterosked.-rob. Heteroske.-rob. Heterosked.-rob.
## Observations                         4,372           4,372            4,389
## R2                                 0.56595         0.56593          0.56589
## Adj. R2                            0.56305         0.56313          0.56320
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

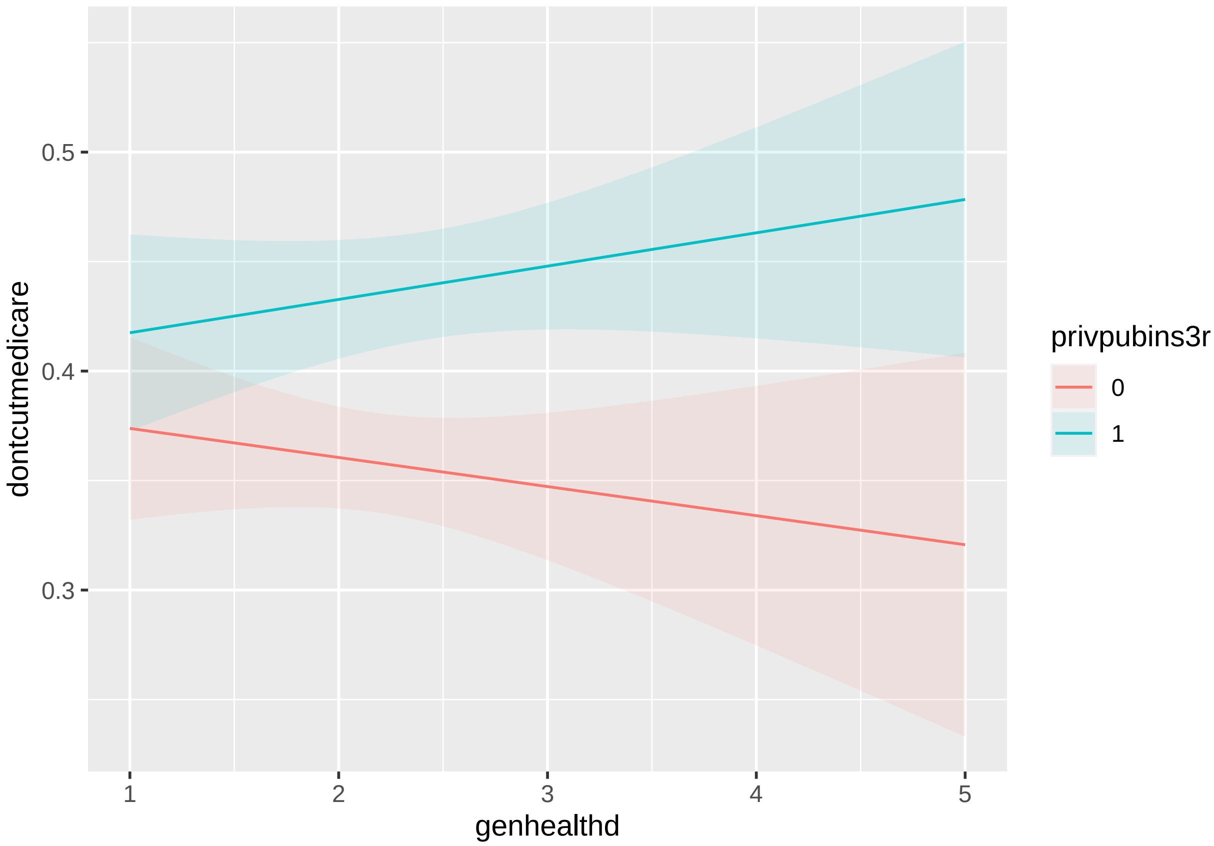
- Question: is this a problem?
  - Authors might think it a problem
  - But not really if we care only about CATE (Complier Average Treatment Effect)

Marginal Plot with IV

```
m_health4<- feols(suppafford~rep+ind+con+mod+ideostrength+hcsocial+
                    fininsur+healthcaresupport+child18+male+married+labor+mobility+
                    homeowner+religimp+employed+votereg+vote08+black+hispanic2+military+
                    educ+fincome+newsint+publicemp+bornagain|privpubins3r*genhealthd~byr4647*genhealthd
                  se='hc1',data=subset(data,is.na(byr4647)==0&is.na(privpubins3r)==0))
plot_predictions(m_health4, condition = c('genhealthd','privpubins3r'),conf_level = 0.90)
```




```
m_health5<- feols(dontcutmedicare~rep+ind+con+mod+ideostrength+hcsocial+
                    fininsur+healthcaresupport+child18+male+married+labor+mobility+
                    homeowner+religimp+employed+votereg+vote08+black+hispanic2+military+
                    educ+fincome+newsint+publicemp+bornagain|privpubins3r*genhealthd~byr4647*genhealthd
                  se='hc1',data=subset(data,is.na(byr4647)==0&is.na(privpubins3r)==0))
plot_predictions(m_health5, condition = c('genhealthd','privpubins3r'),conf_level = 0.90)
```



