Microeconometrics (Causal Inference)
Weeks 7 and 8 - Instrumental variables

Joshua D. Merfeld
KDI School of Public Policy and Management

2023-10-25

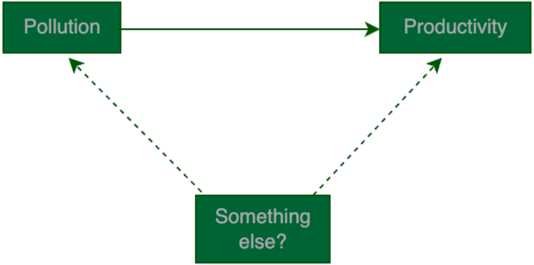
What are we doing today?

- Introduction to IVs
  - Requirements/assumptions
- IVs and RCTs
- In a world of LATE
- Weak instruments

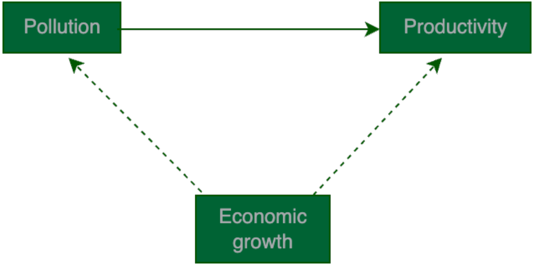
- Instrumental variables (IVs) are a way to estimate causal effects when we have endogeneity
  - The endogeneity can take many forms: omitted variables, measurement error, simultaneity, etc.

- Consider my paper: effects of pollution on agricultural productivity
  - What's the problem with simply regression productivity on pollution?

Endogeneity in the pollution example



Endogeneity in the pollution example



- What we really want to estimate is this:

$$productivity_{it} = \beta_0 + \beta_1 pollution_{it} + \epsilon_{it} \tag{1}$$

where β_1 is the causal effect of pollution on productivity.

- Endogeneity is defined as $cov(pollution_{it}, \epsilon_{it}) \neq 0$
  - That is, the error term is correlated with the endogenous variable
  - A common example is omitted variables

Putting structure on this

$$productivity_{it} = \beta_0^* + \beta_1^* pollution_{it} + \epsilon_{it}^* \tag{1}$$

- When we estimate this, due to the way OLS works, the residuals and pollution will be orthogonal
  - That is, $cov(pollution_{it}, \epsilon_{it}^*) = 0$
  - This is a property of OLS

- However, the issue is that under endogeneity, $\beta_1^* \neq \beta_1$
  - That is, the OLS estimate of β_1 is biased *for the true structural parameter*

- Another way to think about it is that what we want to estimate is this:

$$productivity_{it} = \beta_0 + \beta_1 pollution_{it} + \beta_2 X + \epsilon_{it} \tag{2}$$

- But if we don't properly control for everything – in this case X – we are really estimating this:

$$productivity_{it} = \tilde{\beta_0} + \tilde{\beta_1} pollution_{it} + \eta_{it}, \tag{3}$$

where $\eta_{it} = \beta_2 X_{it} + \epsilon_{it}$.

Differences in differences?

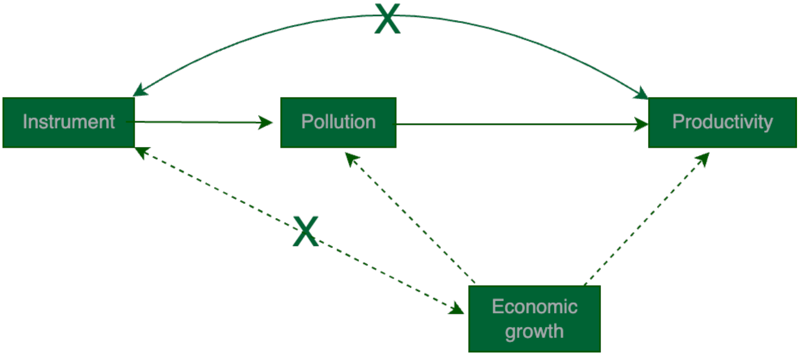
- One solution is to use a differences-in-differences (DiD) approach
- This requires the assumption of parallel trends
  - That is, the trends in the outcome variable would have been the same in the absence of the treatment
- But what if changing economic growth is leading to changes in both pollution and productivity?
  - Then the parallel trends assumption is violated since areas with more pollution are also experiencing faster economic growth

- If you're willing to make assumptions about what the omitted variables are, maybe you could control for theme
- But this is a strong assumption
  - No matter what we do, we'll have to make assumptions, though

Enter: instruments

- Let's take a different approach
- We'll use an instrument
  - A variable that is correlated with the endogenous variable (pollution) but is not correlated with the error term

Instrument in the pollution example



- I very purposefully created the example so that the instrument is correlated with pollution
  - But it's not *directly* correlated with productivity
  - And it's not correlated with the omitted variable (the error term... will show you this in a second)

- Let's look at these more formally

Back to our problem

$$productivity_{it} = \tilde{\beta}_0 + \tilde{\beta}_1 pollution_{it} + \eta_{it} \quad (3)$$

- Can we estimate a version of this equation – that is, without controlling for X_{it} – and still get causal effects?
- Maybe, if we can find a valid instrument.
- So what makes an instrument valid?

What else can instruments help with?

- It turns out IVs can also help with measurement error
  - If we have a variable that is measured with error, we can use an instrument to correct for this
- From Hansen, consider the model:

$$X = Q + u, \tag{4}$$

where X is the variable we observe, Q is the variable we want to measure, and u is measurement error.

