Week 7: Instrumental Variables I

PUBL0050 Causal Inference

Julia de Romémont

Term 2 2023-24

UCL Departement of Political Science

This Lecture

Non-Compliance in Experiments

Identification with Instrumental Variables

Characterising the LATE

Two Stage Least Squares

Non-Compliance in Experiments

Running example

Do televised debates change voters' opinions?

In 2005, in the days leading up to the final TV debate between New York mayoral incumbent Michael Bloomberg and his challenger Fernando Ferrer, researchers (Bertrand et. al. 2010) randomly assigned 1000 individuals to two groups: a treatment group encouraged to watch the debate, and a control group encouraged to watch a "placebo" program. Did watching the debate change voters' views of the two candidates?

- Outcome (Y): Opinion change (1 if changed opinion on either candidate)
- Treatment assignment (Z): Encouragement to watch (1 if encouraged)
- Treatment (D): Watched the debate (1 if watched, 0 otherwise)

Encouragement designs and non-compliance

In contrast to other experiments we have studied, in this example (and many others), there is an imperfect match between:

- the units that are assigned the treatment ($Z_i = 1$) and
- the units that received the treatment ($D_i = 1$)

This type of mismatch is known as **non-compliance**.

One-sided non-compliance

When some units assigned to treatment **do not** receive the treatment **or** some units assigned to control **do** receive the treatment

Two-sided non-compliance

When some units assigned to treatment **do not** receive the treatment **and** some units assigned to control **do** receive the treatment

One-sided or two-sided?

1. Medical trial where some patients refuse to take prescribed tablets
   - **One-sided**: Not all treated units receive the treatment
2. Field experiment in which houses are randomly assigned to canvassers, but some people are not home
   - **One-sided**: Not all treated units receive the treatment
3. School where admission is by lottery, but some winning students go to another school and some losing students bribe their way in
   - **Two-sided**: Not all treated units receive the treatment, some control units do receive the treatment

Non-compliance in the TV debate experiment

We can look for evidence of non-compliance in our data:

	Did not watch	Watched
Not encouraged	415	80
Encouraged	320	185

Implications:

- We have evidence of two-sided non-compliance
- In the control group, 16% did receive the treatment
- In the treatment group, 63% did not receive the treatment

Two ways of dealing with non-compliance

When faced with experimental data marked by non-compliance, we might consider two potential ways of estimating causal effects:

1. Difference in group means based on treatment
    - Problem: Selection bias persists
2. Difference in group means based on treatment assignment
    - Problem: may not be the quantity of interest!

Approach 1: DIGM based on treatment (D_i)

Identification under *randomization* of treatment intake D_i

$$\begin{aligned} E[Y_i|D_i=1] &= E[Y_{1i}|D_i=1] = E[Y_{1i}] \\ E[Y_i|D_i=0] &= E[Y_{0i}|D_i=0] = E[Y_{0i}] \end{aligned}$$

$$\tau_{\text{ATE}} = E[Y_{1i} - Y_{0i}] = E[Y_{1i}] - E[Y_{0i}] = \underbrace{E[Y_i|D_i=1] - E[Y_i|D_i=0]}_{\text{Difference in Means}}$$

But now, while the **treatment assignment** is randomized, **treatment intake** is not!

- $D_i = 0$ includes units who did not watch the debate, even though encouraged
- $D_i = 1$ includes units who watched the debate, even though not encouraged

Implication: $E[Y_{0i}|D_i=1] \neq E[Y_{0i}|D_i=0] \rightarrow$ selection bias persists

Approach 2: DIGM based on treatment assignment (Z_i)

By contrast, we can identify the effect of the treatment assignment.

Identification under *randomization* of treatment assignment Z_i

If we redefine Y_{1i} and Y_{0i} to be the potential outcomes for i under different treatment assignments (Z_i), then, given random assignment of Z_i:

$$\begin{aligned} E[Y_i|Z_i=1] &= E[Y_{1i}|Z_i=1] = E[Y_{1i}] \\ E[Y_i|Z_i=0] &= E[Y_{0i}|Z_i=0] = E[Y_{0i}] \end{aligned}$$

$$E[Y_{1i}-Y_{0i}] = \underbrace{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}_{\text{Difference in Means}} = \tau_{\text{ITT}}$$

where τ_{ITT} is the **Intention to Treat** effect.

Approach 2: DIGM based on treatment assignment (Z_i)

What is the τ_{ITT}?

- The ITT is the causal effect of the **offer** of treatment
  - Not everyone took the treatment when offered it.
  - Not everyone who took the treatment was offered it.

Are we interested in τ_{ITT}?

