# Instrumental Variables

Han Zhang

# Outline

# Recommend Readings

- Joshua D. Angrist and Jorn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricists Companion* . Princeton University Press, 2009. (Chapter 4)

# Short overview so far

- If we assume selection on observables/ignorability, matching and regression both identifies causal effect

# Short overview so far

- If we assume selection on observables/ignorability, matching and regression both identifies causal effect
  - But we are making a very strong assumptions: no unobserved confounders!

# Short overview so far

- If we assume selection on observables/ignorability, matching and regression both identifies causal effect
  - But we are making a very strong assumptions: no unobserved confounders!
  - In Swiss naturalization example, it is more likely to be true

# Short overview so far

- If we assume selection on observables/ignorability, matching and regression both identifies causal effect
  - But we are making a very strong assumptions: no unobserved confounders!
  - In Swiss naturalization example, it is more likely to be true
  - But in most real-life cases, ignorability is too strong

# Short overview so far

- If we assume selection on observables/ignorability, matching and regression both identifies causal effect
  - But we are making a very strong assumptions: no unobserved confounders!
  - In Swiss naturalization example, it is more likely to be true
  - But in most real-life cases, ignorability is too strong
- Instead, a better approach acknowledge that there are unobserved confounders

# Short overview so far

- If we assume selection on observables/ignorability, matching and regression both identifies causal effect
  - But we are making a very strong assumptions: no unobserved confounders!
  - In Swiss naturalization example, it is more likely to be true
  - But in most real-life cases, ignorability is too strong
- Instead, a better approach acknowledge that there are unobserved confounders
  - And try to see eliminate unobserved confounders by adding less strong assumptions

# Short overview so far

- If we assume selection on observables/ignorability, matching and regression both identifies causal effect
  - But we are making a very strong assumptions: no unobserved confounders!
  - In Swiss naturalization example, it is more likely to be true
  - But in most real-life cases, ignorability is too strong
- Instead, a better approach acknowledge that there are unobserved confounders
  - And try to see eliminate unobserved confounders by adding less strong assumptions
  - Natural experiments

# Short overview so far

- If we assume selection on observables/ignorability, matching and regression both identifies causal effect
  - But we are making a very strong assumptions: no unobserved confounders!
  - In Swiss naturalization example, it is more likely to be true
  - But in most real-life cases, ignorability is too strong
- Instead, a better approach acknowledge that there are unobserved confounders
  - And try to see eliminate unobserved confounders by adding less strong assumptions
  - Natural experiments
  - Fixed effects/DiD: exploit natural groups; use within-group difference to cancel out group-level unobserved confounders

# Short overview so far

- If we assume selection on observables/ignorability, matching and regression both identifies causal effect
  - But we are making a very strong assumptions: no unobserved confounders!
  - In Swiss naturalization example, it is more likely to be true
  - But in most real-life cases, ignorability is too strong
- Instead, a better approach acknowledge that there are unobserved confounders
  - And try to see eliminate unobserved confounders by adding less strong assumptions
  - Natural experiments
  - Fixed effects/DiD: exploit natural groups; use within-group difference to cancel out group-level unobserved confounders
  - General principle: approximates experiment ideal

# Instrumental Variable

- $Y = \alpha + \rho D + \epsilon$;

# Instrumental Variable

- $Y = \alpha + \rho D + \epsilon$;
- $\rho$ identifies ATT/ATE if selection on observables are true

# Instrumental Variable

- $Y = \alpha + \rho D + \epsilon$;
- $\rho$ identifies ATT/ATE if selection on observables are true
- When there are unobserved confoundings, $\rho$ will be a biased estimate of $ATT$

# Instrumental Variable

- $Y = \alpha + \rho D + \epsilon$;
- $\rho$ identifies ATT/ATE if selection on observables are true
- When there are unobserved confoundings, $\rho$ will be a biased estimate of $ATT$
- In the linear regression framework, the presence of unobserved confounding means that

# Instrumental Variable

- $Y = \alpha + \rho D + \epsilon$;
- $\rho$ identifies ATT/ATE if selection on observables are true
- When there are unobserved confoundings, $\rho$ will be a biased estimate of $ATT$
- In the linear regression framework, the presence of unobserved confounding means that
    - Zero conditional mean assumption does not hold; $E(\epsilon|X) \neq 0$

# Instrumental Variable

- $Y = \alpha + \rho D + \epsilon$;
- $\rho$ identifies ATT/ATE if selection on observables are true
- When there are unobserved confoundings, $\rho$ will be a biased estimate of $ATT$
- In the linear regression framework, the presence of unobserved confounding means that
  - Zero conditional mean assumption does not hold; $E(\epsilon|X) \neq 0$
  - Or, $X$ is an endogenous regressor

# Instrumental Variable

- $Y = \alpha + \rho D + \epsilon$;
- $\rho$ identifies ATT/ATE if selection on observables are true
- When there are unobserved confoundings, $\rho$ will be a biased estimate of $ATT$
- In the linear regression framework, the presence of unobserved confounding means that
  - Zero conditional mean assumption does not hold; $E(\epsilon|X) \neq 0$
  - Or, $X$ is an endogenous regressor
- Instrumental variables (IV) recognizes that unobserved confounders indeed exist

# Instrumental Variable

- $Y = \alpha + \rho D + \epsilon$;
- $\rho$ identifies ATT/ATE if selection on observables are true
- When there are unobserved confoundings, $\rho$ will be a biased estimate of $ATT$
- In the linear regression framework, the presence of unobserved confounding means that
  - Zero conditional mean assumption does not hold; $E(\epsilon|X) \neq 0$
  - Or, $X$ is an endogenous regressor
- Instrumental variables (IV) recognizes that unobserved confounders indeed exist
- IV exploits exogenous variation that drive the treatment but do not otherwise affect the outcome.

# IV setup

- IV Setup (assuming a linear regression!)

# IV setup

- IV Setup (assuming a linear regression!)
  - Second stage: $Y = \alpha + \rho D + \epsilon$

# IV setup

- IV Setup (assuming a linear regression!)
  - Second stage: $Y = \alpha + \rho D + \epsilon$
    - This is what you would normally run without IV

# IV setup

- IV Setup (assuming a linear regression!)
  - Second stage: $Y = \alpha + \rho D + \epsilon$
    - This is what you would normally run without IV
  - First stage: $X = \gamma + \beta Z + \eta$

# IV setup

- IV Setup (assuming a linear regression!)
  - Second stage: $Y = \alpha + \rho D + \epsilon$
    - This is what you would normally run without IV
  - First stage: $X = \gamma + \beta Z + \eta$
- $Z$ is an IV

# IV setup

- IV Setup (assuming a linear regression!)
  - Second stage: $Y = \alpha + \rho D + \epsilon$
    - This is what you would normally run without IV
  - First stage: $X = \gamma + \beta Z + \eta$
- $Z$ is an IV
  - It drives treatment assignment $D$

# IV setup

- IV Setup (assuming a linear regression!)
  - Second stage: $Y = \alpha + \rho D + \epsilon$
    - This is what you would normally run without IV
  - First stage: $X = \gamma + \beta Z + \eta$
- $Z$ is an IV
  - It drives treatment assignment $D$
  - But it does not have a direct impact on outcome $Y$