Further Examination

```
##                     m_health_iv1     m_health_iv2     m_health_iv3
## Dependent Var.: dontcutmedicare  dontcutmedicare  dontcutmedicare
##
## privpubins3r    0.0910 (0.0651) 0.0718. (0.0373) 0.0805. (0.0476)
## _______________ _______________ ________________ ________________
## S.E. type       Heteroske.-rob. Heterosked.-rob. Heterosked.-rob.
## Observations                476            1,529            1,516
## R2                      0.48813          0.43730          0.35137
## Adj. R2                 0.45728          0.42718          0.33960
##
##                     m_health_iv4    m_health_iv5
## Dependent Var.:  dontcutmedicare dontcutmedicare
##
## privpubins3r    0.1973. (0.1126) 0.2315 (0.3311)
## _______________ ________________ _______________
## S.E. type       Heterosked.-rob. Heteroske.-rob.
## Observations                 651             157
## R2                       0.29488         0.42924
## Adj. R2                  0.26432         0.30978
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                     m_health_iv1    m_health_iv2    m_health_iv3
## Dependent Var.:       suppafford      suppafford      suppafford
```

```
##
## privpubins3r    0.1031. (0.0576) 0.0303 (0.0308) 0.0108 (0.0404)
## _______________ ________________ _______________ _______________
## S.E. type       Heterosked.-rob. Heteroske.-rob. Heteroske.-rob.
## Observations                 479           1,542           1,537
## R2                       0.62632         0.60834         0.55416
## Adj. R2                  0.60395         0.60135         0.54619
##
##                    m_health_iv4    m_health_iv5
## Dependent Var.:      suppafford      suppafford
##
## privpubins3r    0.1059 (0.0939) 0.4228 (0.3194)
## _______________ _______________ _______________
## S.E. type       Heteroske.-rob. Heteroske.-rob.
## Observations                659             155
## R2                      0.54174         0.50909
## Adj. R2                 0.52213         0.40473
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

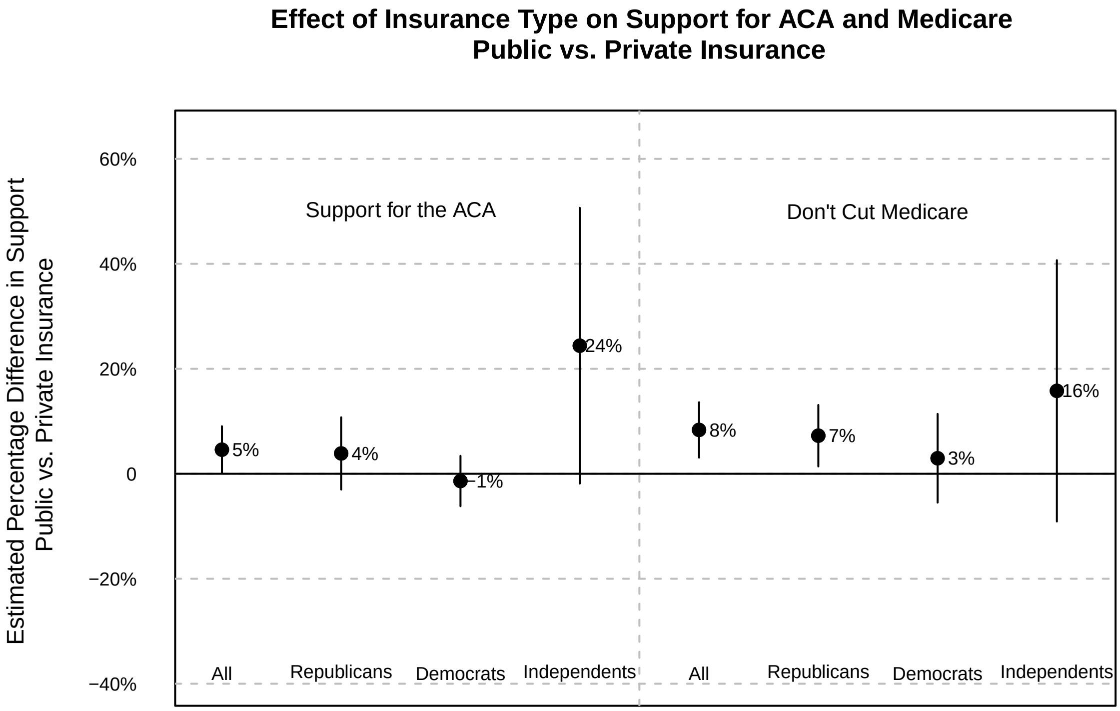
Other heterogeneity

- Partisanship