- Assume that $cov(u, Q) = 0$, so that the measurement error is *random*, i.e. uncorrelated with the true value of Q.
  - This is known as classical measurement error

- We want to estimate:

$$Y = \beta_0 + \beta_1 Q + \epsilon, \tag{5}$$

but what we really estimate is:

$$Y = \tilde{\beta}_0 + \tilde{\beta}_1 X + \tilde{\epsilon} = \tilde{\beta}_0 + \tilde{\beta}_1 (Q + u) + \tilde{\epsilon} \tag{6}$$

- This is what we get:

$$\tilde{\beta}_1 = \beta_1 \left(1 - \frac{\mathbb{E}(u^2)}{\mathbb{E}(X^2)}\right) \tag{7}$$

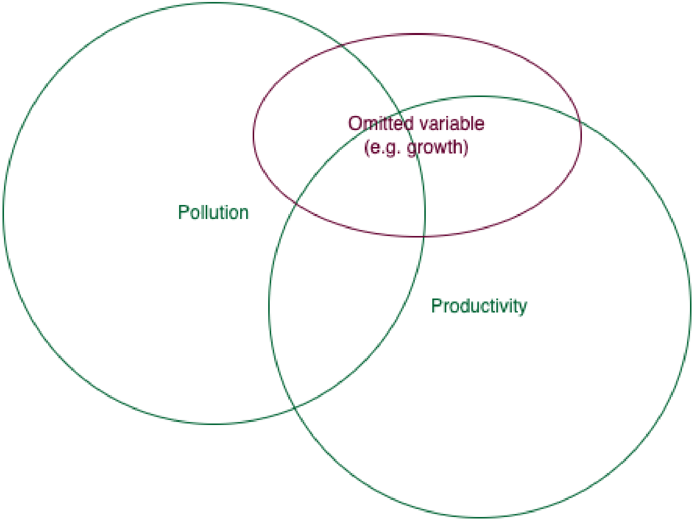
- By definition, $\mathbb{E}(X^2) > \mathbb{E}(u^2)$, so $\tilde{\beta}_1 < \beta_1$.
  - Why is this true?
  - That is, the OLS estimate of β_1 is biased *towards zero*
  - This is called attenutation bias, but is only guaranteed with the measurement error is classical (random)

$$productivity_{it} = \tilde{\beta_0} + \tilde{\beta_1} pollution_{it} + \eta_{it} \tag{3}$$

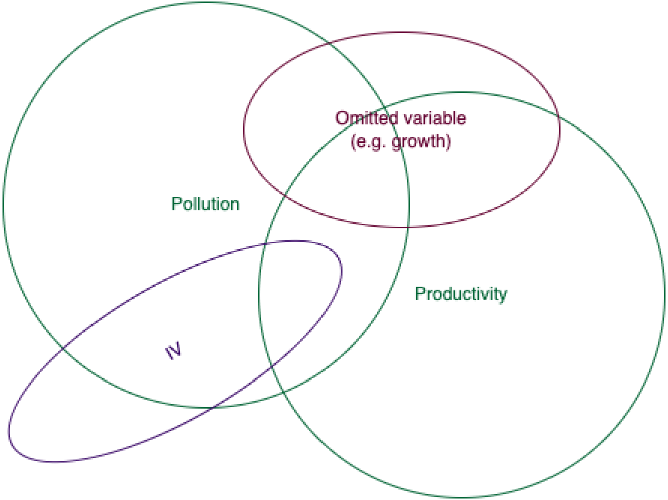
1. The instrument must be correlated with the endogenous variable (pollution)
2. The instrument must not be correlated with the error term (η_{it})
   - Note that this implies two things:
     - The instrument must not be correlated with any omitted variable (here X_{it})
     - The instrument must not directly affect the outcome ($productivity_{it}$)

- If we can find a valid instrument, we can use it to estimate the causal effect of pollution on productivity
- The simplest example uses two stages:
  1. $pollution_{it} = \pi_0 + \pi_1 instrument_{it} + \nu_{it}$
  2. $productivity_{it} = \phi_0 + \phi_1 pollution_{it} + \zeta_{it}$
- We can then estimate ϕ_1 using OLS
  - Note that only under certain circumstances will $\phi_1 = \beta_1$
  - More on this later