# IV setup

- IV Setup (assuming a linear regression!)
  - Second stage: $Y = \alpha + \rho D + \epsilon$
    - This is what you would normally run without IV
  - First stage: $X = \gamma + \beta Z + \eta$
- $Z$ is an IV
  - It drives treatment assignment $D$
  - But it does not have a direct impact on outcome $Y$
- $Z$ is kind of like a researcher who manipulates $D$ but not outcome

# IV setup

- IV Setup (assuming a linear regression!)
  - Second stage: $Y = \alpha + \rho D + \epsilon$
    - This is what you would normally run without IV
  - First stage: $X = \gamma + \beta Z + \eta$
- $Z$ is an IV
  - It drives treatment assignment $D$
  - But it does not have a direct impact on outcome $Y$
- $Z$ is kind of like a researcher who manipulates $D$ but not outcome
  - Later you will become more clear what this means

# IV example

- Angrist and Krueger, 1991, QJE:

# IV example

- Angrist and Krueger, 1991, QJE:
- $Y$: future earnings

# IV example

- Angrist and Krueger, 1991, QJE:
- $Y$: future earnings
- $D$: years of schooling

# IV example

- Angrist and Krueger, 1991, QJE:
- $Y$: future earnings
- $D$: years of schooling
- $Z$: season of birth.

# IV example

- Angrist and Krueger, 1991, QJE:
- $Y$: future earnings
- $D$: years of schooling
- $Z$: season of birth.
  - Compulsory schooling laws require children to enter school in the calendar in which they turn 6

# IV example

- Angrist and Krueger, 1991, QJE:
- $Y$: future earnings
- $D$: years of schooling
- $Z$: season of birth.
    - Compulsory schooling laws require children to enter school in the calendar in which they turn 6
        - $Z = 0$: born in 1st quarter; enter school in Sep around 6.5

# IV example

- Angrist and Krueger, 1991, QJE:
- $Y$: future earnings
- $D$: years of schooling
- $Z$: season of birth.
    - Compulsory schooling laws require children to enter school in the calendar in which they turn 6
        - $Z = 0$: born in 1st quarter; enter school in Sep around 6.5
        - $Z = 1$: born in 4th quarter; enter school in Sep when they were less than 6

# IV example

- Angrist and Krueger, 1991, QJE:
- $Y$: future earnings
- $D$: years of schooling
- $Z$: season of birth.
    - Compulsory schooling laws require children to enter school in the calendar in which they turn 6
        - $Z = 0$: born in 1st quarter; enter school in Sep around 6.5
        - $Z = 1$: born in 4th quarter; enter school in Sep when they were less than 6
    - US compulsory schooling laws are in terms of age (16), not number of years of schooling completed. You can drop out on your 16th birthday (even if in the middle of the school year).

# IV example

- Angrist and Krueger, 1991, QJE:
- $Y$: future earnings
- $D$: years of schooling
- $Z$: season of birth.
  - Compulsory schooling laws require children to enter school in the calendar in which they turn 6
    - $Z = 0$: born in 1st quarter; enter school in Sep around 6.5
    - $Z = 1$: born in 4th quarter; enter school in Sep when they were less than 6
  - US compulsory schooling laws are in terms of age (16), not number of years of schooling completed. You can drop out on your 16th birthday (even if in the middle of the school year).
  - So those born in 4th quarter are compelled to take more educations than those born in 1st quarter

# IV assumptions

- First stage: $D = \gamma + \beta Z + \eta$

# IV assumptions

- First stage: $D = \gamma + \beta Z + \eta$
- Second stage: $Y = \alpha + \rho D + \epsilon$

# IV assumptions

- First stage: $D = \gamma + \beta Z + \eta$
- Second stage: $Y = \alpha + \rho D + \epsilon$
- IV Assumptions:

# IV assumptions

- First stage: $D = \gamma + \beta Z + \eta$
- Second stage: $Y = \alpha + \rho D + \epsilon$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor of $D$, or alternatively saying, in the first-stage equation, there are no unobserved confounders (non-testable)

# IV assumptions

- First stage: $D = \gamma + \beta Z + \eta$
- Second stage: $Y = \alpha + \rho D + \epsilon$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor of $D$, or alternatively saying, in the first-stage equation, there are no unobserved confounders (non-testable)
     - $E(\eta|Z) = 0$

# IV assumptions

- First stage: $D = \gamma + \beta Z + \eta$
- Second stage: $Y = \alpha + \rho D + \epsilon$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor of $D$, or alternatively saying, in the first-stage equation, there are no unobserved confounders (non-testable)
     - $E(\eta|Z) = 0$
     - $E(\epsilon|Z) = 0$

# IV assumptions

- First stage: $D = \gamma + \beta Z + \eta$
- Second stage: $Y = \alpha + \rho D + \epsilon$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor of $D$, or alternatively saying, in the first-stage equation, there are no unobserved confounders (non-testable)
     - $E(\eta|Z) = 0$
     - $E(\epsilon|Z) = 0$
  2. Exclusion restriction: (non-testable)

# IV assumptions

- First stage: $D = \gamma + \beta Z + \eta$
- Second stage: $Y = \alpha + \rho D + \epsilon$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor of $D$, or alternatively saying, in the first-stage equation, there are no unobserved confounders (non-testable)
     - $E(\eta|Z) = 0$
     - $E(\epsilon|Z) = 0$
  2. Exclusion restriction: (non-testable)
     - $Z$ does not has a direct effect on $Y$

# IV assumptions

- First stage: $D = \gamma + \beta Z + \eta$
- Second stage: $Y = \alpha + \rho D + \epsilon$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor of $D$, or alternatively saying, in the first-stage equation, there are no unobserved confounders (non-testable)
     - $E(\eta|Z) = 0$
     - $E(\epsilon|Z) = 0$
  2. Exclusion restriction: (non-testable)
     - $Z$ does not has a direct effect on $Y$
     - That is, in the true data generating process, $Z$ does not appear in the second stage model

# IV assumptions

- First stage: $D = \gamma + \beta Z + \eta$
- Second stage: $Y = \alpha + \rho D + \epsilon$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor of $D$, or alternatively saying, in the first-stage equation, there are no unobserved confounders (non-testable)
     - $E(\eta|Z) = 0$
     - $E(\epsilon|Z) = 0$
  2. Exclusion restriction: (non-testable)
     - $Z$ does not has a direct effect on $Y$
     - That is, in the true data generating process, $Z$ does not appear in the second stage model
  3. First stage relevance: (testable)

# IV assumptions

- First stage: $D = \gamma + \beta Z + \eta$
- Second stage: $Y = \alpha + \rho D + \epsilon$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor of $D$, or alternatively saying, in the first-stage equation, there are no unobserved confounders (non-testable)
     - $E(\eta|Z) = 0$
     - $E(\epsilon|Z) = 0$
  2. Exclusion restriction: (non-testable)
     - $Z$ does not has a direct effect on $Y$
     - That is, in the true data generating process, $Z$ does not appear in the second stage model
  3. First stage relevance: (testable)
     - $Z$'s effect on $D$ should not be 0 ($\beta \neq 0$)

# IV assumptions

- First stage: $D = \gamma + \beta Z + \eta$
- Second stage: $Y = \alpha + \rho D + \epsilon$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor of $D$, or alternatively saying, in the first-stage equation, there are no unobserved confounders (non-testable)
     - $E(\eta|Z) = 0$
     - $E(\epsilon|Z) = 0$
  2. Exclusion restriction: (non-testable)
     - $Z$ does not has a direct effect on $Y$
     - That is, in the true data generating process, $Z$ does not appear in the second stage model
  3. First stage relevance: (testable)
     - $Z$'s effect on $D$ should not be 0 ($\beta \neq 0$)
  4. Constant effect $\rho$