```
##                        m_pty_rep    m_pty_rep_iv        m_pty_dem
## Dependent Var.:       suppafford      suppafford       suppafford
##
## privpubins3r    -0.0007 (0.0176) 0.0388 (0.0354) -0.0024 (0.0118)
## _______________ ________________ _______________ ________________
## S.E. type       Heterosked.-rob. Heteroske.-rob. Heterosked.-rob.
## Observations               1,951           1,951            1,988
## R2                       0.14017         0.13783          0.10187
## Adj. R2                  0.12900         0.12664          0.09042
##
##                     m_pty_dem_iv       m_pty_ind     m_pty_ind_iv
## Dependent Var.:       suppafford      suppafford       suppafford
##
## privpubins3r    -0.0138 (0.0246) 0.0684 (0.0517) 0.2436. (0.1385)
## _______________ ________________ _______________ ________________
## S.E. type       Heterosked.-rob. Heteroske.-rob. Heterosked.-rob.
## Observations               1,988             450              450
## R2                       0.10146         0.17963          0.15780
## Adj. R2                  0.09001         0.13126          0.10815
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                         m_pty_rep     m_pty_rep_iv        m_pty_dem
## Dependent Var.:   dontcutmedicare  dontcutmedicare  dontcutmedicare
##
## privpubins3r    0.0411** (0.0150) 0.0726* (0.0301) 0.0498* (0.0210)
## _______________ _________________ ________________ ________________
## S.E. type       Heteroskeda.-rob. Heterosked.-rob. Heterosked.-rob.
## Observations                1,926            1,926            1,979
## R2                        0.11425          0.11216          0.13863
```

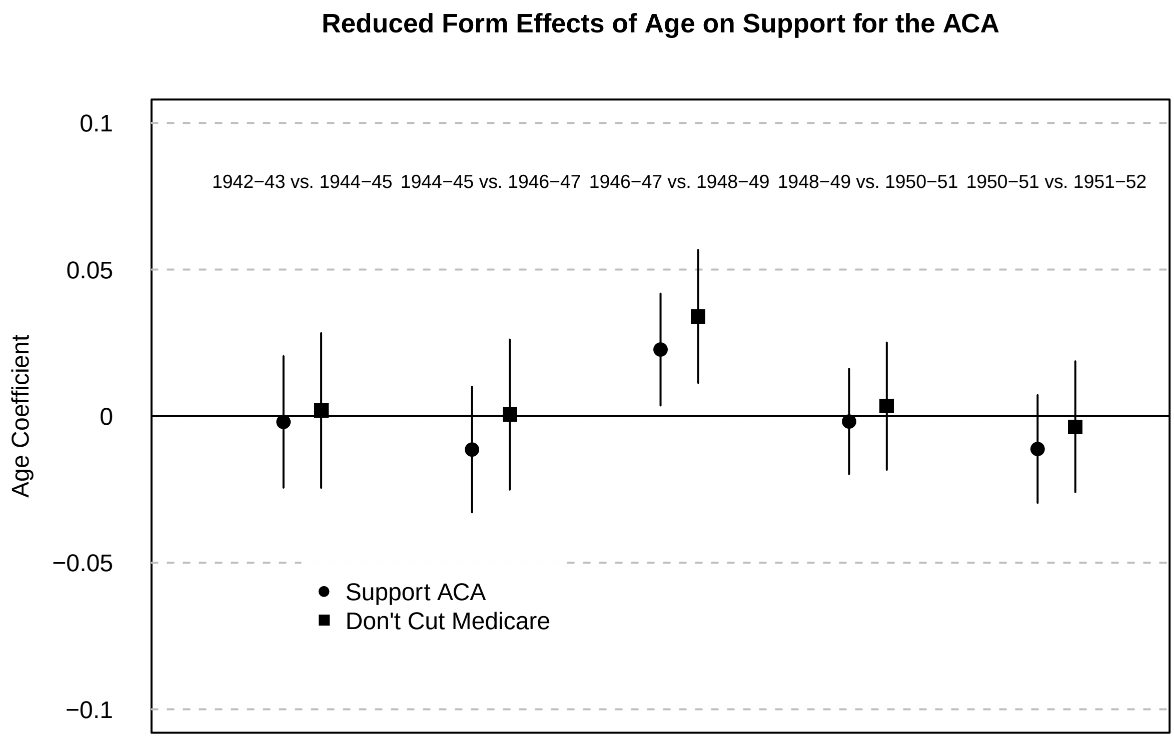
```
## Adj. R2                    0.10260          0.10048          0.12760
##
##                      m_pty_dem_iv         m_pty_ind    m_pty_ind_iv
## Dependent Var.: dontcutmedicare  dontcutmedicare dontcutmedicare
##
## privpubins3r    0.0296 (0.0434) 0.0886. (0.0478) 0.1582 (0.1304)
## _______________ _______________ ________________ _______________
## S.E. type       Heteroske.-rob. Heterosked.-rob. Heteroske.-rob.
## Observations              1,979              442             442
## R2                      0.13823          0.14253         0.13859
## Adj. R2                 0.12720          0.09100         0.08682
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