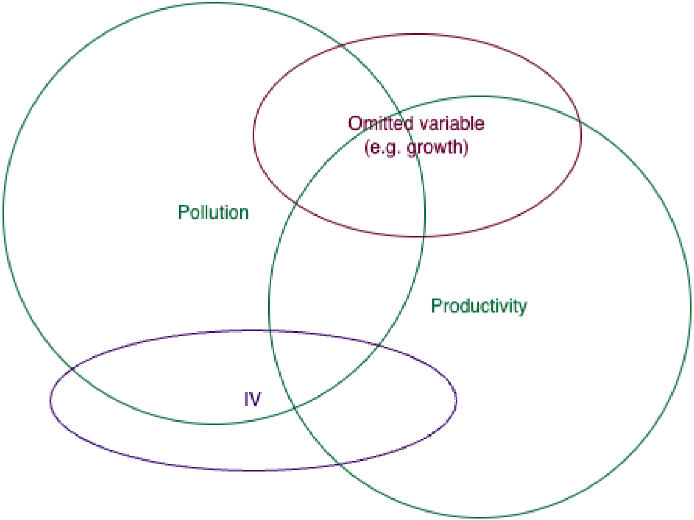
The intuition with venn diagrams



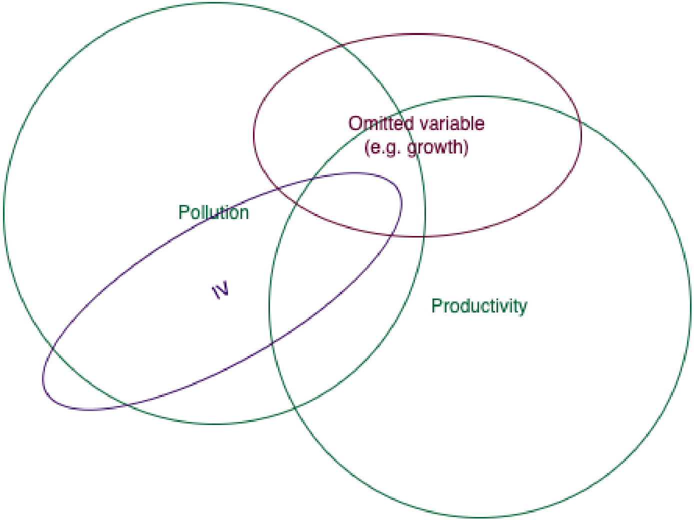
The IV only affects productivity through pollution



This doesn't work. Direct effects on productivity!



This doesn't work. Correlated with growth!



$$\text{Stage 1}: \quad T_{it} = \pi_0 + \pi_1 Z_{it} + \nu_{it}$$
$$\text{Stage 2}: \quad Y_{it} = \phi_0 + \phi_1 T_{it} + \zeta_{it}$$

- Requirements:
    - $cov(Z_{it}, T_{it}) \neq 0$
    - $cov(Z_{it}, \zeta_{it}) = 0$

- We first regress T on the instrument to get $\hat{T}_{it}$
- Then, we use the predicted values of T to estimate the effects on Y
    - If the IV is valid, these predicted values *are unrelated to the omitted variables!*

Some comments

$$\text{Stage 1}: \quad T_{it} = \pi_0 + \pi_1 Z_{it} + \nu_{it}$$

$$cov(Z_{it}, T_{it}) \neq 0 \tag{8}$$

- This is the first requirement
- We can test this!
  - F-test of all *excluded instruments* in the first stages
  - I say all excluded instruments because you can technically have more than one

Some comments

$$
\begin{aligned}
\text{Stage 1}: \quad T_{it} &= \pi_0 + \pi_1 Z_{it} + \nu_{it} \\
\text{Stage 2}: \quad Y_{it} &= \phi_0 + \phi_1 T_{it} + \zeta_{it}
\end{aligned}
$$

$$
cov(Z_{it}, \zeta_{it}) = 0 \tag{9}
$$

- This is the second requirement
- We cannot explicitly test this
  - This is an identifying *assumption*
  - We need this to be true to attribute causality to the second stage

$$\text{Stage 1}: \quad T_{it} = \pi_0 + \pi_1 Z_{it} + \nu_{it}$$
$$\text{Stage 2}: \quad Y_{it} = \phi_0 + \phi_1 T_{it} + \zeta_{it}$$

$$cov(Z_{it}, \zeta_{it}) = 0 \tag{10}$$

- Note that we will use Z_{it} to predict T_{it}.
  - We cannot actually observe $cov(Z_{it}, \zeta_{it})$
- So if $cov(Z_{it}, \zeta_{it}) \neq 0 \ldots$
  - Then this correlation will be contained in the predicted values, $\hat{T}_{it}$
  - i.e. the predicted values will still be endogenous

- Economists have long been interested in supply and demand
  - Obviously. . .
- How does a change in supply affect prices?
  - Not a straightforward question to answer, because prices are determined jointly by supply and demand
  - We can't determine what is changing when we observe market prices
  - One option: an instrument that moves only one side of the market
- Small note: this is how IVs originally came about in economics

Favara and Imbs, 2015 (*American Economic Review*)

- How does the availability of credit affect house prices?

- They use a change in deregulation of banks in the US
  - This deregulation led to an increase in credit supply
  - But it did not affect credit demand, since it was a supply-side change

- Idea: show the change in credit availability for banks affected by the change
  - And no change for banks not affected by the change

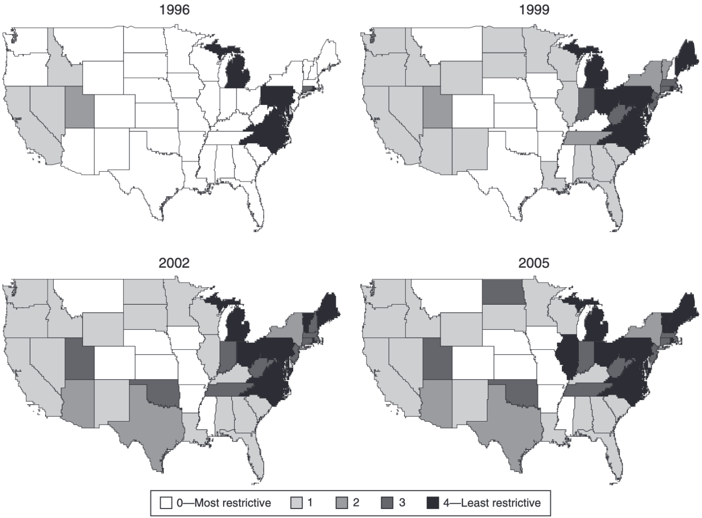


Figure B1. Rice-Strahan (2010) Deregulation Index by State and Year

Two stages: predict credit supply, then predict house prices

$$\text{Stage 1: } credit_{ct} = \delta_0 + \delta_1 deregulation_{ct} + \delta_2 X_{ct} + \alpha_c + \gamma_t + \nu_{ct} \tag{11}$$

$$\text{Stage 2: } price_{ct} = \beta_0 + \beta_1 credit_{ct} + \beta_2 X_{ct} + \phi_c + \eta_t + \zeta_{ct} \tag{12}$$