- If we care about whether some program made a difference to average outcomes, then maybe
- If we want to know about the effectiveness of the treatment, not the treatment assignment, maybe not

A third approach

- **Instrumental variables** offer an alternative approach to producing unbiased estimates of causal effects in experiments marked by non-compliance.
- Although the ITT might be interesting in its own right, IV allows us to use the ITT (the effect of treatment assignment) to estimate **an** ATE (the effect of treatment).
  - Note that IV will not give us **the** ATE - it will give us an average treatment effect for a particular subgroup
- This can be especially useful when randomizing treatment intake is either impossible or unethical.[1]

[1] See MHE p. 165 for a good example

Historical Aside

- Instrumental variables have a long history in economics particularly in the context of structural equation models.[2]
- Typically, the IV estimates causal effects in SEMs are based on an assumption of constant treatment effects, which is somewhat unsatisfying.
- We will situate IV approaches within the **potential outcomes framework**:
  - Allows for heterogeneous effects
  - Makes clear the assumptions needed for causal interpretation
- Note that the conclusions drawn from this presentation apply regardless of how individual authors motivate their IV strategies!

[2] See Angrist et al 1996

Identification with Instrumental Variables

Potential outcome model for instrumental variables

Definition: Instrument

Z_i: Binary instrument for *{unit}* i.

$$Z_i = \begin{cases} 1 & \text{if unit } i \text{ "encouraged" to receive treatment} \\ 0 & \text{if unit } i \text{ "encouraged" to receive control} \end{cases}$$

Definition: Potential Treatments

D_{zi} indicates *potential* treatment status given $Z_i = z$. E.g.

- $D_{1i} = 1$ encouraged to take treatment and takes treatment
- $D_{0i} = 1$ not encouraged to take treatment but takes it anyway

Assumption

Observed treatments (D_i) are connected to the **potential treatments** (D_{1i}, D_{0i}) as follows:

$$D_i = Z_i \cdot D_{1i} + (1 - Z_i) \cdot D_{0i} \quad \text{so} \quad D_i = \begin{cases} D_{1i} & \text{if } Z_i = 1 \\ D_{0i} & \text{if } Z_i = 0 \end{cases}$$

Potential outcome model for instrumental variables

Given these, we can classify units by their potential treatments:

Definitions

- ***Compliers***: take treatment when assigned, do not take it when not assigned
$$D_{1i} > D_{0i} \quad \text{(i.e. } D_{0i} = 0 \text{ and } D_{1i} = 1\text{)}$$
- ***Always-takers***: always take the treatment, whether assigned or not
$$D_{1i} = D_{0i} = 1$$
- ***Never-takers***: never take the treatment, whether assigned or not
$$D_{1i} = D_{0i} = 0$$
- ***Defiers***: do not take treatment when assigned, take it when not assigned
$$D_{1i} < D_{0i} \quad \text{(i.e. } D_{0i} = 1 \text{ and } D_{1i} = 0\text{)}$$

Potential outcome model for instrumental variables

	Not assigned to treatment ($Z_i = 0$)	Assigned to Treatment ($Z_i = 1$)
Not treated ($D_i = 0$)	Never-Taker (N) or Complier (C)	Never-Taker (N) or Defier (D)
Treated ($D_i = 1$)	Always-Taker (A) or Defier (D)	Always-Taker (A) or Complier (C)

- Problem: we cannot know the type of any given individual! **Why?**
- We only observe either D_{1i} or D_{0i} for a given unit $\rightarrow$ FPOCI
  - When $Z_i = 1$, $D_i = 1$ for both Always-Takers and Compliers
  - When $Z_i = 0$, $D_i = 1$ for both Always-Takers and Defiers
- However, we can identify the *proportion* of each type (under certain assumptions).

Assumptions for estimating compliance proportions

Assumption I: Independence of the instrument

$(Y_{0i}, Y_{1i}, D_{0i}, D_{1i}) \perp\!\!\!\perp Z_i$

i.e. the **instrument** is assigned at random (or, as good as random)

Implies that the causal effect of Z_i on Y_i and on D_i are identified.

Assumption II: First stage

$0 < P(Z_i = 1) < 1$ & $P(D_1 = 1) \neq P(D_0 = 1)$

i.e. the instrument Z_i induces some variation in D_i (there need to be some compliers!).

Assumption III: Monotonicity

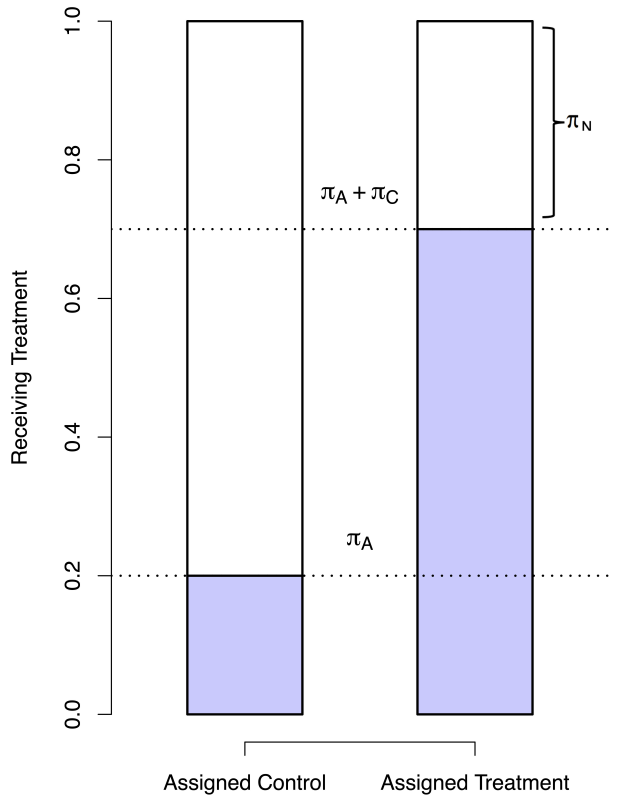
$D_{1i} \geq D_{0i}$

i.e. there are no defiers.