- Political knowledge

Robustness Check

- Placebo test: maybe there is also difference between other ages?



Takeaway

- Cross-sectional data + smart design
- Examine every assumption and offer justification
- Visualization

New package! IV with Bootstrap SE

ivDiag, Lal, Apoorva and Lockhart, Mackenzie William and Xu, Yiqing and Zu, Ziwen, How Much Should We Trust Instrumental Variable Estimates in Political Science? Practical Advice based on Over 60 Replicated Studies (August 14, 2021).

```
#devtool::install_github("apoorvalal/ivDiag")
library(ivDiag)
df <- data
D <- "privpubins3r"
Y1 <- "suppafford"
Y2 <- "dontcutmedicare"
Z <- "byr4647"
controls <-c("rep","ind","con","mod","ideostrength","hcsocial","fininsur",
                    "healthcaresupport","child18","male","married","labor",
                    "mobility","homeowner","religimp","employed","votereg",
                    "vote08","black","hispanic2","military","educ","fincome",
                   "newsint","publicemp","bornagain")
#cl<- NULL
#FE<- NULL
#weights<-NULL

m1<-ivDiag(data=df, Y=Y1, D=D, Z=Z,controls = controls)
```

```
## Bootstrapping:
## Parallelising 1000 reps on 7 cores
## Bootstrap took 42.455 sec.
##
## AR Test Inversion:
```

```
m1
```

```
## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0093 0.0109 0.8542 -0.0121   0.0307   0.393
## Boot.c   0.0093 0.0110 0.8444 -0.0121   0.0312   0.378
## Boot.t   0.0093 0.0109 0.8542 -0.0126   0.0312   0.388
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0459 0.0229 2.0095  0.0011   0.0908  0.0445
## Boot.c   0.0459 0.0221 2.0798  0.0024   0.0865  0.0400
## Boot.t   0.0459 0.0229 2.0095  0.0044   0.0875  0.0370
##
## $est_rf
##           Coef SE.t p.value   SE.b CI.b 2.5% CI.b 97.5% p.value.b
## byr4647 0.0202 0.01  0.0441 0.0097     0.001      0.038      0.04
##
## $est_fs
##           Coef   SE.t p.value   SE.b CI.b 2.5% CI.b 97.5% p.value.b
## byr4647 0.4401 0.0127       0 0.0128     0.416     0.4647         0
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    1272.162    1194.659          NA    1177.954    1194.659
##
## $tF.cF
##         F        cF      Coef        SE         t    CI2.5%   CI97.5%   p-value
## 1194.6594    1.9600    0.0459    0.0229    2.0095    0.0011    0.0908    0.0445
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    4.0595    1.0000 4387.0000    0.0440
##
## $AR$ci
## [1] "[0.0016, 0.0908]"
##
## $AR$bounded
## [1] TRUE
##
##
## $p_iv
## [1] 1
##
## $N
## [1] 4389
##
## $N_cl
```

```
## NULL
##
## $df
## [1] 4361
```

```r
m2<-ivDiag(data=df, Y=Y2, D=D, Z=Z,controls = controls)
```

```
## Bootstrapping:
## Parallelising 1000 reps on 7 cores
## Bootstrap took 40.718 sec.
##
## AR Test Inversion:
```

```r
m2
```

```
## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.055 0.0127 4.3131  0.0300   0.0799       0
## Boot.c   0.055 0.0130 4.2286  0.0294   0.0790       0
## Boot.t   0.055 0.0127 4.3131  0.0298   0.0801       0
##
## $est_2sls
##             Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0836 0.0270 3.1016  0.0308   0.1365  0.0019
## Boot.c   0.0836 0.0264 3.1624  0.0330   0.1360  0.0020
## Boot.t   0.0836 0.0270 3.1016  0.0318   0.1354  0.0020
##
## $est_rf
##            Coef  SE.t p.value   SE.b CI.b 2.5% CI.b 97.5% p.value.b
## byr4647 0.0371 0.012  0.0019 0.0118    0.0147     0.0601     0.002
##
## $est_fs
##            Coef   SE.t p.value   SE.b CI.b 2.5% CI.b 97.5% p.value.b
## byr4647 0.4435 0.0128       0 0.0124    0.4191     0.4678         0
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    1281.094    1202.212          NA    1276.511    1202.212
##
## $tF.cF
##         F        cF      Coef        SE         t    CI2.5%   CI97.5%   p-value
## 1202.2116    1.9600    0.0836    0.0270    3.1016    0.0308    0.1365    0.0019
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    9.6308    1.0000 4345.0000    0.0019
##
## $AR$ci
## [1] "[0.0313, 0.1365]"
##
## $AR$bounded
## [1] TRUE
##
##
## $p_iv
```

```
## [1] 1
##
## $N
## [1] 4347
##
## $N_cl
## NULL
##
## $df
## [1] 4319
```