# IV assumptions

- First stage: $D = \gamma + \beta Z + \eta$
- Second stage: $Y = \alpha + \rho D + \epsilon$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor of $D$, or alternatively saying, in the first-stage equation, there are no unobserved confounders (non-testable)
     - $E(\eta|Z) = 0$
     - $E(\epsilon|Z) = 0$
  2. Exclusion restriction: (non-testable)
     - $Z$ does not has a direct effect on $Y$
     - That is, in the true data generating process, $Z$ does not appear in the second stage model
  3. First stage relevance: (testable)
     - $Z$'s effect on $D$ should not be 0 ($\beta \neq 0$)
  4. Constant effect $\rho$
     - It will be more clear what this assumption means later

# Are the assumptions met for Angrist and Krueger (1991)?

1. Exogeneity: $Z$ is as-if randomly assigned.
   - non-testable

# Are the assumptions met for Angrist and Krueger (1991)?

1. Exogeneity: $Z$ is as-if randomly assigned.
   - non-testable
   - Reasonable? Will parents selectively give birth in certain months because they think children will be more beneficial?

# Durbin-Wu-Hausman test

- Durbin-Wu-Hausman test (or Hausman's specification test)

# Durbin-Wu-Hausman test

- Durbin-Wu-Hausman test (or Hausman's specification test)
- Traditionally this was believed to be a way to "test whether a regressor is exogenous or endogenous"

# Durbin-Wu-Hausman test

- Durbin-Wu-Hausman test (or Hausman's specification test)
- Traditionally this was believed to be a way to "test whether a regressor is exogenous or endogenous"
  - You may still see some of this saying in the old textbook/articles

# Durbin-Wu-Hausman test

- Durbin-Wu-Hausman test (or Hausman's specification test)
- Traditionally this was believed to be a way to "test whether a regressor is exogenous or endogenous"
  - You may still see some of this saying in the old textbook/articles
- Basically, DWH test compares OLS estimate (second-stage) with 2SLS estimate

# Durbin-Wu-Hausman test

- Durbin-Wu-Hausman test (or Hausman's specification test)
- Traditionally this was believed to be a way to "test whether a regressor is exogenous or endogenous"
  - You may still see some of this saying in the old textbook/articles
- Basically, DWH test compares OLS estimate (second-stage) with 2SLS estimate
  - The Null is that the two are the same

# Durbin-Wu-Hausman test

- Durbin-Wu-Hausman test (or Hausman's specification test)
- Traditionally this was believed to be a way to "test whether a regressor is exogenous or endogenous"
  - You may still see some of this saying in the old textbook/articles
- Basically, DWH test compares OLS estimate (second-stage) with 2SLS estimate
  - The Null is that the two are the same
  - The Alternative is that the two are different

# Durbin-Wu-Hausman test

- Durbin-Wu-Hausman test (or Hausman's specification test)
- Traditionally this was believed to be a way to "test whether a regressor is exogenous or endogenous"
  - You may still see some of this saying in the old textbook/articles
- Basically, DWH test compares OLS estimate (second-stage) with 2SLS estimate
  - The Null is that the two are the same
  - The Alternative is that the two are different
- If you do not pass DWH test, your IV is bad and probably should not be used.

# Durbin-Wu-Hausman test

- Durbin-Wu-Hausman test (or Hausman's specification test)
- Traditionally this was believed to be a way to "test whether a regressor is exogenous or endogenous"
    - You may still see some of this saying in the old textbook/articles
- Basically, DWH test compares OLS estimate (second-stage) with 2SLS estimate
    - The Null is that the two are the same
    - The Alternative is that the two are different
- If you do not pass DWH test, your IV is bad and probably should not be used.
- But if you pass DWH test (p-value $< 0.05$), it does not mean that you have a valid IV

# Durbin-Wu-Hausman test

- Durbin-Wu-Hausman test (or Hausman's specification test)
- Traditionally this was believed to be a way to "test whether a regressor is exogenous or endogenous"
  - You may still see some of this saying in the old textbook/articles
- Basically, DWH test compares OLS estimate (second-stage) with 2SLS estimate
  - The Null is that the two are the same
  - The Alternative is that the two are different
- If you do not pass DWH test, your IV is bad and probably should not be used.
- But if you pass DWH test (p-value $< 0.05$), it does not mean that you have a valid IV
- You still need to have a story on why IV is as-if randomly assigned

# Are the assumptions met for Angrist and Krueger (1991)?

1. Exclusion restriction: $Z$ does not has a direct effect on $D$
   - non-testable

# Are the assumptions met for Angrist and Krueger (1991)?

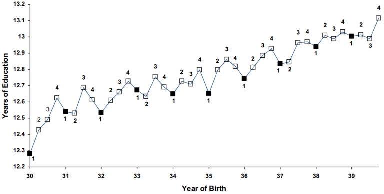
1. Exclusion restriction: $Z$ does not has a direct effect on $D$
   - non-testable
   - Could birthdays affect earnings directly, beyond their effect on years of schooling?

# Are the assumptions met for Angrist and Krueger (1991)?

1. Exclusion restriction: $Z$ does not has a direct effect on $D$
    - non-testable
    - Could birthdays affect earnings directly, beyond their effect on years of schooling?
        - e.g., those who were born in 1st quarter are older so that they will be more mentally developed compared with those born in 4th quarter

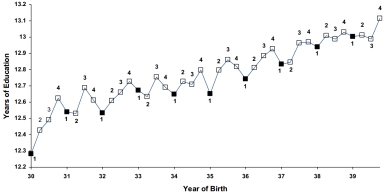
# Are the assumptions met for Angrist and Krueger (1991)?

1. First stage relevance: instrument $Z$ should has an impact on $D$
   - Plot $D$ against $Z$; visually inspect correlation (MHE, Fig. 4.1.1)



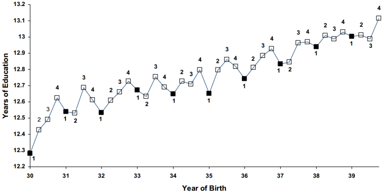
# Are the assumptions met for Angrist and Krueger (1991)?