Estimating compliance proportions

With these assumptions:

- $\pi_A = E[D_i = 1|Z_i = 0]$
  (No defiers)
- $\pi_C =$
  $E[D_i = 1|Z_i = 1] - E[D_i = 1|Z_i = 0]$
  (Independence and first stage)
- $\pi_N = 1 - E[D_i = 1|Z_i = 1]$
  (No defiers)



$\Rightarrow$ With these, and one further assumption, we can estimate the ATE for the **compliers**

- i.e. **only for those who are actually responsive to treatment**

Assumption to go from ITT to LATE

Assumption IV: Exclusion restriction

$$Y(D_i = 1, Z_i = 1) = Y(D_i = 1, Z_i = 0)$$

and

$$Y(D_i = 0, Z_i = 1) = Y(D_i = 0, Z_i = 0)$$

i.e. the treatment assignment **only** affects Y by affecting the treatment received

Estimand (LATE)

$\tau_{LATE} = E[Y_{1i} - Y_{0i} | D_{1i} > D_{0i}]$
*is defined as the **L**ocal **A**verage **T**reatment **E**ffect for Compliers.*

- This estimand varies with the particular instrument Z
- Whether compliers are interesting depends on instrument

ITT and LATE

To see why we can recover the LATE for compliers, notice that we can express the ITT as a weighted average of the unit-type specific LATEs:

$$\begin{aligned} ITT &= LATE_C\pi_C + LATE_D\pi_D + LATE_A\pi_A + LATE_N\pi_N \\ &= LATE_C\pi_C + LATE_D 0 + LATE_A\pi_A + LATE_N\pi_N \quad \text{(No defiers)} \\ &= LATE_C\pi_C + LATE_D 0 + 0\pi_A + 0\pi_N \quad \text{(Exclusion restriction)} \\ &= LATE_C\pi_C \end{aligned}$$

This implies that we can recover the LATE for compliers via:

$$LATE_C = \frac{ITT}{\pi_C} = E[Y_{1i} - Y_{0i} | D_{1i} > D_{0i}]$$

Interpretation: LATE is the average causal effect for units whose treatment status is entirely determined by the instrument

What have we assumed to get here?

1. Independence of Z_i
   - I.e. random assignment of Z_i
   - Uncontroversial when dealing with a randomized experiment
2. Some compliers (first stage)
   - Generally uncontroversial, and typically easy to check empirically
   - Testable! Regress D_i on Z_i, and check for significance
3. No defiers (monotonicity)
   - Normally uncontroversial (most people aren't *that* weird), but untestable
4. Exclusion restriction
   - Impossible to check empirically
   - Often controversial with observational data as it may often suggest a very unrealistic data generating process
   - Much more on this next week...

Estimating the LATE (I)

Estimand (LATE)

$\tau_{LATE} = E[Y_{1i} - Y_{0i} | D_{1i} > D_{0i}]$
*i.e. the **L**ocal **A**verage **T**reatment **E**ffect (for compliers)*

Estimator (Wald Estimator)

If Z_i is randomized (independence), an unbiased estimator for the LATE is:

$$\begin{aligned} LATE &= \frac{ITT}{\pi_C} = \frac{E[Y_i|Z_i=1] - E[Y_i|Z_i=0]}{E[D_i|Z_i=1] - E[D_i|Z_i=0]} \\ &= \frac{\textit{Effect of } Z_i \textit{ on } Y_i}{\textit{Effect of } Z_i \textit{ on } D_i} = \frac{\bar{Y}_{Z_i=1} - \bar{Y}_{Z_i=0}}{\bar{D}_{Z_i=1} - \bar{D}_{Z_i=0}} \end{aligned}$$

An equivalent "ratio of coefficients" approach:

$$\begin{aligned} \textit{First stage:} \quad D_i &= \alpha_1 + \beta_1 Z_i + \epsilon_{1i} \\ \textit{Reduced form:} \quad Y_i &= \alpha_2 + \beta_2 Z_i + \epsilon_{2i} \\ LATE &= \frac{\hat{\beta}_2}{\hat{\beta}_1} = \frac{\textit{Effect of } Z_i \textit{ on } Y_i}{\textit{Effect of } Z_i \textit{ on } D_i} \end{aligned}$$

Estimating the LATE in practice

```
# Estimating LATE by the ratio of two differences in means
## Step 1: Effect of Z on Y
y_diff_in_means <-
  mean(debate$changed_opinion[debate$encouraged == 1]) -
  mean(debate$changed_opinion[debate$encouraged == 0])

y_diff_in_means
```

```
## [1] 0.05706571
```

Interpretation: The ITT suggests that assignment to treatment caused a 5.7 percentage point increase, on average, in the probability of changing opinion.