Replication #2: López-Cariboni CPS 2022

López-Cariboni 2022 Political Regimes and Informal Social Insurance.

- Research Question: What is the political motivation for deliberate nonenforcement of the law?
- Argument: Democracies * Negative Economic Shock -> Electricity loss
- Data: 110 developing countries 1970-2014
- IV: regional democratic diffusion * Negative Economic Shock

```
library(readr)
dt<- read_csv("Section3 Instrument I/Section3/dt_replication.csv")
iv_dem <- ivreg(outgap.tdl ~ l.outgap.tdl
    + l.outgap.gdp.hamilton * l.democracy#Estimator
    + as.factor(iso3c)
    + as.factor(year)
    | .
    - l.outgap.gdp.hamilton*l.democracy#Estimator
    + l.outgap.gdp.hamilton*l.wreg.democracy#IV
    ,
    data=dt, na.action=na.omit)
summary(iv_dem)
```

```
##
## Call:
## ivreg(formula = outgap.tdl ~ l.outgap.tdl + l.outgap.gdp.hamilton *
##     l.democracy + as.factor(iso3c) + as.factor(year) | . - l.outgap.gdp.hamilton *
##     l.democracy + l.outgap.gdp.hamilton * l.wreg.democracy, data = dt,
##     na.action = na.omit)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -2.793263 -0.111132 -0.008912  0.084379  3.961454
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -0.0963814  0.0755814  -1.275  0.20233
## l.outgap.tdl                  0.6423725  0.0142227  45.165  < 2e-16 ***
## l.outgap.gdp.hamilton         0.0022054  0.0068398   0.322  0.74714
## l.democracy                   0.0361513  0.1020100   0.354  0.72307
## as.factor(iso3c)ALB           0.1134801  0.1086984   1.044  0.29657
## as.factor(iso3c)ARE           0.0256484  0.0774112   0.331  0.74042
## as.factor(iso3c)ARG           0.0063618  0.1169037   0.054  0.95660
## as.factor(iso3c)ARM          -0.0521678  0.0904658  -0.577  0.56421
```

```
## as.factor(iso3c)AZE                 0.0271074  0.0904714   0.300  0.76448
## as.factor(iso3c)BGD                 0.0228123  0.0975748   0.234  0.81516
## as.factor(iso3c)BGR                -0.0158363  0.1169311  -0.135  0.89228
## as.factor(iso3c)BHR                 0.0758740  0.0803829   0.944  0.34529
## as.factor(iso3c)BIH                -0.0649655  0.1390382  -0.467  0.64035
## as.factor(iso3c)BLR                -0.0006109  0.0904769  -0.007  0.99461
## as.factor(iso3c)BOL                -0.0226212  0.1213126  -0.186  0.85209
## as.factor(iso3c)BRA                 0.0100936  0.1127514   0.090  0.92867
## as.factor(iso3c)BRN                 0.0260661  0.0802893   0.325  0.74547
## as.factor(iso3c)BWA                -0.0284961  0.1321771  -0.216  0.82932
## as.factor(iso3c)CHL                 0.0250005  0.1030990   0.242  0.80842
## as.factor(iso3c)CHN                 0.0266289  0.0757279   0.352  0.72513
## as.factor(iso3c)CIV                 0.0334140  0.0758785   0.440  0.65971
## as.factor(iso3c)CMR                 0.0366175  0.0757152   0.484  0.62869
## as.factor(iso3c)COD                 0.0645921  0.0758840   0.851  0.39473
## as.factor(iso3c)COG                 0.0942302  0.0835898   1.127  0.25971
## as.factor(iso3c)COL                -0.0158241  0.1319608  -0.120  0.90456
## as.factor(iso3c)CRI                 0.0065985  0.1322744   0.050  0.96022
## as.factor(iso3c)CUB                 0.0337645  0.0757317   0.446  0.65574
## as.factor(iso3c)CYP                -0.0178146  0.1298952  -0.137  0.89092
## as.factor(iso3c)CZE                -0.0641254  0.1335383  -0.480  0.63112
## as.factor(iso3c)DOM                -0.0303150  0.1324636  -0.229  0.81900
## as.factor(iso3c)DZA                 0.0310692  0.0757261   0.410  0.68163
