3. Instrumental Variables

Econometrics II
Winter 2019
Osaka U

Shuhei Kitamura

Summary of last lecture

Treatment effect

Average treatment effect (ATE)
- Independence assumption
- Average Effect of Treatment on the Treated (ATT)

Is the effect statistically significant?
- Statistical significance
- Confidence interval

Is the effect large?
- Effect size

Conducting an experiment

    Power calculation, MDE

Regression analysis
- Conditional expectation function
- Conditional independence assumption (CIA)
- Regression and matching
- Omitted-variable bias
- Bad control
- Measurement error

Outline

Imperfect compliance

Local average treatment effect (LATE)

Intention to treat (ITT) effect

Two-stage least squares (2SLS)

IV assumptions

IV solves measurement error

Imperfect compliance

So far we have implicitly assumed that all individuals who are treated are treated, and all individuals who are not treated are not treated.

However, this is not always the case in practice.

- Examples: Going to a school by getting an acceptance letter to that school, using a mosquito net by getting it for free, etc.

We call the individuals who do not comply with the experiment non-compliers. The existence of such individuals makes us consider the case of randomized experiments with imperfect compliance.

KIPP schooling

During the 1990s, the veterans of Teach for America started charter schools called the KIPP (Knowledge Is Power Program).

The first KIPP school in New England started in Lynn, MA.

The school was oversubscribed after 2005. Therefore, charter seats were allocated by lottery as required by the Massachusetts law.
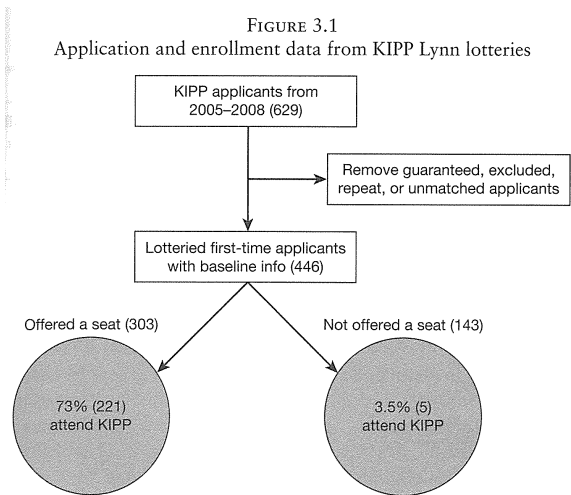
Thus, those who were and were not offered a seat were randomly determined.

KIPP schooling (cont.)

Between 2005-2008, 303 were offered a seat, while 143 weren't.

- Out of those who were offered, 221 (73%) appeared at KIPP the following school year, and out of those who weren't, 5 (3.5%) found their way into KIPP.

KIPP schooling (cont.)

FIGURE 3.1
Application and enrollment data from KIPP Lynn lotteries

KIPP applicants from
2005–2008 (629)

Remove guaranteed, excluded,
repeat, or unmatched applicants

Lotteried first-time applicants
with baseline info (446)

Offered a seat (303)

73% (221)
attend KIPP

Not offered a seat (143)

3.5% (5)
attend KIPP

Note: Numbers of Knowledge Is Power Program (KIPP) applicants are
shown in parentheses.

Source: Metrics.

KIPP schooling (cont.)

If the assignment was truly random, demographic characteristics should not be systematically different between those who won the lottery and those who did not.

Let's check it.

KIPP schooling (cont.)

TABLE 3.1
Analysis of KIPP lotteries

| | Lynn public fifth graders (1) | KIPP applicants | | | |
		KIPP Lynn lottery winners (2)	Winners vs. losers (3)	Attended KIPP (4)	Attended KIPP vs. others (5)
Panel A. Baseline characteristics					
Hispanic	.418	.510	−.058 (.058)	.539	.012 (.054)
Black	.173	.257	.026 (.047)	.240	−.001 (.043)
Female	.480	.494	−.008 (.059)	.495	−.009 (.055)
Free/Reduced price lunch	.770	.814	−.032 (.046)	.828	.011 (.042)
Baseline (4th grade) math score	−.307	−.290	.102 (.120)	−.289	.069 (.109)
Baseline (4th grade) verbal score	−.356	−.386	.063 (.125)	−.368	.088 (.114)
Panel B. Outcomes					
Attended KIPP	.000	.787	.741 (.037)	1.000	1.000 —
Math score	−.363	−.003	.355 (.115)	.095	.467 (.103)
Verbal score	−.417	−.262	.113 (.122)	−.211	.211 (.109)
Sample size	3,964	253	371	204	371