Estimating the LATE in practice

```
# Estimating LATE by the ratio of two differences in means
## Step 2: Effect of Z on D
d_diff_in_means <-
  mean(debate$watched[debate$encouraged == 1]) -
  mean(debate$watched[debate$encouraged == 0])

d_diff_in_means
```

```
## [1] 0.2047205
```

Interpretation: Assignment to treatment caused a 20 percentage point increase, on average, in the probability of watching the debate.

Estimating the LATE in practice

```
# Estimating LATE by the ratio of two differences in means
## Step 3: Effect of Z on Y divided by Effect of Z on D
y_diff_in_means / d_diff_in_means
```

```
## [1] 0.2787494
```

Interpretation: Actually *watching* the debate caused a 28 percentage point increase in the probability of changing opinion, on average **for the compliers**.

Estimating the LATE in practice

We can retrieve exactly the same result using the 'ratio of coefficients' method:

```
# Estimating LATE by the ratio of first stage and reduced
# form coefficients
## Step 1: Effect of Z on Y
first_stage <- lm(watched ~ encouraged,
                  data = debate)

## Step 2: Effect of Z on D
reduced_form <- lm(changed_opinion ~ encouraged,
                   data = debate)

## Step 3: Ratio of effects
coef(reduced_form)["encouraged"]/
  coef(first_stage)["encouraged"]
```

```
## encouraged
##  0.2787494
```

Implications of the IV formula

$$\begin{array}{lrcl} \text{First stage:} & D_i & = & \alpha_1 + \beta_1 Z_i + \epsilon_{1i} \\ \text{Reduced form:} & Y_i & = & \alpha_2 + \beta_2 Z_i + \epsilon_{2i} \end{array} \quad \Rightarrow \quad \text{LATE} = \frac{\hat{\beta}_2}{\hat{\beta}_1}$$

- When the effect of Z_i in the reduced form in zero, the LATE is also zero.

$$\text{LATE} = \frac{\hat{\beta}_2}{\hat{\beta}_1} = \frac{0}{\hat{\beta}_1} = 0$$

- When the effect of Z_i in the first stage is zero, the reduced form should also be zero.

$$\hat{\beta}_2 = \text{LATE} \cdot \hat{\beta}_1 = \text{LATE} \cdot 0 = 0$$

Implication: If the first stage is zero but we have a significant reduced form, then the instrument must be affecting the outcome through something other than the treatment.

- This would be a violation of the **exclusion restriction**

Questions

Task: Recall our example of the TV debate experiment. Spend a few minutes considering the following questions.

- What does the intention-to-treat (ITT) effect mean here?
- Is the no-defier assumption reasonable?
- What is the exclusion restriction in this example? Do you think it holds?

Characterising the LATE

How "local" is the LATE?[3]

- We definied the LATE as the average causal effect of the treatment for the population of compliers.
- We cannot tell which individual units are compliers, so we cannot say anything directly about 'what they look like'.
- To the extent that the compliers are unlike other units, our LATE may not have very high **external validity**.
- Although we cannot observe D_{0i} and D_{1i} for any individual, we can describe how, **on average**, compliers differ from never-takers and always takers.

[3] See Marbach & Hangartner 2020

Some always and never-takers are observable

- Units assigned to the control group ($Z_i = 0$) who take the treatment ($D_i = 1$) are "observable" always-takers
  - Assuming there are no defiers!
- Units assigned to the treatment group ($Z_i = 1$) who do not take the treatment ($D_i = 0$) are "observable" never-takers
  - Assuming there are no defiers!
- Because Z_i is randomised, observable and unobservable never-/always-takers (NT/AT) will have the same covariate distributions
  - As long as randomisation was successful!

Who are the compliers?

- We cannot directly observe covariate distributions for compliers because
  - control group compliers are indistinguishable from control never-takers
  - treatment group compliers are indistinguishable from treatment always-takers
- We can, however, calculate their covariate means by subtracting the means for the never-takers and always-takers from the means of the entire sample

Using the AT & NT to learn about compliers

Example: % women among never-takers, always-takers and compliers

1. Estimate the proportion of the entire sample who are female:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

2. Estimate the proportion of the always-takers who are female:

$$\hat{\mu}_{at} = \frac{1}{K_{at}} \sum_{i=1}^{K_{at}} x_i$$

   where K_{at} is the total number of units for whom $Z = 0$ and $D = 1$

3. Estimate the proportion of the never-takers who are female:

$$\hat{\mu}_{nt} = \frac{1}{K_{nt}} \sum_{i=1}^{K_{nt}} x_i$$

   where K_{nt} is the total number of units for whom $Z = 1$ and $D = 0$

Using the AT & NT to learn about compliers

Example: % women among never-takers, always-takers and compliers

4. Estimate the sample fraction of never-takers, always-takers, and compliers

$$\begin{aligned} \pi_{nt} &= E[D_i = 0 | Z_i = 1] \\ \pi_{at} &= E[D_i = 1 | Z_i = 0] \\ \pi_c &= E[D_i = 1 | Z_i = 1] - \pi_{at} \end{aligned}$$

5. Estimate the proportion of compliers who are female:

$$\hat{\mu} = \hat{\pi}_{co}\hat{\mu}_{co} + \hat{\pi}_{nt}\hat{\mu}_{nt} + \hat{\pi}_{at}\hat{\mu}_{at}$$

$$\Leftrightarrow$$

$$\hat{\mu}_{co} = \frac{1}{\hat{\pi}_{co}}\hat{\mu} - \frac{\hat{\pi}_{nt}}{\hat{\pi}_{co}}\hat{\mu}_{nt} - \frac{\hat{\pi}_{at}}{\hat{\pi}_{co}}\hat{\mu}_{at}$$

Implication: The covariate mean for compliers is the (weighted) difference between the sample mean and the mean for always-takers and compliers.