## as.factor(iso3c)ECU                -0.0010899  0.1189519  -0.009  0.99269
## as.factor(iso3c)EGY                 0.0270464  0.0757228   0.357  0.72098
## as.factor(iso3c)ERI                 0.0318902  0.0975178   0.327  0.74368
## as.factor(iso3c)EST                -0.0900163  0.1369359  -0.657  0.51100
## as.factor(iso3c)ETH                 0.0585880  0.0810231   0.723  0.46967
## as.factor(iso3c)GAB                 0.0305084  0.0757263   0.403  0.68707
## as.factor(iso3c)GEO                -0.0180608  0.1032328  -0.175  0.86113
## as.factor(iso3c)GHA                 0.0720550  0.0948404   0.760  0.44746
## as.factor(iso3c)GTM                 0.0010470  0.1231700   0.009  0.99322
## as.factor(iso3c)HND                 0.0043575  0.1174728   0.037  0.97041
## as.factor(iso3c)HRV                -0.0484238  0.1396635  -0.347  0.72883
## as.factor(iso3c)HTI                 0.0290805  0.0757351   0.384  0.70102
## as.factor(iso3c)HUN                -0.0397930  0.1346905  -0.295  0.76768
## as.factor(iso3c)IDN                 0.0246921  0.0887359   0.278  0.78083
## as.factor(iso3c)IND                 0.0083989  0.1299221   0.065  0.94846
## as.factor(iso3c)IRN                 0.0311597  0.0759282   0.410  0.68155
## as.factor(iso3c)IRQ                 0.0313712  0.0757278   0.414  0.67871
## as.factor(iso3c)ISR                -0.0190116  0.1321169  -0.144  0.88559
## as.factor(iso3c)JAM                -0.0094266  0.1323173  -0.071  0.94321
## as.factor(iso3c)JOR                 0.0034274  0.0779095   0.044  0.96491
## as.factor(iso3c)KAZ                -0.0093361  0.0904686  -0.103  0.91781
## as.factor(iso3c)KEN                 0.0266775  0.0848121   0.315  0.75313
## as.factor(iso3c)KGZ                 0.0192991  0.0904736   0.213  0.83110
## as.factor(iso3c)KHM                 0.0137153  0.0998556   0.137  0.89076
## as.factor(iso3c)KOR                -0.0209094  0.1067859  -0.196  0.84477
## as.factor(iso3c)KWT                 0.0330063  0.0999465   0.330  0.74124
## as.factor(iso3c)LBN                 0.0529039  0.0878224   0.602  0.54695
## as.factor(iso3c)LBY                 0.2307820  0.1131075   2.040  0.04140 *
## as.factor(iso3c)LKA                 0.0122397  0.0961313   0.127  0.89869
## as.factor(iso3c)LTU                -0.0167540  0.1396229  -0.120  0.90450
## as.factor(iso3c)LVA                -0.0748631  0.1400601  -0.535  0.59303
```

```
## as.factor(iso3c)MAR                 0.0369274  0.0757190   0.488  0.62580
## as.factor(iso3c)MDA                -0.0246903  0.1394300  -0.177  0.85946
## as.factor(iso3c)MEX                 0.0193335  0.0872039   0.222  0.82456
## as.factor(iso3c)MKD                -0.0112741  0.1334112  -0.085  0.93266
## as.factor(iso3c)MLT                -0.0008635  0.1321714  -0.007  0.99479
## as.factor(iso3c)MMR                 0.0303363  0.0757495   0.400  0.68883
## as.factor(iso3c)MNE                 0.0161723  0.1907842   0.085  0.93245
## as.factor(iso3c)MNG                -0.0047879  0.1305046  -0.037  0.97074
## as.factor(iso3c)MOZ                -0.0359686  0.0876690  -0.410  0.68163
## as.factor(iso3c)MUS                -0.0324916  0.1304515  -0.249  0.80332
## as.factor(iso3c)MYS                 0.0227709  0.0757211   0.301  0.76365
## as.factor(iso3c)NAM                 0.0367590  0.0920690   0.399  0.68973
## as.factor(iso3c)NER                 0.0360422  0.1382634   0.261  0.79436
## as.factor(iso3c)NGA                 0.0193410  0.0774048   0.250  0.80271
## as.factor(iso3c)NIC                -0.0240256  0.1162458  -0.207  0.83627
## as.factor(iso3c)NPL                 0.0146632  0.0913608   0.160  0.87250
## as.factor(iso3c)OMN                 0.0214063  0.0757140   0.283  0.77741
## as.factor(iso3c)PAK                 0.0008132  0.0932093   0.009  0.99304
## as.factor(iso3c)PAN                 0.0157893  0.1015391   0.155  0.87644
## as.factor(iso3c)PER                 0.0127395  0.1014688   0.126  0.90010
## as.factor(iso3c)PHL                 0.0019961  0.1112535   0.018  0.98569
## as.factor(iso3c)POL                -0.0589610  0.1334053  -0.442  0.65854
## as.factor(iso3c)PRY                 0.0014858  0.0834637   0.018  0.98580
## as.factor(iso3c)QAT                -0.0152580  0.1171755  -0.130  0.89641
## as.factor(iso3c)ROU                -0.0116447  0.1334438  -0.087  0.93047
## as.factor(iso3c)RUS                -0.0038277  0.0925213  -0.041  0.96700
## as.factor(iso3c)SAU                 0.0047119  0.0757561   0.062  0.95041
## as.factor(iso3c)SDN                 0.0225678  0.0770094   0.293  0.76950
## as.factor(iso3c)SEN                 0.0292358  0.0872501   0.335  0.73759
## as.factor(iso3c)SGP                 0.0142157  0.0757210   0.188  0.85109
## as.factor(iso3c)SLV                 0.0074315  0.1149239   0.065  0.94845
## as.factor(iso3c)SUR                 0.0575049  0.1515375   0.379  0.70436
## as.factor(iso3c)SVK                -0.0298885  0.1352356  -0.221  0.82510
## as.factor(iso3c)SVN                -0.0235229  0.1334988  -0.176  0.86015
## as.factor(iso3c)TGO                 0.0418883  0.0784486   0.534  0.59341
## as.factor(iso3c)THA                 0.0095160  0.1049851   0.091  0.92778
## as.factor(iso3c)TJK                 0.0349217  0.0905161   0.386  0.69967
## as.factor(iso3c)TKM                 0.0158482  0.0905251   0.175  0.86104
## as.factor(iso3c)TTO                 0.0351470  0.1314149   0.267  0.78914
## as.factor(iso3c)TUN                 0.0310974  0.0757267   0.411  0.68135
## as.factor(iso3c)TUR                 0.0055834  0.1250410   0.045  0.96439
## as.factor(iso3c)TZA                -0.0051167  0.0877885  -0.058  0.95353
## as.factor(iso3c)UKR                -0.0234375  0.1340214  -0.175  0.86119
## as.factor(iso3c)URY                 0.0064197  0.1126347   0.057  0.95455
## as.factor(iso3c)UZB                 0.0181445  0.0905778   0.200  0.84124
## as.factor(iso3c)VEN                -0.0169432  0.1137131  -0.149  0.88156
## as.factor(iso3c)VNM                -0.0034668  0.0835089  -0.042  0.96689
## as.factor(iso3c)YEM                 0.0101083  0.0908929   0.111  0.91146
## as.factor(iso3c)ZAF                 0.0163869  0.0962576   0.170  0.86483
## as.factor(iso3c)ZMB                 0.0272836  0.0791966   0.345  0.73049
## as.factor(iso3c)ZWE                 0.0385686  0.0757188   0.509  0.61053
## as.factor(year)1978                 0.0477726  0.0564564   0.846  0.39751
## as.factor(year)1979                 0.0384090  0.0565107   0.680  0.49676
## as.factor(year)1980                 0.0262766  0.0564452   0.466  0.64159
```

```