1. First stage relevance: instrument $Z$ should has an impact on $D$
   - Plot $D$ against $Z$; visually inspect correlation (MHE, Fig. 4.1.1)
     - On average, those who were born in the 4th quarter has more schooling than those born in the 1st quarter



# Are the assumptions met for Angrist and Krueger (1991)?

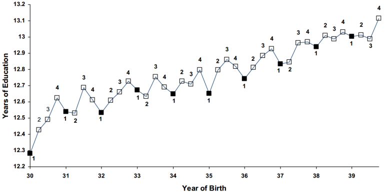
1. First stage relevance: instrument $Z$ should has an impact on $D$
    - Plot $D$ against $Z$; visually inspect correlation (MHE, Fig. 4.1.1)
        - On average, those who were born in the 4th quarter has more schooling than those born in the 1st quarter




- Or, regress $D$ on $Z$;

# Are the assumptions met for Angrist and Krueger (1991)?

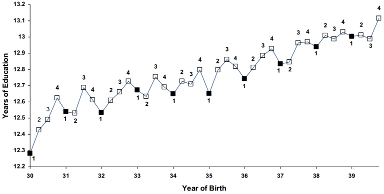
1. First stage relevance: instrument $Z$ should has an impact on $D$
    - Plot $D$ against $Z$; visually inspect correlation (MHE, Fig. 4.1.1)
        - On average, those who were born in the 4th quarter has more schooling than those born in the 1st quarter




- Or, regress $D$ on $Z$;
    - The traditional rule-of-thumb is that first-stage regression's F-statistics $> 10$

# Are the assumptions met for Angrist and Krueger (1991)?

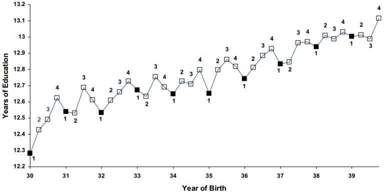
1. First stage relevance: instrument $Z$ should has an impact on $D$
    - Plot $D$ against $Z$; visually inspect correlation (MHE, Fig. 4.1.1)
        - On average, those who were born in the 4th quarter has more schooling than those born in the 1st quarter




- Or, regress $D$ on $Z$;
    - The traditional rule-of-thumb is that first-stage regression's F-statistics $> 10$
    - For Angrist and Krueger (1991), F-statistics $= 24.1 > 10$

# Are the assumptions met for Angrist and Krueger (1991)?

1. First stage relevance: instrument $Z$ should has an impact on $D$
   - Plot $D$ against $Z$; visually inspect correlation (MHE, Fig. 4.1.1)
     - On average, those who were born in the 4th quarter has more schooling than those born in the 1st quarter




- Or, regress $D$ on $Z$;
  - The traditional rule-of-thumb is that first-stage regression's F-statistics $> 10$
  - For Angrist and Krueger (1991), F-statistics $= 24.1 > 10$
  - Some recent work suggests that F-statistics $> 100$

# Math of IV

- We are interested in $\rho$, but directly regressing $Y$ on $D$ yields a biased estimate of $\rho$ (because $E(\epsilon|D) \neq 0$)

$$\begin{aligned} Y &= \alpha + \rho D + \epsilon \\ &= \alpha + \rho\left(\gamma + \beta Z + \eta\right) + \epsilon \\ &= \left(\alpha + \rho\gamma\right) + \rho\beta Z + \rho\eta + \epsilon \end{aligned} \tag{1}$$

# Math of IV

- We are interested in $\rho$, but directly regressing $Y$ on $D$ yields a biased estimate of $\rho$ (because $E(\epsilon|D) \neq 0$)
- Substitute $D$ by the first-stage equation

$$\begin{aligned} Y &= \alpha + \rho D + \epsilon \\ &= \alpha + \rho\left(\gamma + \beta Z + \eta\right) + \epsilon \\ &= (\alpha + \rho\gamma) + \rho\beta Z + \rho\eta + \epsilon \end{aligned} \tag{1}$$

# Math of IV

- We are interested in $\rho$, but directly regressing $Y$ on $D$ yields a biased estimate of $\rho$ (because $E(\epsilon|D) \neq 0$)
- Substitute $D$ by the first-stage equation

$$\begin{aligned} Y &= \alpha + \rho D + \epsilon \\ &= \alpha + \rho\left(\gamma + \beta Z + \eta\right) + \epsilon \\ &= (\alpha + \rho\gamma) + \rho\beta Z + \rho\eta + \epsilon \end{aligned} \tag{1}$$

- Now the second-stage becomes regressing $Y$ on instrument $Z$; sometimes this regression is called reduced form regression

# Math of IV

- We are interested in $\rho$, but directly regressing $Y$ on $D$ yields a biased estimate of $\rho$ (because $E(\epsilon|D) \neq 0$)
- Substitute $D$ by the first-stage equation

$$\begin{aligned} Y &= \alpha + \rho D + \epsilon \\ &= \alpha + \rho(\gamma + \beta Z + \eta) + \epsilon \\ &= (\alpha + \rho\gamma) + \rho\beta Z + \rho\eta + \epsilon \end{aligned} \tag{1}$$

- Now the second-stage becomes regressing $Y$ on instrument $Z$; sometimes this regression is called reduced form regression
- In reduced-form regression, the errors are mean independent of $Z$: $E(\rho\eta + \epsilon|Z) = 0$)

# Math of IV

- We are interested in $\rho$, but directly regressing $Y$ on $D$ yields a biased estimate of $\rho$ (because $E(\epsilon|D) \neq 0$)
- Substitute $D$ by the first-stage equation

$$\begin{aligned} Y &= \alpha + \rho D + \epsilon \\ &= \alpha + \rho(\gamma + \beta Z + \eta) + \epsilon \\ &= (\alpha + \rho\gamma) + \rho\beta Z + \rho\eta + \epsilon \end{aligned} \tag{1}$$

- Now the second-stage becomes regressing $Y$ on instrument $Z$; sometimes this regression is called reduced form regression
- In reduced-form regression, the errors are mean independent of $Z$: $E(\rho\eta + \epsilon|Z) = 0)$
  - Because $E(\eta|Z) = 0)$

# Math of IV

- We are interested in $\rho$, but directly regressing $Y$ on $D$ yields a biased estimate of $\rho$ (because $E(\epsilon|D) \neq 0$)
- Substitute $D$ by the first-stage equation

$$\begin{aligned} Y &= \alpha + \rho D + \epsilon \\ &= \alpha + \rho\left(\gamma + \beta Z + \eta\right) + \epsilon \\ &= \left(\alpha + \rho\gamma\right) + \rho\beta Z + \rho\eta + \epsilon \end{aligned} \tag{1}$$

- Now the second-stage becomes regressing $Y$ on instrument $Z$; sometimes this regression is called reduced form regression
- In reduced-form regression, the errors are mean independent of $Z$: $E(\rho\eta + \epsilon|Z) = 0$)
    - Because $E(\eta|Z) = 0$)
    - And $E(\epsilon|Z) = 0$

# Math of IV

- We are interested in $\rho$, but directly regressing $Y$ on $D$ yields a biased estimate of $\rho$ (because $E(\epsilon|D) \neq 0$)
- Substitute $D$ by the first-stage equation

$$\begin{aligned} Y &= \alpha + \rho D + \epsilon \\ &= \alpha + \rho\left(\gamma + \beta Z + \eta\right) + \epsilon \\ &= \left(\alpha + \rho\gamma\right) + \rho\beta Z + \rho\eta + \epsilon \end{aligned} \tag{1}$$

- Now the second-stage becomes regressing $Y$ on instrument $Z$; sometimes this regression is called reduced form regression
- In reduced-form regression, the errors are mean independent of $Z$: $E(\rho\eta + \epsilon|Z) = 0)$
  - Because $E(\eta|Z) = 0)$
  - And $E(\epsilon|Z) = 0$
- Hence regressing $Y$ on instrument $Z$ identifies $\rho\beta$

# Math of IV

$$Y = (\alpha + \rho\gamma) + \rho\beta Z + \rho\eta + \epsilon$$

- Regress $Y$ on $Z$, $\widehat{\rho\beta}$ is a consistent estimate of $\rho\beta$

# Math of IV

$$Y = (\alpha + \rho\gamma) + \rho\beta Z + \rho\eta + \epsilon$$

- Regress $Y$ on $Z$, $\widehat{\rho\beta}$ is a consistent estimate of $\rho\beta$
- Regression $D$ on $Z$, $\hat{\beta}$ is a consistent estimate of $\beta$