Describing compliers in R

```
library(ivdesc)
female <- ivdesc(X = debate$female, D = debate$watched,
                 Z = debate$encouraged)
female$mu.lo <- female$mu - 1.96*female$mu_se
female$mu.hi <- female$mu + 1.96*female$mu_se
female$group <- c("Sample mean","Compliers",
                  "Never Takers", "Always Takers")
```

group	mu	mu_se	pi	pi_se	mu.lo	mu.hi
Sample mean	0.611	0.015	1.000	0.000	0.581	0.641
Compliers	0.567	0.073	0.205	0.027	0.423	0.710
Never Takers	0.628	0.027	0.634	0.020	0.576	0.680
Always Takers	0.600	0.056	0.162	0.017	0.490	0.710

Code for other covariates

```
# Race
white <- ivdesc(X = debate$white, D = debate$watched,
                Z = debate$encouraged)
white$mu.lo <- white$mu - 1.96*white$mu_se
white$mu.hi <- white$mu + 1.96*white$mu_se
white$group <- c("Sample mean","Compliers","Never Takers", "Always Takers")

# College Education
grad <- ivdesc(X = debate$graduated, D = debate$watched,
                Z = debate$encouraged)
grad$mu.lo <- grad$mu - 1.96*grad$mu_se
grad$mu.hi <- grad$mu + 1.96*grad$mu_se
grad$group <- c("Sample mean","Compliers","Never Takers", "Always Takers")
```

Code for figures

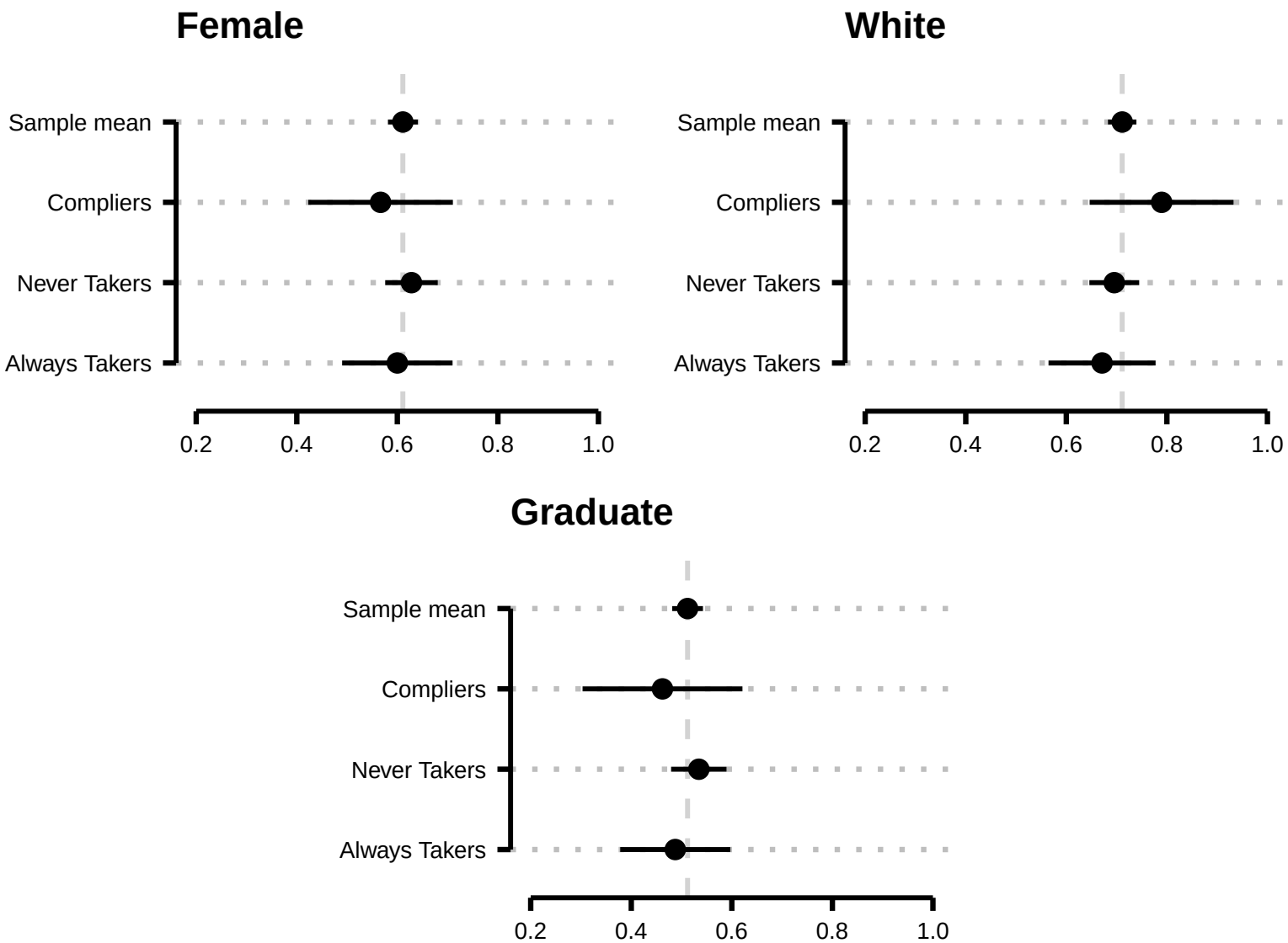
```
library(ggplot2)
library(ggthemes)
library(ggpubr)
# install.packages("lemon")

p.female <- ggplot(female, aes(x=mu,y=group)) +
  geom_vline(aes(xintercept = mu[group=="Sample mean"]),
             col="lightgray",linetype="dashed") +
  geom_point(size=1) + geom_linerange(aes(xmin = mu.lo, xmax=mu.hi)) +
  xlim(0.2,1) + labs(title = "Female") + theme_clean() +
  lemon::coord_capped_cart(left = "both",bottom = "both") +
  theme(axis.title = element_blank(),
        plot.background = element_rect(color=NA))

## Repeat for white and graduates (code omitted here)
p.white <- ggplot(white, aes(x=mu,y=group)) +
  ... + labs(title = "White") + ...
p.grad <- ggplot(grad, aes(x=mu,y=group)) +
  ... + labs(title = "Graduate") +...

## Combine in a grid
ggarrange(p.female,p.white,p.grad)
```