- They instrument for *credit* using *deregulation*
  - *deregulation* is correlated with *credit* but not with ζ_{ct}, according to the authors
  - (Let's ignore whether this is true for now since it's so contextual)

- They control for X_{ct}, which is a vector of controls
- This is also a two-way fixed effects specification:
  - α_c and γ_t (ϕ_c and η_t in stage 2) are county and year fixed effects

Replication data: hmda_merged.dta

```r
library(haven)
df <- read_dta("hmda_merged.dta")
head(df)
```

```
## # A tibble: 6 x 99
##    year county state_n yryear_1994 yryear_1995 yryear_1996 yryear_1997
##   <dbl>  <dbl>   <dbl>       <dbl>       <dbl>       <dbl>       <dbl>
## 1  1994   1001       1           1           0           0           0
## 2  1995   1001       1           0           1           0           0
## 3  1996   1001       1           0           0           1           0
## 4  1997   1001       1           0           0           0           1
## 5  1998   1001       1           0           0           0           0
## 6  1999   1001       1           0           0           0           0
## # i 92 more variables: yryear_1998 <dbl>, yryear_1999 <dbl>, yryear_2000 <dbl>,
## #   yryear_2001 <dbl>, yryear_2002 <dbl>, yryear_2003 <dbl>, yryear_2004 <dbl>,
## #   yryear_2005 <dbl>, Dl_nloans_b <dbl>, LDl_nloans_b <dbl>,
## #   Dl_vloans_b <dbl>, LDl_vloans_b <dbl>, Dl_nden_b <dbl>, LDl_nden_b <dbl>,
## #   Dl_lir_b <dbl>, LDl_lir_b <dbl>, Dl_nsold_b <dbl>, LDl_nsold_b <dbl>,
## #   Dl_nloans_pl <dbl>, LDl_nloans_pl <dbl>, Dl_vloans_pl <dbl>,
## #   LDl_vloans_pl <dbl>, Dl_nden_pl <dbl>, LDl_nden_pl <dbl>, ...
```

```r
# key controls: LDl_hpi Dl_inc LDl_inc Dl_pop LDl_pop Dl_her_v LDl_her_v
# instrument: Linter_bra
# endogenous variables: Dl_nloans_b Dl_vloans_b Dl_lir_b
# weights: w1
# restriction: border counties only (border==1)
# county and year FE
# cluster on state
```

- It is common to estimate the reduced form of the first stage
  - This is a regression of the outcome of interest on the instrument

- In this case, this equals

$$price_{ct} = B_0 + B_1 deregulation_{ct} + B_2 X_{ct} + \cdots \tag{13}$$

```r
bordercounties <- df %>% filter(border==1)
summary(feols(Dl_hpi ~ Linter_bra + LDl_hpi + Dl_inc + LDl_inc + Dl_pop + LDl_pop + Dl_her_v + LDl_her_v | county + year,
        data = bordercounties, weights = bordercounties$w1,
        cluster = "state_n"))
```

```
## OLS estimation, Dep. Var.: Dl_hpi
## Observations: 2,937
## Weights: bordercounties$w1
## Fixed-effects: county: 267,  year: 11
## Standard-errors: Clustered (state_n)
##             Estimate Std. Error   t value   Pr(>|t|)
## Linter_bra  0.004217   0.001822  2.314494 2.6813e-02 *
## LDl_hpi     0.530888   0.041265 12.865486 1.2778e-14 ***
## Dl_inc      0.144029   0.046402  3.103911 3.8332e-03 **
## LDl_inc     0.033606   0.046377  0.724637 4.7363e-01
## Dl_pop      0.428247   0.149652  2.861615 7.1620e-03 **
## LDl_pop     0.410567   0.172030  2.386604 2.2713e-02 *
## Dl_her_v   -0.004457   0.003411 -1.306403 2.0018e-01
## LDl_her_v  -0.003473   0.002327 -1.492225 1.4486e-01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.006641     Adj. R2: 0.47468
##                  Within R2: 0.34867
```

First stage

```
bordercounties <- df %>% filter(border==1)
reg1 <- feols(Dl_nloans_b ~ Linter_bra + LDl_hpi + Dl_inc + LDl_inc + Dl_pop + LDl_pop + Dl_her_v + LDl_her_v | county + year,
              data = bordercounties, weights = bordercounties$w1,
              cluster = "state_n")
reg2 <- feols(Dl_vloans_b ~ Linter_bra + LDl_hpi + Dl_inc + LDl_inc + Dl_pop + LDl_pop + Dl_her_v + LDl_her_v | county + year,
              data = bordercounties, weights = bordercounties$w1,
              cluster = "state_n")
reg3 <- feols(Dl_lir_b ~ Linter_bra + LDl_hpi + Dl_inc + LDl_inc + Dl_pop + LDl_pop + Dl_her_v + LDl_her_v | county + year,
              data = bordercounties, weights = bordercounties$w1,
              cluster = "state_n")
```