Notes: This table describes baseline characteristics of Lynn fifth graders and reports estimated offer effects for Knowledge Is Power Program (KIPP) Lynn applicants. Means appear in columns (1), (2), and (4). Column (3) shows differences between lottery winners and losers. These are coefficients from regressions that control for risk sets, namely, dummies for year and grade of application and the presence of a sibling applicant. Column (5) shows differences between KIPP students and applicants who did not attend KIPP. Standard errors are reported in parentheses.

LATE

However; individual's choice to attend/not to attend the KIPP school was not random, even though the assignment was random.

You can apply the Instrumental Variables (IV) method to such cases. Since

$$
\begin{aligned}
& \text{Effect of offers on scores} \\
&= \text{Effect of offers on attendance} \\
&\times \text{Effect of attendance on scores,} \qquad (1)
\end{aligned}
$$

The causal effect of KIPP attendance on scores is

$$
= \frac{\overset{\displaystyle \text{Effect of attendance on scores}}{\text{Effect of offers on scores}}}{\text{Effect of offers on attendance}} \qquad (2)
$$

What happens to the denominator if there is no compliance issue?

LATE (cont.)

This effect is called a Local Average Treatment Effect (LATE).

$$= \frac{\text{Effect of attendance on scores}}{\text{Effect of offers on attendance}}$$

Effect of offers on scores

The numerator is called the reduced form, while the denominator is called the *first stage*.

- You will know why they are called as such soon.

Using mathematical notations, LATE is written by

$$\lambda = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}, \tag{3}$$

Where Z_i is called an instrument.

- In the KIPP example, the instrument takes the value of one if a pupil was offered a seat, and zero otherwise.
- Z_i should satisfy several assumptions. We will look into them later.

The sample version of (3) is called an *IV estimator*.

- The IV estimator recovers LATE.

KIPP schooling (cont.)

How can we interpret λ in words?

There are four types of pupils in the KIPP example

		Lottery losers ($Z_i = 0$)	
		Doesn't attend ($D_i = 0$)	Attend ($D_i = 1$)
Lottery Winners ($Z_i = 1$)	Doesn't attend ($D_i = 0$)	Never-takers	Defiers
	Attend ($D_i = 1$)	Compliers	Always-takers

LATE captures the effect on *compliers*.

KIPP schooling (cont.)

It needn't capture causal effects on never-takers or always-takers without stronger assumptions than, e.g., a constant causal effect for everybody.

What about defiers?

- The average effect can be zero because of defiers! The *monotonicity* assumption is required to exclude the possibility of defiers, which assumes that no one behaves in the opposite direction if treated.
- Defiers are unlikely to exist in the KIPP example = the monotonicity assumption is likely to be satisfied.

LATE (cont.)

Thus LATE is rewritten by

$$\lambda = E[Y_i(1) - Y_i(0)|C_i = 1], \tag{4}$$

Where $C_i = 1$ means compliers.

Recall the definition of ATT (or TOT)

$$E[Y_i(1) - Y_i(0)|D_i = 1]. \tag{5}$$

LATE $\neq$ ATT as long as $C_i \neq D_i$.

- ATT is the weighted average of effects on compliers and always-takers.

LATE (cont.)

In the KIPP example, pupils who attended the KIPP may include both compliers and always-takers.

- If always-takers exist, LATE $\neq$ ATT.

The effect on compliers might be quite different from the effect on always-takers.

- LATE can be very big, compared with ATT, if the effect on always-takers is expected to be small.
- Do you think that this is likely the case for the KIPP example?