Describing compliers in R



So - how "local" is the LATE?

Implications: Compared to the distribution of these covariates in the sample, compliers are ...

- ...somewhat less likely to be female
- ...somewhat more likely to be white
- ...somewhat less likely to be graduates

Note that the complier group depends on the instrument! Different IVs will lead to different estimands.

LATE special cases

1. When treatment intake, D_i, is itself randomized, then:
   - $Z_i = D_i$ for all i
     - i.e every individual is a complier
   - $\rightarrow \tau_{LATE} = \tau_{ATE}$
2. When treatment effects are homogenous, then:
   - $\tau_i = \tau$ for all i
     - i.e. every unit is affected by treatment in the same way
   - This is the traditional constant-effects IV assumption
   - $\rightarrow \tau_{LATE} = \tau_{ATE}$
3. When non-compliance is one-sided such that $D_{0i} = 0$ for all i
   - $E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}] = E[Y_{1i} - Y_{0i}|D_i = 1]$
     - i.e. there are no always-takers
   - $\rightarrow \tau_{LATE} = \tau_{ATT}$

Two Stage Least Squares

The approach so far

Our approach up to here has assumed that we have a single binary Z_i and a binary D_i. With this and other assumptions, we can estimate:

- the **intention-to-treat (ITT)** via:

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$$

- the **proportion of compliers (π_C)** via:

$$E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$$

- the **LATE** via:

$$\frac{ITT}{\pi_C}$$

However, it would be nice to be able to generalise this to include:

- Non-binary treatments/instruments
- Covariates (for when Z_i is not randomly assigned)
- More than one instrument

Estimating the LATE (II)

Estimand (LATE)

$\tau_{LATE} = E[Y_{1i} - Y_{0i} | D_{1i} > D_{0i}]$
*i.e. the **L**ocal **A**verage **T**reatment **E**ffect (for compliers)*

Estimator (Two Stage Least Squares (2SLS))

$$\begin{aligned} \textit{First stage:} \quad D_i &= \alpha_1 + \beta_1 Z_i + \epsilon_{1i} \\ \textit{Second stage:} \quad Y_i &= \alpha_2 + \beta_2 \hat{D}_i + \epsilon_{2i} \\ LATE &= \hat{\beta}_2 \end{aligned}$$

where $\hat{D}_i$ is the fitted value for each unit from the first stage model.

"2 stage least squares" because we estimate OLS models in 2 stages.

2SLS intuition

1. $\hat{\beta}_1$ measures the average amount of change in D induced by a unit change in Z
2. Variation in $\hat{D}_i$ reflects only the predicted amount of D caused by a unit's value of Z
3. $\hat{\beta}_2$ represents an unbiased estimate of D on Y for those for whom a shift in Z causes higher values of D
4. $\hat{\beta}_2$ should be interpreted as the effect of a one-unit increase in $\hat{D}_i$ on the outcome Y, **for the compliers**

2SLS example in R

```
# Wald estimator
first_stage <- lm(watched ~ encouraged, data = debate)
reduced_form <- lm(changed_opinion ~ encouraged, data = debate)

coef(reduced_form)[2]/coef(first_stage)[2]
```

```
## encouraged
##  0.2787494
```