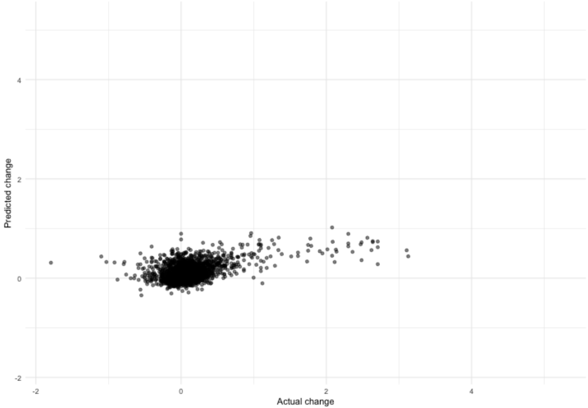
First stage

	Loans	Loan volume	Loan-to-inc. ratio
IV	0.034***	0.034**	0.034***
	(0.011)	(0.013)	(0.012)
House price (lag)	0.280	0.647**	0.653**
	(0.261)	(0.251)	(0.248)
Inc. p.c.	1.37**	1.56***	1.01*
	(0.555)	(0.486)	(0.518)
Inc. p.c. (lag)	0.310	0.682*	0.467
	(0.345)	(0.370)	(0.357)
Population	5.43***	5.48***	4.99***
	(1.34)	(1.56)	(1.65)
Population (lag)	0.115	0.996	0.918
	(1.35)	(1.61)	(1.64)
Herf. index	-0.105***	-0.087**	-0.087**
	(0.033)	(0.033)	(0.034)
Herf. index (lag)	-0.120**	-0.134**	-0.142**
	(0.044)	(0.055)	(0.057)
Observations	2,914	2,914	2,914
F-test for instrument	8.986	6.917	7.803

Standard errors clustered on state in parentheses.

Note: F-test differs from results in paper due to differences in how xtreg calculates standard errors.

First stage predictions vs. actual values. . . what do you notice?



First stage predictions vs. actual values... what do you notice?

	min	max	SD
Actual	-1.792	3.128	0.326
Predicted	-0.346	1.020	0.158

- Note how much less variance there is in the predicted values than the actual values
  - This is the point of using an instrument!
  - We are able to isolate the variation in the endogenous variable that is not correlated with the error term
    - This is of course only a subset of the total variation in the endogenous variable
- This will be important later

We cannot simply use the predicted values in the second stage... standard errors will be wrong!

```r
# create a macro for the main regression controls (to avoid repetition and save space)
setFixest_fml(..controls = ~ LDl_hpi + Dl_inc + LDl_inc + Dl_pop + LDl_pop + Dl_her_v + LDl_her_v)
# Let's use feols to estimate the two stages
reg1 <- feols(Dl_hpi ~ ..controls | county + year | Dl_nloans_b ~ Linter_bra,
              data = df, weights = df$w1,
              cluster = "state_n")
reg2 <- feols(Dl_hpi ~ ..controls | county + year | Dl_vloans_b ~ Linter_bra,
              data = df, weights = df$w1,
              cluster = "state_n")
reg3 <- feols(Dl_hpi ~ ..controls | county + year | Dl_lir_b ~ Linter_bra,
              data = df, weights = df$w1,
              cluster = "state_n")
```

fixest will give us the correct standard errors, however (first stage)

```r
# first stage:
etable(
      reg1, reg2, reg3,
      stage = 1,
      se.below = TRUE,
      depvar = FALSE,
      signif.code = c("***" = 0.01, "**" = 0.05, "*" = 0.1),
      digits = "r3",
      digits.stats = "r0",
      fitstat = c("ivwald", "n"), # make sure to use ivwald for first-stage F-test
      coefstat = "se",
      group = list(controls = "LDl_hpi"),
      keep = "Linter_bra"
    )
```

`fixest` will give us the correct standard errors, however (first stage)

	Loans	Loan volume	Loan-to-inc. ratio
IV (deregulation index)	0.024**	0.025**	0.028**
	(0.011)	(0.012)	(0.012)
controls	Yes	Yes	Yes
Fixed-Effects:			
county	Yes	Yes	Yes
year	Yes	Yes	Yes
Wald (1st stage)	4.332	4.700	5.243
Observations	11,107	11,107	11,107

Standard errors clustered on state in parentheses.

Note: The Wald (similar to F-test) values do not equal the values in the paper due to differences in how xtreg calculates standard errors.

fixest will give us the correct standard errors, however (second stage)

```r
# second stage:
etable(
      reg1, reg2, reg3,
      stage = 2,
      se.below = TRUE,
      depvar = FALSE,
      signif.code = c("***" = 0.01, "**" = 0.05, "*" = 0.1),
      digits = "r3",
      digits.stats = "r3",
      fitstat = c("ivwald", "n"), # make sure to use ivwald for first-stage F-test
      coefstat = "se",
      group = list(controls = "LDl_hpi"),
      keep = c("Dl_nloans_b", "Dl_vloans_b", "Dl_lir_b")
    )
```

`fixest` will give us the correct standard errors, however (second stage)

	(1)	(2)	(3)
Loans	0.141		
	(0.129)		
Loan volume		0.135	
		(0.117)	
Loan-to-inc. ratio			0.120
			(0.099)
controls	Yes	Yes	Yes
Fixed-Effects:			
county	Yes	Yes	Yes
year	Yes	Yes	Yes
Wald (1st stage)	4.332	4.700	5.243
Observations	11,107	11,107	11,107

Standard errors clustered on state in parentheses. The Wald (similar to F-test) values do not equal the values in the paper due to differences in how xtreg calculates standard errors. The standard errors here are more conservative.