# Math of IV

$$Y = (\alpha + \rho\gamma) + \rho\beta Z + \rho\eta + \epsilon$$

- Regress $Y$ on $Z$, $\widehat{\rho\beta}$ is a consistent estimate of $\rho\beta$
- Regression $D$ on $Z$, $\hat{\beta}$ is a consistent estimate of $\beta$
- $\frac{\widehat{\rho\beta}}{\hat{\beta}}$ yields a consistent estimate of $\rho$: indirect least square estimator (Wright, 1928)

# Math of IV

$$Y = (\alpha + \rho\gamma) + \rho\beta Z + \rho\eta + \epsilon$$

- Regress $Y$ on $Z$, $\widehat{\rho\beta}$ is a consistent estimate of $\rho\beta$
- Regression $D$ on $Z$, $\hat{\beta}$ is a consistent estimate of $\beta$
- $\frac{\widehat{\rho\beta}}{\hat{\beta}}$ yields a consistent estimate of $\rho$: indirect least square estimator (Wright, 1928)
- Because we are interested in the effect of $D$ on $Y$, but we run two other regression $Y$ on $Z$, and $D$ on $Z$, and then take their ratios

# Indirect Least Square and multiple instruments

- If we have two instruments $Z_1, Z_2$,

# Indirect Least Square and multiple instruments

- If we have two instruments $Z_1, Z_2$,
- Using $Z_1$ or $Z_2$ results in a different estimate of $\beta$, because there are two ways to write the indirect least square

# Indirect Least Square and multiple instruments

- If we have two instruments $Z_1, Z_2$,
- Using $Z_1$ or $Z_2$ results in a different estimate of $\beta$, because there are two ways to write the indirect least square
- Treatment effect depends on your instrument

# Indirect Least Square and multiple instruments

- If we have two instruments $Z_1, Z_2$,
- Using $Z_1$ or $Z_2$ results in a different estimate of $\beta$, because there are two ways to write the indirect least square
- Treatment effect depends on your instrument
    - More on this later

# Indirect Least Square and multiple instruments

- If we have two instruments $Z_1, Z_2$,
- Using $Z_1$ or $Z_2$ results in a different estimate of $\beta$, because there are two ways to write the indirect least square
- Treatment effect depends on your instrument
  - More on this later
- Which treatment effect do we use? Not so easy under the indirect least square

# Wald estimator (Wald, 1940)

- In the case of binary instrument, assuming the same linear model:

$$
\begin{aligned}
\hat{\rho} &= \frac{\hat{E}(Y|Z=1) - \hat{E}(Y|Z=0)}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)}, \text{(Wald Estimator)} && (2) \\
&= \frac{\widehat{\rho\beta}}{\hat{\beta}} \text{(Equals to Indirect Least Square)} && (3)
\end{aligned}
$$

# Wald estimator (Wald, 1940)

- In the case of binary instrument, assuming the same linear model:
- $E(D|Z=1)=\gamma+\beta+\eta$ and $E(D|Z=0)=\gamma+\eta$

$$\hat{\rho}=\frac{\hat{E}(Y|Z=1)-\hat{E}(Y|Z=0)}{\hat{E}(D|Z=1)-\hat{E}(D|Z=0)}, \text{(Wald Estimator)} \tag{2}$$

$$=\frac{\widehat{\rho\beta}}{\hat{\beta}} \text{(Equals to Indirect Least Square)} \tag{3}$$

## Wald estimator (Wald, 1940)

- In the case of binary instrument, assuming the same linear model:
- $E(D|Z=1)=\gamma+\beta+\eta$ and $E(D|Z=0)=\gamma+\eta$
- $\beta=E(D|Z=1)-E(D|Z=0)$

$$
\begin{aligned}
\hat{\rho} &= \frac{\hat{E}(Y|Z=1)-\hat{E}(Y|Z=0)}{\hat{E}(D|Z=1)-\hat{E}(D|Z=0)}, \text{(Wald Estimator)} \quad (2) \\
&= \frac{\widehat{\rho\beta}}{\hat{\beta}} \text{(Equals to Indirect Least Square)} \quad (3)
\end{aligned}
$$

# Wald estimator (Wald, 1940)

- In the case of binary instrument, assuming the same linear model:
- $E(D|Z=1) = \gamma + \beta + \eta$ and $E(D|Z=0) = \gamma + \eta$
- $\beta = E(D|Z=1) - E(D|Z=0)$
- Similarly, we can get $\rho\beta = E(Y|Z=1) - E(Y|Z=0)$

$$\hat{\rho} = \frac{\hat{E}(Y|Z=1) - \hat{E}(Y|Z=0)}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)}, \text{(Wald Estimator)} \tag{2}$$

$$= \frac{\widehat{\rho\beta}}{\hat{\beta}} \text{(Equals to Indirect Least Square)} \tag{3}$$

# Wald estimator (Wald, 1940)

- In the case of binary instrument, assuming the same linear model:
- $E(D|Z=1) = \gamma + \beta + \eta$ and $E(D|Z=0) = \gamma + \eta$
- $\beta = E(D|Z=1) - E(D|Z=0)$
- Similarly, we can get $\rho\beta = E(Y|Z=1) - E(Y|Z=0)$
- Wald Estimator

$$\hat{\rho} = \frac{\hat{E}(Y|Z=1) - \hat{E}(Y|Z=0)}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)}, \text{(Wald Estimator)} \tag{2}$$

$$= \frac{\widehat{\rho\beta}}{\hat{\beta}} \text{(Equals to Indirect Least Square)} \tag{3}$$

# Wald estimator (Wald, 1940)

- In the case of binary instrument, assuming the same linear model:
- $E(D|Z=1) = \gamma + \beta + \eta$ and $E(D|Z=0) = \gamma + \eta$
- $\beta = E(D|Z=1) - E(D|Z=0)$
- Similarly, we can get $\rho\beta = E(Y|Z=1) - E(Y|Z=0)$
- Wald Estimator

$$\hat{\rho} = \frac{\hat{E}(Y|Z=1) - \hat{E}(Y|Z=0)}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)}, \text{(Wald Estimator)} \tag{2}$$

$$= \frac{\widehat{\rho\beta}}{\hat{\beta}} \text{(Equals to Indirect Least Square)} \tag{3}$$

- Note: Wald estimator is a more general non-parametric estimator; it equals to Indirect Least Square in binary case only when the linear model assumption is true

# 2SLS estimator

- To solve the multiple instrument problem, we slightly change how we combine the terms

$$\begin{aligned} Y &= \alpha + \rho D + \epsilon \\ &= \alpha + \rho\left(\gamma + \beta Z + \eta\right) + \epsilon \\ &= \alpha + \rho\left(\gamma + \beta Z\right) + \left(\eta + \rho\epsilon\right) \end{aligned}$$

# 2SLS estimator

- To solve the multiple instrument problem, we slightly change how we combine the terms

$$
\begin{aligned}
Y &= \alpha + \rho D + \epsilon \\
&= \alpha + \rho\left(\gamma + \beta Z + \eta\right) + \epsilon \\
&= \alpha + \rho\left(\gamma + \beta Z\right) + \left(\eta + \rho\epsilon\right)
\end{aligned}
$$