ITT

The numerator of LATE is called the Intention to Treat (ITT) effect.

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]. \tag{6}$$

It captures the causal effect on those who are *assigned to* treatment.

With perfect compliance, LATE = ITT.

- How do we know?
- Recall LATE = ATT with perfect compliance. Taken together, LATE = ITT = ATT with perfect compliance.

MDVE

Domestic violence is a serious social issue.

What would be the most effective way for police to deter further domestic assaults?

- Arrest?
- Advise?
- Separate?
- Else?

MDVE (cont.)

The Minneapolis Domestic Violence Experiment (MDVE) is a pathbreaking experiment conducted in the early 1980s.

There were three treatments

- Arrest
- Separate (for 8 hours)
- Advise

Officers acted based on the color of the form on top of a pad when they encountered a situation which meets experimental criteria (= probable cause to believe that a suspect committed misdemeanor assault against a victim in the past 4 hours).

- Report forms are randomly colored.

Further rules

- Cases of life-threatening or severe injury were excluded.
- Both suspect and victim had to be present.

MDVE (cont.)

In practice, officers often deviated from the responses indicated by the color of the form.

- When a suspect attempted to assault an officer, both parties were injured, etc.

Thus, this is an experiment with imperfect compliance!

- Assignment = the color of the form, treatment = officers' actions

We examine the effect on the reoccurrence of a domestic assault within 6 months.

MDVE (cont.)

Let's compute first-stage, ITT, and LATE!

Launch RStudio.

Type

```
mdve <- mmdata::mdve
mdve <- mdve[T_FINAL != 4,]

with(mdve, CrossTable(T_RANDOM, T_FINAL, prop.c=FALSE,
prop.t=FALSE, prop.chisq=FALSE))
```

T_RANDOM is assignment and T_FINAL is police's action.

$1 =$ Arrest, $2 =$ Advise, and $3 =$ Separate in the table.

MDVE (cont.)

Recode Advise and Separate as Coddle.

Type (I skip the code for making coddle variables for brevity)

```
with(mdve, CrossTable(z_coddle, d_coddle, prop.c=FALSE,
prop.t=FALSE, prop.chisq=FALSE))
```

In the table, $0 =$ Arrest and $1 =$ Coddle.

In the following analysis, we consider Arrest as a control group and Coddle as a treatment group.

MDVE (cont.)

Let's compute the first stage.

Recall the definition of the first stage

$$E[D_i|Z_i = 1] - E[D_i|Z_i = 0], \tag{7}$$

Where Z_i is assigned treatment and D_i is delivered treatment.

MDVE (cont.)

Type

```
exp_d_z1 <- with(mdve, length(d_coddle[z_coddle==1 &
d_coddle==1])/length(d_coddle[z_coddle==1]))
exp_d_z0 <- with(mdve, length(d_coddle[z_coddle==0 &
d_coddle==1])/length(d_coddle[z_coddle==0]))

exp_d_z1
exp_d_z0

first_stage <- exp_d_z1 - exp_d_z0
first_stage
```

MDVE (cont.)

Next, let's compute ITT.

Recall the definition of ITT

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0], \tag{8}$$

Where Y_i is outcome (a zero-one dummy, reflecting the reoccurrence of a domestic assault within 6 months) in this example.

Type (I skip the part for making the outcome variable for brevity)

```
with(mdve_use, CrossTable(z_coddle, y, prop.c=FALSE,
prop.t=FALSE, prop.chisq=FALSE))
```

MDVE (cont.)

Type

```
exp_y_z1 <- with(mdve, length(y[z_coddle==1 &
y==1])/length(y[z_coddle == 1]))
exp_y_z0 <- with(mdve, length(y[z_coddle==0 &
y==1])/length(y[z_coddle == 0]))

exp_y_z1
exp_y_z0

itt <- exp_y_z1 - exp_y_z0
itt
```

MDVE (cont.)

Finally, let's compute LATE.

Recall the definition of LATE

$$\lambda = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}. \tag{9}$$

Type

```
late <- itt/first_stage
late
```

Do you think that LATE is larger than ATT?