Note the syntax for `fixest`

```
feols(y ~ x | fe1 + fe2 | endogenousvar ~ z, ...)
```

```
feols(y ~ x | fe1 + fe2 | endogenousvar1 + endogenousvar2 ~ z1 + z2,
...)
```

- All controls should be in the first stage, as well as the second
  - `fixest` does this for us automatically
- The package also automatically calculates correct standard errors in the second stage
  - For the "generated regressor"

- With just a single instrument and a single endogenous variable, there is a single first stage
- Let's continue with our outcome Y, our endogenous variable X, and our exogenous variables Z (which includes the instrument)
- It turns out that we can write $\hat{\beta}_{IV}$ as:

$$\hat{\beta}_{IV} = \left((Z'Z)^{-1}(Z'X)\right)^{-1}\left((Z'Z)^{-1}(Z'Y)\right) \tag{14}$$

$$\hat{\beta}_{IV} = \left((Z'Z)^{-1}(Z'X)\right)^{-1}\left((Z'Z)^{-1}(Z'Y)\right) \tag{14}$$

- We can immediately see two things:
  - The requirement that Z predicts X is necessary to invert the first term
  - The IV estimate *scales the reduced form by the first stage*

$$\hat{\beta}_{IV} = \left((Z'Z)^{-1}(Z'X)\right)^{-1}\left((Z'Z)^{-1}(Z'Y)\right) \tag{14}$$
$$= (Z'X)^{-1}(Z'Z)(Z'Z)^{-1}(Z'Y) \tag{15}$$
$$= (Z'X)^{-1}(Z'Y) \tag{16}$$

Binary instrument and binary treatment

- Let's consider a binary instrument and a binary treatment
  - Z and D are binary, i.e. $Z, D \in \{0, 1\}$

- It turns out there is a very real case where we can find a valid instrument that is binary
  - Treatment assignment in an RCT!

- Banerjee et al. (2015): The Miracle of Microfinance? Evidence from a Randomized Evaluation (*AEJ: Applied*)
- They are interested in the effects of access to credit on outcomes
  - They randomly assign households (sort of) to microcredit *access*
- Z: whether or not the household was offered microcredit
  - This is a binary instrument
- X: whether or not the household received credit
  - This is a binary endogenous variable

RCTs and IV

- Banerjee et al. (2015): The Miracle of Microfinance? Evidence from a Randomized Evaluation (*AEJ: Applied*)
- They are interested in the effects of access to credit on outcomes
  - They randomly assign households (sort of) to microcredit *access*
- Z: whether or not the household was offered microcredit
  - This is a binary instrument
- X: whether or not the household received credit
  - This is a binary endogenous variable

Effects of the program on outcomes in endline 1

```r
df <- read_dta("banerjeeetal.dta")
# create a macro for the main regression controls (to avoid repetition and save space)
setFixest_fml(..controls = ~ area_pop_base + area_debt_total_base + area_business_total_base + area_exp_pc_mean_base +
                           area_literate_head_base + area_literate_base)
# they control for baseline values of NEIGHBORHOOD means of these variables
```

- They estimate:

$$y_{in} = \beta_0 + \beta_1 Z_n + \sum_{k=1}^{K} \gamma_k X_k + \varepsilon_n, \tag{17}$$

where Z_i is the treatment variable (microcredit access) and standard errors are clustered at the areaid (neighborhood)

```r
reg1 <- feols(any_biz_1 ~ treatment + ..controls,
              data = df, weights = df$w1,
              cluster = "areaid")
reg2 <- feols(bizassets_1 ~ treatment + ..controls,
              data = df, weights = df$w1,
              cluster = "areaid")
reg3 <- feols(bizprofit_1 ~ treatment + ..controls,
              data = df, weights = df$w1,
              cluster = "areaid")
table <- etable(reg1, reg2, reg3,
                digits = 3, fitstat = c("n"), se.below = TRUE, depvar = FALSE,
                # change significance codes to the norm
                signif.code = c("***" = 0.01, "**" = 0.05, "*" = 0.1),
                group = list(controls = "area_pop_base"), keep = "treatment")
```

Reduced form, clean table

	Any biz?	Biz assets	Biz profits
treatment	0.005	421.4	345.7
	(0.019)	(310.8)	(315.9)
controls	Yes	Yes	Yes
Observations	6,186	6,186	6,186

Standard errors clustered on neighborhood in parentheses.

First stage

```r
reg1 <- feols(anymfi_1 ~ treatment + ..controls,
              data = df, weights = df$w1,
              cluster = "areaid")
reg2 <- feols(anyloan_1 ~ treatment + ..controls,
              data = df, weights = df$w1,
              cluster = "areaid")
table <- etable(reg1, reg2,
                digits = 3, fitstat = c("n"), se.below = TRUE, depvar = FALSE,
                # change significance codes to the norm
                signif.code = c("***" = 0.01, "**" = 0.05, "*" = 0.1),
                group = list(controls = "area_pop_base"), keep = "treatment")
```

First stage, clean table

	Any MFI loan?	Any loan?
treatment	0.083***	-0.018
	(0.026)	(0.013)
controls	Yes	Yes
Observations	6,186	6,186

Standard errors clustered on neighborhood in parentheses.