- Errors $\eta + \rho\epsilon$ are mean independent of $\gamma + \beta Z \implies$ regressing $Y$ on $\gamma + \beta Z$ gives consistent estimate of $\rho$

# 2SLS estimator

- To solve the multiple instrument problem, we slightly change how we combine the terms

$$\begin{aligned} Y &= \alpha + \rho D + \epsilon \\ &= \alpha + \rho(\gamma + \beta Z + \eta) + \epsilon \\ &= \alpha + \rho(\gamma + \beta Z) + (\eta + \rho\epsilon) \end{aligned}$$

- Errors $\eta + \rho\epsilon$ are mean independent of $\gamma + \beta Z \implies$ regressing $Y$ on $\gamma + \beta Z$ gives consistent estimate of $\rho$
- $\gamma + \beta Z$ can be approximated by the predicted values of the first stage, $\hat{D} = \hat{\gamma} + \hat{\beta} Z$

# 2SLS estimator

- To solve the multiple instrument problem, we slightly change how we combine the terms

$$\begin{aligned} Y &= \alpha + \rho D + \epsilon \\ &= \alpha + \rho\left(\gamma + \beta Z + \eta\right) + \epsilon \\ &= \alpha + \rho\left(\gamma + \beta Z\right) + \left(\eta + \rho\epsilon\right) \end{aligned}$$

- Errors $\eta + \rho\epsilon$ are mean independent of $\gamma + \beta Z \implies$ regressing $Y$ on $\gamma + \beta Z$ gives consistent estimate of $\rho$
- $\gamma + \beta Z$ can be approximated by the predicted values of the first stage, $\hat{D} = \hat{\gamma} + \hat{\beta} Z$
- Hence, 2SLS estimator (Two-stage least square):

# 2SLS estimator

- To solve the multiple instrument problem, we slightly change how we combine the terms

$$\begin{aligned} Y &= \alpha + \rho D + \epsilon \\ &= \alpha + \rho(\gamma + \beta Z + \eta) + \epsilon \\ &= \alpha + \rho(\gamma + \beta Z) + (\eta + \rho\epsilon) \end{aligned}$$

- Errors $\eta + \rho\epsilon$ are mean independent of $\gamma + \beta Z \implies$ regressing $Y$ on $\gamma + \beta Z$ gives consistent estimate of $\rho$
- $\gamma + \beta Z$ can be approximated by the predicted values of the first stage, $\hat{D} = \hat{\gamma} + \hat{\beta} Z$
- Hence, 2SLS estimator (Two-stage least square):
  1. Fit the first-stage regression using all instruments ($Z_1$ and $Z_2$, for example)

$$D = \gamma + \beta_1 Z_1 + \beta_2 Z_2 + \eta \tag{4}$$

# 2SLS estimator

- To solve the multiple instrument problem, we slightly change how we combine the terms

$$
\begin{aligned}
Y &= \alpha + \rho D + \epsilon \\
&= \alpha + \rho(\gamma + \beta Z + \eta) + \epsilon \\
&= \alpha + \rho(\gamma + \beta Z) + (\eta + \rho\epsilon)
\end{aligned}
$$

- Errors $\eta + \rho\epsilon$ are mean independent of $\gamma + \beta Z \implies$ regressing $Y$ on $\gamma + \beta Z$ gives consistent estimate of $\rho$
- $\gamma + \beta Z$ can be approximated by the predicted values of the first stage, $\hat{D} = \hat{\gamma} + \hat{\beta} Z$
- Hence, 2SLS estimator (Two-stage least square):
  1. Fit the first-stage regression using all instruments ($Z_1$ and $Z_2$, for example)

  $$D = \gamma + \beta_1 Z_1 + \beta_2 Z_2 + \eta \tag{4}$$

  2. Regress $Y$ on predicted $D$ (or fitted values)

  $$Y = \alpha + \rho\hat{D} + \epsilon$$

# 2SLS estimator

- To solve the multiple instrument problem, we slightly change how we combine the terms

$$\begin{aligned} Y &= \alpha + \rho D + \epsilon \\ &= \alpha + \rho(\gamma + \beta Z + \eta) + \epsilon \\ &= \alpha + \rho(\gamma + \beta Z) + (\eta + \rho\epsilon) \end{aligned}$$

- Errors $\eta + \rho\epsilon$ are mean independent of $\gamma + \beta Z \implies$ regressing $Y$ on $\gamma + \beta Z$ gives consistent estimate of $\rho$
- $\gamma + \beta Z$ can be approximated by the predicted values of the first stage, $\hat{D} = \hat{\gamma} + \hat{\beta} Z$
- Hence, 2SLS estimator (Two-stage least square):
  1. Fit the first-stage regression using all instruments ($Z_1$ and $Z_2$, for example)
  $$D = \gamma + \beta_1 Z_1 + \beta_2 Z_2 + \eta \tag{4}$$
  2. Regress $Y$ on predicted $D$ (or fitted values)
  $$Y = \alpha + \rho\hat{D} + \epsilon$$

- Coefficient of $D$ from the 2nd stage consistently estimate $\rho$

## Standard Errors

- The standard error estimate of $\rho$ is more complex

## Standard Errors

- The standard error estimate of $\rho$ is more complex
- 2SLS suggests that you can regress $Y$ on predicted $D$; coefficient of $D$ is your estimate of $\rho$

$$Y = \alpha + \rho \hat{D} + \epsilon$$

# Standard Errors

- The standard error estimate of $\rho$ is more complex
- 2SLS suggests that you can regress $Y$ on predicted $D$; coefficient of $D$ is your estimate of $\rho$

$$Y = \alpha + \rho\hat{D} + \epsilon$$

- This gives you a consistent point estimate of $\rho$, but not a consistent estimate of the standard error of $\rho$

# Standard Errors

- The standard error estimate of $\rho$ is more complex
- 2SLS suggests that you can regress $Y$ on predicted $D$; coefficient of $D$ is your estimate of $\rho$

$$Y = \alpha + \rho \hat{D} + \epsilon$$

- This gives you a consistent point estimate of $\rho$, but not a consistent estimate of the standard error of $\rho$
- Directly regress $Y$ on $\hat{D}$ assumes that standard error are related to $\hat{D}$

# Standard Errors

- The standard error estimate of $\rho$ is more complex
- 2SLS suggests that you can regress $Y$ on predicted $D$; coefficient of $D$ is your estimate of $\rho$

$$Y = \alpha + \rho\hat{D} + \epsilon$$

- This gives you a consistent point estimate of $\rho$, but not a consistent estimate of the standard error of $\rho$
- Directly regress $Y$ on $\hat{D}$ assumes that standard error are related to $\hat{D}$
    - But $\hat{D}$ is itself estimated and hence this estimation uncertainty is not taken into consideration

# Standard Errors

- The standard error estimate of $\rho$ is more complex
- 2SLS suggests that you can regress $Y$ on predicted $D$; coefficient of $D$ is your estimate of $\rho$

$$Y = \alpha + \rho\hat{D} + \epsilon$$

- This gives you a consistent point estimate of $\rho$, but not a consistent estimate of the standard error of $\rho$
- Directly regress $Y$ on $\hat{D}$ assumes that standard error are related to $\hat{D}$
  - But $\hat{D}$ is itself estimated and hence this estimation uncertainty is not taken into consideration
- Software (e.g. R's `ivreg` or Stata's `ivregress`) will correct this for you.