Two-stage least squares (2SLS)

So far we have considered an experimental setting for which LATE is relatively easily computed.

In less clean settings, we may want to add control variables.

We introduce a method called two-stage least squares (2SLS), which is a generalization of IV (though still IV) because it allows to

- Add controls,
- Use multiple instruments.

Two-stage least squares (2SLS) (cont.)

Consider a model

$$Y_i = \alpha + \beta D_i + \varepsilon_i. \qquad (10)$$

If D_i is not random, the OLS estimate of β is likely biased.

- This is often called a structural regression equation in the IV framework.

Our goal is to get a causal effect β using the IV (2SLS) method.

Two-stage least squares (2SLS) (cont.)

The reduced-form regression equation is written by

$$Y_i = \alpha_1 + \rho Z_i + \varepsilon_{1i}, \tag{11}$$

Where Z_i is assigned treatment, which is assumed to be randomly assigned and affect the outcome Y_i only through D_i.

The reduced-form effect is

$$\rho = E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0], \tag{12}$$

Which is ITT (as we already know)!

Two-stage least squares (2SLS) (cont.)

Next, the first-stage regression equation is written by

$$D_i = a + \phi Z_i + u_i, \tag{13}$$

Where D_i is delivered treatment.

The first-stage effect is

$$\phi = E[D_i|Z_i = 1] - E[D_i|Z_i = 0]. \tag{14}$$

Two-stage least squares (2SLS) (cont.)

Using two equations, we get LATE

$$\lambda = \rho/\phi = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}. \tag{15}$$

Two-stage least squares (2SLS) (cont.)

An alternative expression is

$$
\begin{aligned}
\lambda = \rho/\phi &= \frac{Cov(Y_i, Z_i)/Var(Z_i)}{Cov(D_i, Z_i)/Var(Z_i)} \qquad (16) \\
&= \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)} \qquad (17)
\end{aligned}
$$

The ratio of covariances is called the IV formula. It's sample analogue is the IV estimator.

- Recall that the IV estimator recovers LATE.
- You get a similar expression as (17) for regressions with control variables. (In that case, Z_i is replaced by the residual from a regression of Z_i on X_i.)

The dependent variables in (11) and (13), i.e., Y_i and D_i, are called endogenous variables, while Z_i in both equations is called an exogenous variable (or an instrument).

- Control variables are sometimes called included instruments, while the instrument like Z_i is sometimes called excluded instruments.

Two-stage least squares (2SLS) (cont.)

The 2SLS method estimates LATE in a slightly different way.

First, run the first-stage regression, and get fitted values

$$\hat{D}_i = a + \phi Z_i. \tag{18}$$

Then, the second-stage regression equation is

$$Y_i = \alpha_2 + \lambda_{2SLS}\hat{D}_i + \varepsilon_{2i}. \tag{19}$$

This λ_{2SLS} is equal to λ.

- Prove it.

MDVE (cont.)

Let's compute the first-stage, ITT, and LATE using regressions.

Launch RStudio.

Type

```
reg1 <- lm(d_coddle ~ z_coddle, data=mdve)
stargazer(reg1, type="text")

first_stage
```

What do you find?

MDVE (cont.)

Type

```
reg2 <- lm(y ~ z_coddle, data=mdve)
stargazer(reg2, type="text")

itt
```

What do you find?

MDVE (cont.)

Type

```
reg3 <- ivreg(y ~ d_coddle | z_coddle, data=mdve)
stargazer(reg1, reg2, reg3, type="text")

late
```

What do you find?

Identification

So far we have assumed that Z_i is randomly assigned and affects the outcome Y_i only through D_i.

- This means that Z_i satisfies the independence and exclusion restriction assumptions.

More generally, what kind of assumptions do we need to validate the IV (2SLS) method?

Four assumptions for IV (2SLS) (very important)

1. First stage
   - The first-stage effect exists.
2. Independence
   - The instrument is randomly assigned.
3. Exclusion restriction
   - The instrument affects outcomes only through the treatment variable.
4. Monotonicity
   - No one behaves in the opposite direction if treated.