IV results

```
reg1 <- feols(any_biz_1 ~ ..controls | anymfi_1 ~ treatment,
              data = df, weights = df$w1,
              cluster = "areaid")
reg2 <- feols(bizassets_1 ~ ..controls | anymfi_1 ~ treatment,
              data = df, weights = df$w1,
              cluster = "areaid")
reg3 <- feols(bizprofit_1 ~ ..controls | anymfi_1 ~ treatment,
              data = df, weights = df$w1,
              cluster = "areaid")

table <- etable(reg1, reg2, reg3,
                digits = 3, fitstat = c("ivwald", "n"), se.below = TRUE, depvar = FALSE,
                # change significance codes to the norm
                signif.code = c("***" = 0.01, "**" = 0.05, "*" = 0.1),
                group = list(controls = "area_pop_base"), keep = "anymfi_1")
```

	Any biz?	Biz assets	Biz profits
Has MFI loan	0.062	5,092.5	4,177.2
	(0.229)	(4,182.9)	(3,876.0)
controls	Yes	Yes	Yes
Wald (1st stage)	9.8326	9.8326	9.8326
Observations	6,186	6,186	6,186

Standard errors clustered on neighborhood in parentheses.

Putting them together

```r
# reduced form
reg1 <- feols(any_biz_1 ~ treatment + ..controls,
              data = df, weights = df$w1,
              cluster = "areaid")
# first stage
reg2 <- feols(anymfi_1 ~ treatment + ..controls,
              data = df, weights = df$w1,
              cluster = "areaid")
# IV result
reg3 <- feols(any_biz_1 ~ ..controls | anymfi_1 ~ treatment,
              data = df, weights = df$w1,
              cluster = "areaid")
```

- Coefficient on reduced form: 0.00513203819868444
- Coefficient on first stage: 0.0827576895837559
- Coefficient on IV: 0.0620128259329202
    - Can you figure out how this is related to the RF and FS?

Putting them together

```r
# reduced form
reg1 <- feols(any_biz_1 ~ treatment + ..controls,
              data = df, weights = df$w1,
              cluster = "areaid")
# first stage
reg2 <- feols(anymfi_1 ~ treatment + ..controls,
              data = df, weights = df$w1,
              cluster = "areaid")
# IV result
reg3 <- feols(any_biz_1 ~ ..controls | anymfi_1 ~ treatment,
              data = df, weights = df$w1,
              cluster = "areaid")
```

- Coefficient on reduced form: 0.00513203819868440
- Coefficient on first stage: 0.0827576895837590
- Coefficient on IV: 0.0620128259329202
  - Can you figure out how this is related to the RF and FS?
  - This is a ratio: $\frac{\hat{\beta}_{RF}}{\hat{\beta}_{FS}} = \hat{\beta}_{IV}$
  - The IV result *scales the reduced form by the first stage*

Putting them together, the intuition

- The IV estimate is a ratio of two coefficients
  - The reduced form coefficient and the first stage coefficient
- In this example, treatment increases MFI loan take-up by 8.2 percentage points.
  - In other words, the treatment effect is driven by a change in MFI loan take-up among 8.2 percent of households
- If the probability of owning a business goes up by 0.005 (0.5 p.p.), what is the change in probability of owning a business for those who take up the MFI loan?
  - 0.005/0.082! This is the IV estimate

- So this IV estimate is driven by the change in MFI loan take-up among 8.2 percent of households
  - What does this mean for the effect of MFI loans on business ownership?
- Two worlds:
  - Homogeneous treatment effects
  - Heterogeneous treatment effects
- Remember how I said an IV identifies just certain kinds of variation?
  - This will come into play here