# Weak instruments

- Weak instrument: $Z$ is not a good predictor of $D$, or first-stage relevance is not very good

$$\hat{\rho} = \frac{\widehat{\rho\beta}}{\hat{\beta}} = \frac{\hat{E}(Y|Z=1) - \hat{E}(Y|Z=0)}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)}$$

# Weak instruments

- Weak instrument: $Z$ is not a good predictor of $D$, or first-stage relevance is not very good
- Weak instrument can lead to biased estimates

$$\hat{\rho} = \frac{\widehat{\rho\beta}}{\hat{\beta}} = \frac{\hat{E}(Y|Z=1) - \hat{E}(Y|Z=0)}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)}$$

# Weak instruments

- Weak instrument: $Z$ is not a good predictor of $D$, or first-stage relevance is not very good
- Weak instrument can lead to biased estimates
- To get the intuition, use indirect least square or Wald estimator form

$$\hat{\rho} = \frac{\widehat{\rho\beta}}{\hat{\beta}} = \frac{\hat{E}(Y|Z=1) - \hat{E}(Y|Z=0)}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)}$$

# Weak instruments

- Weak instrument: $Z$ is not a good predictor of $D$, or first-stage relevance is not very good
- Weak instrument can lead to biased estimates
- To get the intuition, use indirect least square or Wald estimator form

$$\hat{\rho} = \frac{\widehat{\rho\beta}}{\hat{\beta}} = \frac{\hat{E}(Y|Z=1) - \hat{E}(Y|Z=0)}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)}$$

- Weak instrument $Z$ suggests that $Z$ has little power to distinguish between $\hat{E}(D|Z=1)$ from $\hat{E}(D|Z=0)$

# Weak instruments

- Weak instrument: $Z$ is not a good predictor of $D$, or first-stage relevance is not very good
- Weak instrument can lead to biased estimates
- To get the intuition, use indirect least square or Wald estimator form

$$\hat{\rho} = \frac{\widehat{\rho\beta}}{\hat{\beta}} = \frac{\hat{E}(Y|Z=1) - \hat{E}(Y|Z=0)}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)}$$

- Weak instrument $Z$ suggests that $Z$ has little power to distinguish between $\hat{E}(D|Z=1)$ from $\hat{E}(D|Z=0)$
- In other words, $\hat{E}(D|Z=1) - \hat{E}(D|Z=0)$ is close to 0

# Weak instruments

- Weak instrument: $Z$ is not a good predictor of $D$, or first-stage relevance is not very good
- Weak instrument can lead to biased estimates
- To get the intuition, use indirect least square or Wald estimator form

$$\hat{\rho} = \frac{\widehat{\rho\beta}}{\hat{\beta}} = \frac{\hat{E}(Y|Z=1) - \hat{E}(Y|Z=0)}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)}$$

- Weak instrument $Z$ suggests that $Z$ has little power to distinguish between $\hat{E}(D|Z=1)$ from $\hat{E}(D|Z=0)$
- In other words, $\hat{E}(D|Z=1) - \hat{E}(D|Z=0)$ is close to 0
- Alternatively, first stage regression coefficient $\hat{\beta}$ is close to 0

# Weak instruments

- Weak instrument: $Z$ is not a good predictor of $D$, or first-stage relevance is not very good
- Weak instrument can lead to biased estimates
- To get the intuition, use indirect least square or Wald estimator form

$$\hat{\rho} = \frac{\widehat{\rho\beta}}{\hat{\beta}} = \frac{\hat{E}(Y|Z=1) - \hat{E}(Y|Z=0)}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)}$$

- Weak instrument $Z$ suggests that $Z$ has little power to distinguish between $\hat{E}(D|Z=1)$ from $\hat{E}(D|Z=0)$
- In other words, $\hat{E}(D|Z=1) - \hat{E}(D|Z=0)$ is close to 0
- Alternatively, first stage regression coefficient $\hat{\beta}$ is close to 0
- Small disturbance in the denominator can lead to large bias

# Weak instruments

- Weak instrument: $Z$ is not a good predictor of $D$, or first-stage relevance is not very good
- Weak instrument can lead to biased estimates
- To get the intuition, use indirect least square or Wald estimator form

$$\hat{\rho} = \frac{\widehat{\rho\beta}}{\hat{\beta}} = \frac{\hat{E}(Y|Z=1) - \hat{E}(Y|Z=0)}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)}$$

- Weak instrument $Z$ suggests that $Z$ has little power to distinguish between $\hat{E}(D|Z=1)$ from $\hat{E}(D|Z=0)$
- In other words, $\hat{E}(D|Z=1) - \hat{E}(D|Z=0)$ is close to 0
- Alternatively, first stage regression coefficient $\hat{\beta}$ is close to 0
- Small disturbance in the denominator can lead to large bias
- That's why we require first-stage relevance

# Traditional IV assumptions again

- Second stage: $Y = \alpha + \rho D + \epsilon$

# Traditional IV assumptions again

- Second stage: $Y = \alpha + \rho D + \epsilon$
- First stage: $D = \gamma + \beta Z + \eta$

# Traditional IV assumptions again

- Second stage: $Y = \alpha + \rho D + \epsilon$
- First stage: $D = \gamma + \beta Z + \eta$
- IV Assumptions:

# Traditional IV assumptions again

- Second stage: $Y = \alpha + \rho D + \epsilon$
- First stage: $D = \gamma + \beta Z + \eta$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor (non-testable)

# Traditional IV assumptions again

- Second stage: $Y = \alpha + \rho D + \epsilon$
- First stage: $D = \gamma + \beta Z + \eta$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor (non-testable)
     - $E(\eta|Z) = 0$

# Traditional IV assumptions again

- Second stage: $Y = \alpha + \rho D + \epsilon$
- First stage: $D = \gamma + \beta Z + \eta$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor (non-testable)
     - $E(\eta|Z) = 0$
     - $E(\epsilon|Z) = 0$

# Traditional IV assumptions again

- Second stage: $Y = \alpha + \rho D + \epsilon$
- First stage: $D = \gamma + \beta Z + \eta$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor (non-testable)
     - $E(\eta|Z) = 0$
     - $E(\epsilon|Z) = 0$
  2. Exclusion restriction: (non-testable)

# Traditional IV assumptions again

- Second stage: $Y = \alpha + \rho D + \epsilon$
- First stage: $D = \gamma + \beta Z + \eta$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor (non-testable)
     - $E(\eta|Z) = 0$
     - $E(\epsilon|Z) = 0$
  2. Exclusion restriction: (non-testable)
     - $Z$ does not has a direct effect on $Y$

# Traditional IV assumptions again

- Second stage: $Y = \alpha + \rho D + \epsilon$
- First stage: $D = \gamma + \beta Z + \eta$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor (non-testable)
     - $E(\eta|Z) = 0$
     - $E(\epsilon|Z) = 0$
  2. Exclusion restriction: (non-testable)
     - $Z$ does not has a direct effect on $Y$
     - In linear model, $Z$ does not appear in the second stage model

# Traditional IV assumptions again

- Second stage: $Y = \alpha + \rho D + \epsilon$
- First stage: $D = \gamma + \beta Z + \eta$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor (non-testable)
     - $E(\eta|Z) = 0$
     - $E(\epsilon|Z) = 0$
  2. Exclusion restriction: (non-testable)
     - $Z$ does not has a direct effect on $Y$
     - In linear model, $Z$ does not appear in the second stage model
  3. First stage relevance: (testable)