Compare to the 2SLS estimate:

```
# 2SLS estimator
first_stage <- lm(watched ~ encouraged, data = debate)
debate$fitted_d <- predict(first_stage)
second_stage <- lm(changed_opinion ~ fitted_d, data = debate)

summary(second_stage)$coefficients[2,1:2]
```

```
##   Estimate Std. Error
##  0.2787494  0.1535059
```

2SLS example in R

While this two-step approach clarifies intuition, it will lead to incorrect standard errors. Instead, we use:

```
library(AER)
iv_model <- ivreg(formula = changed_opinion ~ watched,
                  instruments = ~ encouraged,
                  data = debate)
summary(iv_model)
```

```
...
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.37313    0.04346   8.585   <2e-16 ***
## watched      0.27875    0.15299   1.822   0.0688 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4952 on 998 degrees of freedom
## Multiple R-Squared: 0.00992, Adjusted R-squared: 0.008928
...
```

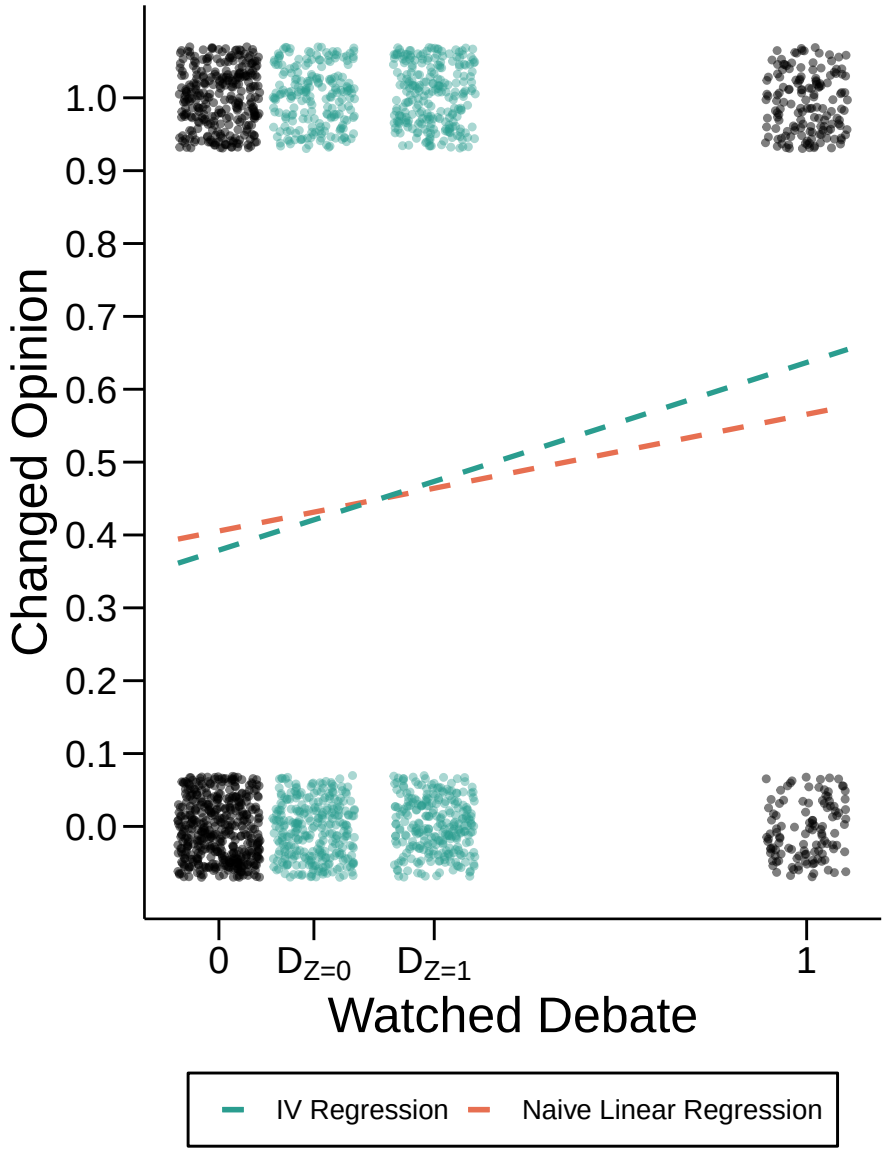
2SLS intuition

Wald estimator:

- Regressing Y_i on Z_i gives the ITT
- Dividing by $E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$ inflates to give the LATE

2SLS:

- Regressing D_i on Z_i gives fitted values $E[D_i|Z_i = 1]$ and $E[D_i|Z_i = 0]$
- Regressing Y_i on $\hat{D}_i$ inflates to give the LATE



Why use 2SLS?

In comparison to the Wald estimator, 2SLS is far more general:

- Can accommodate non-binary instruments and non-binary treatments
- Can accommodate covariates
- Can accommodate multiple instruments

For example to include covariates:

$$\begin{aligned} \text{First stage:} \quad &= \quad D_i = \alpha_1^1 + \beta_1^1 Z_i + \beta_2^1 X_{1i} + \beta_3^1 X_{2i} + \epsilon_{1i} \\ \text{Second stage:} \quad &= \quad Y_i = \alpha_1^2 + \beta_1^2 \hat{D}_i + \beta_2^2 X_{1i} + \beta_3^2 X_{2i} + \epsilon_{2i} \end{aligned}$$

Note: essential to use the *same* covariates in first and second stage models **(more on this next week)**.