First stage

The first stage assumption says that the first-stage effect exists.

At minimum, the first stage should be statistically significant.

Conventionally, the first-stage F-statistic should be *at least 10*.

- If the number of the instrument is one, you can roughly compute F-statistic by squaring t-statistic.

If F-statistic is small, it is said that the instrument is *weak*.

- With weak instruments, the IV estimate is biased or does not exist.
- To see this intuitively, see what happens to λ when $\phi = 0$.

Independence

The independence assumption says that the instrument is randomly assigned, so that it is independent of potential outcomes and potential treatment status.

Put differently, this means that the instrument Z_i is independent of ε_{1i} in (11) and u_i in (13).

- If the independence assumption is satisfied, one can interpret the effect in (11) and (13) as a causal effect.

At minimum, the assumption should be satisfied conditional on control variables.

Exclusion restriction

The exclusion restriction says that the instrument affects outcomes only through the treatment variable.

Put differently, the instrument should be uncorrelated with the error term in the structural regression equation

$$Y_i = \alpha + \beta D_i + \varepsilon_i.$$

Exclusion restriction (cont.)

To see why we need the assumption

$$
\frac{E[Y_i|Z_i=1] - E[Y_i|Z_i=0]}{E[D_i|Z_i=1] - E[D_i|Z_i=0]}
$$

$$
= \frac{\beta\{E[D_i|Z_i=1] - E[D_i|Z_i=0]\} - E[\varepsilon_i|Z_i=1] - E[\varepsilon_i|Z_i=0]}{E[D_i|Z_i=1] - E[D_i|Z_i=0]}
$$

$$
= \beta + \frac{E[\varepsilon_i|Z_i=1] - E[\varepsilon_i|Z_i=0]}{E[D_i|Z_i=1] - E[D_i|Z_i=0]} \tag{20}
$$

If the exclusion restriction is satisfied, the second term becomes zero.

Exclusion restriction (cont.)

Alternatively, you can check the same thing but using the following expressions

$$\frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)}$$
$$= \frac{Cov(\alpha + \beta D_i + \varepsilon_i, Z_i)}{Cov(D_i, Z_i)}$$
$$= \beta + \frac{Cov(\varepsilon_i, Z_i)}{Cov(D_i, Z_i)}. \qquad (21)$$

If the exclusion restriction is satisfied, the second term becomes zero.

Exclusion restriction (cont.)

At minimum, the assumption should be satisfied conditional on control variables.

Although the assumption cannot be tested, its validity can be investigated by looking at reduced-form effects.

- If there is no reduced-form effect, it is likely that there are other channels.
- However, the existence of reduced-form effect does not necessarily imply that the assumption is satisfied.
- You always need to argue the validity of the assumption!

Four assumptions for IV (cont.)

There is no way to directly test these assumptions.

At minimum, you should show

- First-stage F-statistic is sufficiently large
- Reduced-form effects exist.

In addition, you need to give a good argument for why the assumptions are likely to be satisfied in your case.

Minimum rules for empiricists: IV

Show the first-stage F-statistic and reduced-form results.

Argue why your instrument(s) is (are) likely to be valid.

- Independence?
- Exclusion restriction?
- Monotonicity?

Then, compare IV estimates with OLS estimates.

- If the OLS estimate and the IV estimate are very similar. → The OLS estimate is likely to capture a causal effect.
- If the OLS estimate and the IV estimate are different. → The OLS estimate is likely to be biased. You need to provide a possible explanation for this.

Returns to education

Let's estimate the causal effect of education on wages using the IV method.

Consider a regression

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \qquad (22)$$

Where Y_i is wages and X_i is years of schooling. The regression may not give you a causal effect because X_i is likely endogenous.

Is there any good instrument for X_i?

Returns to education (cont.)

A famous instrument used by Angrist and Krueger (1991) is the quarter of birth.

In most states in the U.S., children enter kindergarten in the year they turn 5.

Most states also allow students to drop out school only after they turn 16 without finishing the school year.