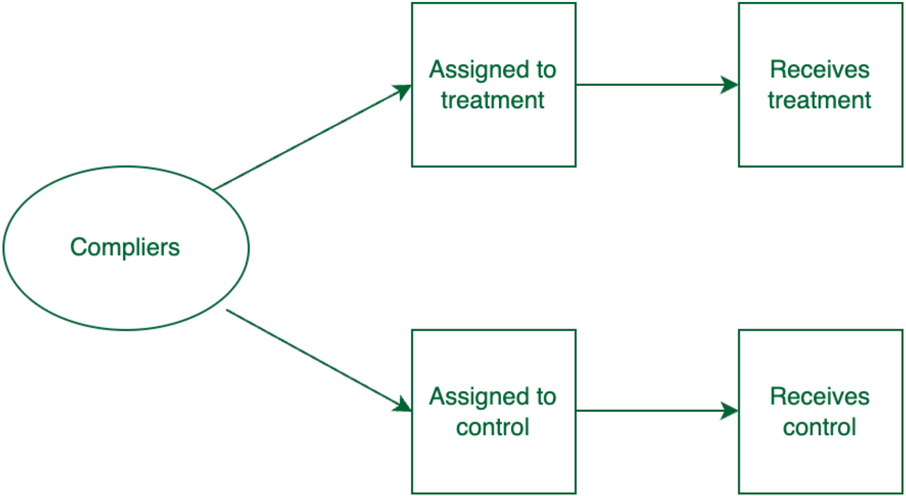
Homogeneous treatment effects

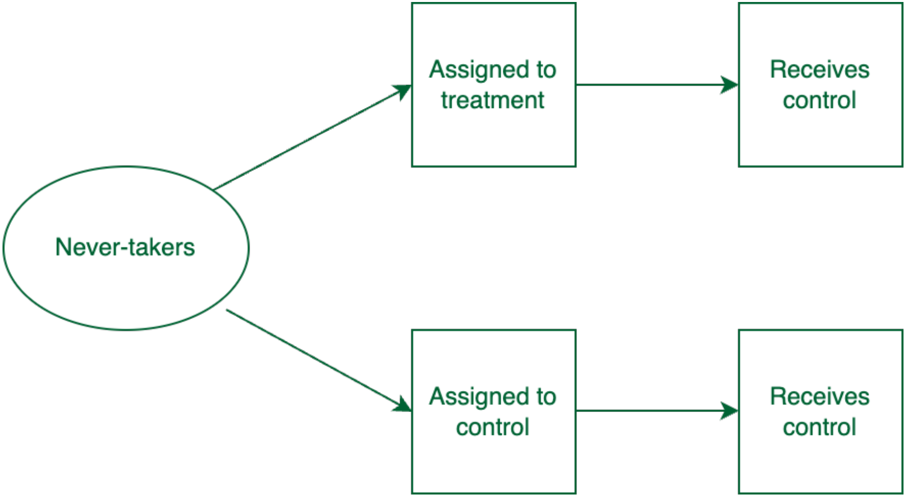
- We had a similar discussion when we talked about DiD
- If everyone has the same treatment effect, then it doesn't matter what variation we isolate
  - All variation will be identifying the same effect
- In this case, the IV is estimated the average treatment effect
- But what if effects are not homogeneous?

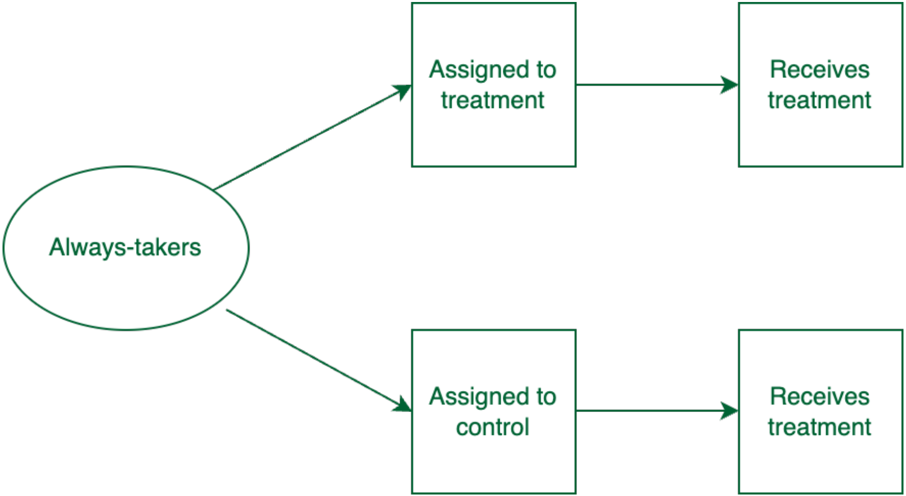
- What if not everyone has the same treatment effect?
  - In other words, what if different types of variation are identifying different effects?
- Imagine a world in which we have an endogenous variable, D
  - Imagine we also have multiple *valid* instruments: Z_1 and Z_2
- If Z_1 and Z_2 are correlated with different "parts" of D, then they can be isolating different variation in D
  - This also means that they IV results can lead to different estimates, even though both instruments are valid!

- We need to define four separate groups:
  - Compliers
  - Always-takers
  - Never-takers
  - Defiers

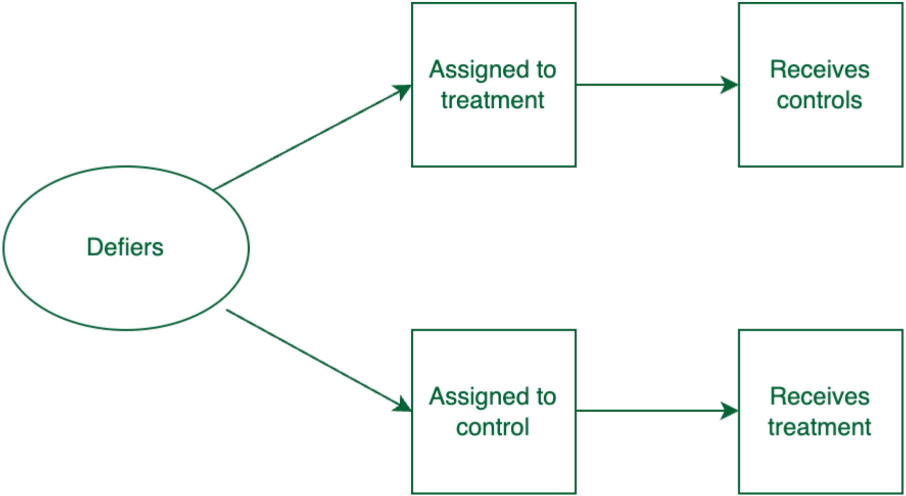
Compliers


Compliers
Assigned to treatment
Receives treatment
Assigned to control
Receives control


Never-takers
Assigned to treatment
Receives control
Assigned to control
Receives control


Always-takers
Assigned to treatment
Receives treatment
Assigned to control
Receives treatment

Defiers


Defiers
Assigned to treatment
Receives controls
Assigned to control
Receives treatment

Comparing the four groups

		Actual status	
		Treatment	Control
Assignment	Treatment	Compliers Always-takers	Never-takers Defiers
	Control	Always-takers Defiers	Compliers Never-takers