Control vs Instrument

- In lecture 4 we saw that, when we **control for a confounder** X_i in a regression to estimate the effect of D_i on Y_i, ...
  - ... β_1 measures the relationship between Y_i and the part of D_i that is **"not explained"** by X_i (i.e. the residuals)
- Instead, when we are **using an instrument** Z_i to estimate the effect of D_i on Y_i,...
  - ... β_1^2 measures the relationship between Y_i and **only** the part of D_i that is **"explained"** by Z_i (i.e. the first stage)

Controlling for a Confounder

```
# Regression anatomy
treat_reg <- lm(watched ~ female, data = debate)
debate$residuals <- resid(treat_reg) ## !!!!!!!!!!
resid_reg <- lm(changed_opinion ~ residuals, data=debate)
# Which is equivalent to:
long_reg <- lm(changed_opinion ~ watched + female, data = debate)

coef(resid_reg)
```

```
## (Intercept)   residuals
##   0.4470000   0.1610594
```

Instrumenting the Treatment

```
# IV
treat_reg <- lm(watched ~ encouraged, data = debate)
debate$fitted <- predict(treat_reg) ## !!!!!!!!!!
second_stage <- lm(changed_opinion ~ fitted, data=debate)
coef(second_stage)
```

```
## (Intercept)      fitted
##   0.3731314   0.2787494
```

Example II

Job Training Partnership Act, 1982

The JPTA was the largest randomized job training scheme ever implemented in the US. This scheme was implemented for people who had previously been disadvantaged in the labour market (unemployed or low waged), and provided job and job-search training. The sample is roughly 20,000 individuals who were randomly assigned to treatment (training) or control (no training) groups.

- Outcome (Y): Wages 30 months after training (US$)
- Treatment assignment (Z): Place on training course (1 if given a place)
- Treatment (D): Participation in training (1 if participated)

Non-compliance in the JTPA experiment

	Not Enrolled	Enrolled	Total
Assigned to Control	3663	54	3717
Assigned to Treatment	2683	4804	7487
Total	6346	4858	11204

What type of non-compliance do we observe here? What are the implications for the LATE?

- Almost perfect one-sided non-compliance
- About 2% of units assigned to control participated in training
- About 40% of units assigned to treatment did not participate
- LATE will be very close to the ATT

Naive estimate in the JTPA experiment

Let's ignore the non-compliance and estimate the effect of the treatment:

```
naive_model <- lm(earnings ~ training, data= jtpa)
coef(naive_model)
```

```
## (Intercept)    training
##   14605.085    2791.088
```

Is this estimate causally identified? If not, what is the likely bias?

- It is not identified!
  - Units assigned to treatment were free to decline (and many did)
  - So this is only an unbiased estimate of ATE if the decision to comply was random (which is very unlikely)
- The bias is probably positive: those who comply with their treatment assignment are probably more motivated

First stage and reduced form in JTPA

```
first_stage <- lm(training ~ assignmt, data= jtpa)
coef(first_stage)
```

```
## (Intercept)    assignmt
##  0.01452785  0.62711767
```

```
reduced_form <- lm(earnings ~ assignmt, data= jtpa)
coef(reduced_form) ## Intention to Treat
```

```
## (Intercept)    assignmt
##   15040.504    1159.433
```

Questions:

- What is the LATE?
- What is the exclusion restriction in this example? Is it plausible?

LATE in the JTPA experiment

The LATE is given by the ratio of the reduced form and first stage coefficients (i.e. $\frac{\hat{\beta}_2}{\hat{\beta}_1}$):

```
# Wald estimator
coef(reduced_form)[2]/coef(first_stage)[2]
```

```
## assignmt
## 1848.829
```

Interpretation: This suggests that the training programme increased wages by $1800 on average for those who complied with the treatment.

Note that the LATE is:

- about 50% larger than the ITT (~$1160)
- about 35% smaller than the naive estimate (~$2800)

LATE or ITT in the JTPA experiment

Which is more important in the JTPA: LATE or ITT?

- The case for LATE
  - Provides the average causal effect for those who complied with treatment
  - If we are most interested in how effective training is for treated individuals, then this is the salient quantity
  - The LATE provides relevant information if we want to know about training effectiveness
- The case for ITT
  - Provides the average causal effect of assigning units to treatment
  - If we are most interested in the benefits of the training programme as a whole, then this is the salient quantity
  - The ITT provides relevant information regarding the cost/benefit calculation of most policy-makers

Conclusion

- Even when treatments are randomly assigned in an experiment, some units will fail to comply with the treatment condition they are assigned.
- IV estimators help to solve the problem of non-compliance when we are dealing with experimental data.
- Next week we will focus on a far more prevalent use of IV $\rightarrow$ to address omitted variable bias in observational data.
- The mechanics of IV will remain the same, but we will be **much** more dubious about whether the relevant assumptions are being met.