This implies

- Children born earlier in the year (e.g. January 1st) have to wait until school starts in September, while they are able to drop out before completing the school year when they turn 16.
- Children born later in the year (e.g., December 31st) are not yet 5 when they enter school, and cannot leave school before turning 16.

There is a good reason to believe that children born earlier in the year *are forced to* have lower average schooling.
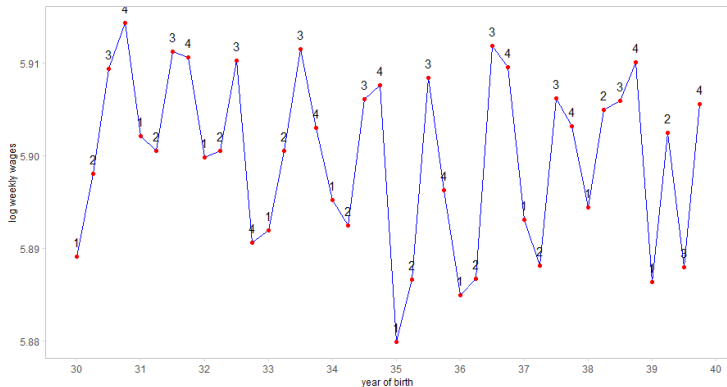
First stage
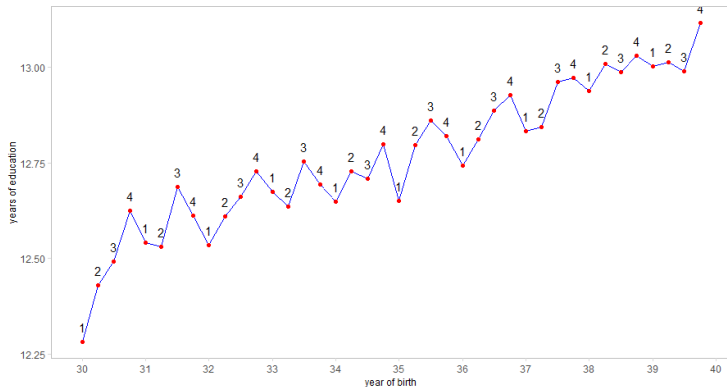
Sample: Individuals who born in 1930-1939 in the 1980 Census.

Reduced form

Sample: Individuals who born in 1930-1939 in the 1980 Census.

IV assumptions

The first-stage assumption is likely to be satisfied.

- We will also test this.

How about other assumptions?

- Independence
- Exclusion restriction
- Monotonicity

Returns to education (cont.)

Individuals born in Q4 are considered to be treated, while the rest are control.

Type

```
qob <- mmdata::qob

qob.f <- factor(qob$qob)
q <- model.matrix(~ qob.f + 0)
qob <- cbind(qob,q)

diff_school <- with(qob, mean(s[qob.f4==1])) - with(qob,
mean(s[qob.f4!=1]))
diff_lnw <- with(qob, mean(lnw[qob.f4==1])) - with(qob,
mean(lnw[qob.f4!=1]))

diff_school
diff_lnw
diff_lnw/diff_school
```

Returns to education (cont.)

Difference in schooling between individuals born in Q4 and the rest is 0.9 years.

Difference in weekly wages between them is 0.7%.

- The effect is interpreted in percent because the dependent variable is logged.

LATE indicates 7% increase in wages.

Returns to education (cont.)

Next, let's compare the results with results from regressions.
Type

```
reg1 <- lm(s ~ qob.f4, data = qob)
reg2 <- lm(lnw ~ qob.f4, data = qob)
reg3 <- ivreg(lnw ~ s | qob.f4, data = qob)
reg4 <- lm(lnw ~ s, data = qob)

stargazer(reg1, reg2, reg3, reg4, type="text")
```

Do you get the same effect?

Is the first-stage assumption likely to be satisfied?

Compare the OLS estimate with the IV estimate. What do you find?

IV solves measurement error

Let's see how an IV solves measurement error.