## as.factor(year)1981             0.1114896  0.0561780   1.985  0.04728 *
## as.factor(year)1982             0.0521906  0.0556874   0.937  0.34873
## as.factor(year)1983             0.1124997  0.0558479   2.014  0.04405 *
## as.factor(year)1984            -0.0073257  0.0561587  -0.130  0.89622
## as.factor(year)1985             0.0390559  0.0566671   0.689  0.49074
## as.factor(year)1986            -0.0021277  0.0561284  -0.038  0.96976
## as.factor(year)1987             0.0333909  0.0569295   0.587  0.55756
## as.factor(year)1988             0.1060890  0.0566294   1.873  0.06111 .
## as.factor(year)1989             0.1647638  0.0574957   2.866  0.00419 **
## as.factor(year)1990             0.0882197  0.0569067   1.550  0.12118
## as.factor(year)1991             0.1050157  0.0572741   1.834  0.06681 .
## as.factor(year)1992             0.0332995  0.0590216   0.564  0.57267
## as.factor(year)1993             0.0073011  0.0602568   0.121  0.90357
## as.factor(year)1994             0.1031109  0.0601549   1.714  0.08661 .
## as.factor(year)1995             0.0390873  0.0613603   0.637  0.52416
## as.factor(year)1996             0.0934199  0.0605137   1.544  0.12274
## as.factor(year)1997             0.0689713  0.0588758   1.171  0.24150
## as.factor(year)1998             0.0538141  0.0597448   0.901  0.36780
## as.factor(year)1999             0.0628991  0.0596550   1.054  0.29179
## as.factor(year)2000             0.1288616  0.0592191   2.176  0.02963 *
## as.factor(year)2001             0.1081492  0.0595244   1.817  0.06933 .
## as.factor(year)2002             0.0666299  0.0602498   1.106  0.26886
## as.factor(year)2003             0.0608043  0.0601372   1.011  0.31205
## as.factor(year)2004             0.0467735  0.0611736   0.765  0.44457
## as.factor(year)2005             0.0730868  0.0610259   1.198  0.23115
## as.factor(year)2006             0.0690148  0.0605076   1.141  0.25413
## as.factor(year)2007             0.1190551  0.0601148   1.980  0.04774 *
## as.factor(year)2008             0.1324141  0.0597912   2.215  0.02686 *
## as.factor(year)2009             0.0777981  0.0614147   1.267  0.20533
## as.factor(year)2010             0.0723541  0.0603809   1.198  0.23089
## as.factor(year)2011             0.0608862  0.0604628   1.007  0.31401
## as.factor(year)2012             0.0250867  0.0618298   0.406  0.68496
## as.factor(year)2013             0.0596776  0.0617932   0.966  0.33424
## as.factor(year)2014             0.0927240  0.0617555   1.501  0.13334
## as.factor(year)2015            -0.0312008  0.3234857  -0.096  0.92317
## as.factor(year)2016            -0.0218120  0.3235011  -0.067  0.94625
## l.outgap.gdp.hamilton:l.democracy -0.0607667  0.0256632  -2.368  0.01795 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3057 on 3096 degrees of freedom
## Multiple R-Squared: 0.4455,  Adjusted R-squared: 0.4183
## Wald test: 16.41 on 152 and 3096 DF,  p-value: < 2.2e-16
```

Another (easier) way to run a 2WFE with IV

```
library(fixest)#Fastest Fixed Effects
iv_dem<- feols(outgap.tdl ~ l.outgap.tdl|iso3c+year|l.outgap.gdp.hamilton * l.democracy~ l.outgap.gdp.ha
etable(iv_dem)
```

```
##                                            iv_dem
## Dependent Var.:                        outgap.tdl
##
## l.outgap.gdp.hamilton               0.0022 (0.0068)
## l.democracy                         0.0362 (0.1020)
```

```
## l.outgap.gdp.hamilton:l.democracy  -0.0608* (0.0257)
## l.outgap.tdl                      0.6424*** (0.0142)
## Fixed-Effects:                    ------------------
## iso3c                                            Yes
## year                                             Yes
## _________________________________ __________________
## S.E. type                                        IID
## Observations                                   3,249
## R2                                           0.44551
## Within R2                                    0.42776
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```