Consider a model

$$Y_i = \alpha + \beta X_i^* + \varepsilon_i \tag{23}$$

Where X_i^* is measured with an error

$$e_i = X_i - X_i^*, \tag{24}$$

Such that $E(e_i) = 0$.

Thus, what we estimate is

$$Y_i = \alpha + \beta X_i + (\varepsilon_i - \beta e_i). \tag{25}$$

Assume $Cov(X_i^*, e_i) = 0$.

Let's consider an instrument Z_i for X_i^*, which is unrelated to ε_i or e_i.

IV solves measurement error (cont.)

The IV formula for β in (25) is

$$\frac{Cov(Y_i, Z_i)}{Cov(X_i, Z_i)} \tag{26}$$

$$= \frac{Cov(\alpha + \beta X_i^* + \varepsilon_i, Z_i)}{Cov(X_i^* + e_i, Z_i)} \tag{27}$$

$$= \frac{\beta Cov(X_i^*, Z_i) + Cov(\varepsilon_i, Z_i)}{Cov(X_i^*, Z_i) + Cov(e_i, Z_i)} \tag{28}$$

$$= \beta, \tag{29}$$

Where I used $Cov(e_i, Z_i) = 0$ and $Cov(\varepsilon_i, Z_i) = 0$.

Summary

Imperfect compliance

Local average treatment effect (LATE)

Intention to treat (ITT) effect

Two-stage least squares (2SLS)

IV assumptions

IV solves measurement error

Appendix

Let's prove that LATE is the effect on compliers using four assumptions.

Define $D_i(1)$ and $D_i(0)$ be potential treatment status when $Z_i = 1$ and $Z_i = 0$, respectively.

Define $Y_i(d, z)$ be the potential outcome when $D_i = d$ and $Z_i = z$.

- $Y_i(1, Z_i) - Y_i(0, Z_i)$ is the causal effect of D_i on Y_i given Z_i.
- $Y_i(D_i, 1) - Y_i(D_i, 0)$ is the causal effect of Z_i on Y_i given D_i.

Four assumptions are formally written by

- (Independence) $Y_i(D_i(1), 1), Y_i(D_i(0), 0), D_i(1), D_i(0) \perp Z_i$
- (Exclusion restriction) $Y_i(d, 1) = Y_i(d, 0) = Y_i(d)$ for all $d = 0, 1$
- (First-stage) $E[D_i(1) - D_i(0)] \neq 0$
- (Monotonicity) $D_i(1) - D_i(0) \geq 0$ for all i or *vice versa*.

Appendix (cont.)

The numerator of LATE is written by

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

$$= E[Y_i(0) + (Y_i(1) - Y_i(0))D_i|Z_i = 1]$$
$$- E[Y_i(0) + (Y_i(1) - Y_i(0))D_i|Z_i = 0] \tag{30}$$

$$= E[Y_i(0) + (Y_i(1) - Y_i(0))D_i(1)]$$
$$- E[Y_i(0) + (Y_i(1) - Y_i(0))D_i(0)] \tag{31}$$

$$= E[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))] \tag{32}$$

$$= E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]\text{Prob}(D_i(1) > D_i(0)) \tag{33}$$

Where the first equation uses exclusion restriction, the second equation uses independence, and the fourth equation uses monotonicity.

- Without monotonicity, we instead get

$$E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]\text{Prob}(D_i(1) > D_i(0))$$
$$- E[Y_i(1) - Y_i(0)|D_i(1) < D_i(0)]\text{Prob}(D_i(1) < D_i(0)). \tag{34}$$

Appendix (cont.)

Similarly, the denominator of LATE is written by

$$E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$$
$$= E[D_i(1) - D_i(0)] \tag{35}$$
$$= \text{Prob}(D_i(1) > D_i(0)). \tag{36}$$

Which is not zero by the first-stage assumption.

Therefore, LATE is rewritten by

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]. \tag{37}$$

This equation says that LATE is the effect on compliers because $D_i(1) = 1$ and $D_i(0) = 0$ for them.

- What are $D_i(1)$ and $D_i(0)$ for always-takers, never-takers, and defiers?