# Traditional IV assumptions again

- Second stage: $Y = \alpha + \rho D + \epsilon$
- First stage: $D = \gamma + \beta Z + \eta$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor (non-testable)
     - $E(\eta|Z) = 0$
     - $E(\epsilon|Z) = 0$
  2. Exclusion restriction: (non-testable)
     - $Z$ does not has a direct effect on $Y$
     - In linear model, $Z$ does not appear in the second stage model
  3. First stage relevance: (testable)
     - $Z$'s effect on $D$ should not be 0 ($\beta \neq 0$)

# Traditional IV assumptions again

- Second stage: $Y = \alpha + \rho D + \epsilon$
- First stage: $D = \gamma + \beta Z + \eta$
- IV Assumptions:
  1. Exogeneity: $Z$ is an exogenous regressor (non-testable)
     - $E(\eta|Z) = 0$
     - $E(\epsilon|Z) = 0$
  2. Exclusion restriction: (non-testable)
     - $Z$ does not has a direct effect on $Y$
     - In linear model, $Z$ does not appear in the second stage model
  3. First stage relevance: (testable)
     - $Z$'s effect on $D$ should not be 0 ($\beta \neq 0$)
  4. Constant effect $\rho$

## Traditional IV assumptions again

- Second stage: $Y = \alpha + \rho D + \epsilon$
- First stage: $D = \gamma + \beta Z + \eta$
- IV Assumptions:
    1. Exogeneity: $Z$ is an exogenous regressor (non-testable)
        - $E(\eta|Z) = 0$
        - $E(\epsilon|Z) = 0$
    2. Exclusion restriction: (non-testable)
        - $Z$ does not has a direct effect on $Y$
        - In linear model, $Z$ does not appear in the second stage model
    3. First stage relevance: (testable)
        - $Z$'s effect on $D$ should not be 0 ($\beta \neq 0$)
    4. Constant effect $\rho$
        - It will be more clear what this assumption entails later

# IV in randomized experiments

- Traditional IV relies heavily on math tricks; model-based

# IV in randomized experiments

- Traditional IV relies heavily on math tricks; model-based
- How do we understand IV in a design-based framework?

# IV in randomized experiments

- Traditional IV relies heavily on math tricks; model-based
- How do we understand IV in a design-based framework?
    - making your study more like experiments?

# IV in randomized experiments

- Traditional IV relies heavily on math tricks; model-based
- How do we understand IV in a design-based framework?
  - making your study more like experiments?
- In many randomized experiments, you can assign treatment, but cannot force people to take the treatment

# IV in randomized experiments

- Traditional IV relies heavily on math tricks; model-based
- How do we understand IV in a design-based framework?
  - making your study more like experiments?
- In many randomized experiments, you can assign treatment, but cannot force people to take the treatment
- Instrument $Z$ is usually considered as treatment assignment

# IV in randomized experiments

- Traditional IV relies heavily on math tricks; model-based
- How do we understand IV in a design-based framework?
    - making your study more like experiments?
- In many randomized experiments, you can assign treatment, but cannot force people to take the treatment
- Instrument $Z$ is usually considered as treatment assignment
    - which determines actual treatment, but not directly related to outcome

# IV in randomized experiments

- Traditional IV relies heavily on math tricks; model-based
- How do we understand IV in a design-based framework?
  - making your study more like experiments?
- In many randomized experiments, you can assign treatment, but cannot force people to take the treatment
- Instrument $Z$ is usually considered as treatment assignment
  - which determines actual treatment, but not directly related to outcome
  - Or put it differently, the effect of treatment assignment on outcome only matters through actual treatment

# IV in randomized experiments

- Sommer and Zenger (1991)

# IV in randomized experiments

- Sommer and Zenger (1991)
- Goal: Study the effect of vitamin A supplements on infant mortality in Indonesia.

# IV in randomized experiments

- Sommer and Zenger (1991)
- Goal: Study the effect of vitamin A supplements on infant mortality in Indonesia.
- $Z$: treatment assignment; The vitamin supplements were randomly assigned to villages

# IV in randomized experiments

- Sommer and Zenger (1991)
- Goal: Study the effect of vitamin A supplements on infant mortality in Indonesia.
- $Z$: treatment assignment; The vitamin supplements were randomly assigned to villages
- $D$: actual treatment: some of the individuals in villages assigned to the treatment group do not actually take the treatment

# IV in randomized experiments

- Sommer and Zenger (1991)
- Goal: Study the effect of vitamin A supplements on infant mortality in Indonesia.
- $Z$: treatment assignment; The vitamin supplements were randomly assigned to villages
- $D$: actual treatment: some of the individuals in villages assigned to the treatment group do not actually take the treatment
- Non-compliance: people do not follow the treatment they were assigned to

# IV in randomized experiments

- Sommer and Zenger (1991)
- Goal: Study the effect of vitamin A supplements on infant mortality in Indonesia.
- $Z$: treatment assignment; The vitamin supplements were randomly assigned to villages
- $D$: actual treatment: some of the individuals in villages assigned to the treatment group do not actually take the treatment
- Non-compliance: people do not follow the treatment they were assigned to
    - This is deviation from perfect randomized experiments in which everyone obeys his assignment and faithful take it

# IV in randomized experiments

- $Z$ is a valid instrument:

# IV in randomized experiments

- $Z$ is a valid instrument:
  - It is randomly assigned (exogeneity)

# IV in randomized experiments

- $Z$ is a valid instrument:
  - It is randomly assigned (exogeneity)
  - It affect outcome only through $D$, actual treatment (exclusion restriction)

# IV in randomized experiments

- $Z$ is a valid instrument:
  - It is randomly assigned (exogeneity)
  - It affect outcome only through $D$, actual treatment (exclusion restriction)
  - And it certainly matters for $D$ (first-stage relevance)

# IV in randomized experiments

- $Z$ is a valid instrument:
  - It is randomly assigned (exogeneity)
  - It affect outcome only through $D$, actual treatment (exclusion restriction)
  - And it certainly matters for $D$ (first-stage relevance)
- Effect of $Z$ on $Y$ then measures Intent-to-treat (ITT) effect

# IV in randomized experiments

- $Z$ is a valid instrument:
  - It is randomly assigned (exogeneity)
  - It affect outcome only through $D$, actual treatment (exclusion restriction)
  - And it certainly matters for $D$ (first-stage relevance)
- Effect of $Z$ on $Y$ then measures Intent-to-treat (ITT) effect
- Effect of $D$ on $Y$ measures actual treatment effect

# IV in randomized experiments

- $Z$ is a valid instrument:
  - It is randomly assigned (exogeneity)
  - It affect outcome only through $D$, actual treatment (exclusion restriction)
  - And it certainly matters for $D$ (first-stage relevance)
- Effect of $Z$ on $Y$ then measures Intent-to-treat (ITT) effect
- Effect of $D$ on $Y$ measures actual treatment effect
- We formally develop the idea using counterfactual framework

# New IV assumptions

1. Exogeneity: $Z$ is as-if randomly assigned
   - Compare with the previous exogenous regressor assumption, which one do you prefer?

2. Exclusion restriction: $Z$ does not has a direct effect on $Y$

3. First-stage relevance

# New IV assumptions

1. Exogeneity: $Z$ is as-if randomly assigned
   - Compare with the previous exogenous regressor assumption, which one do you prefer?
   - Under counterfactual framework, this implies that potential outcomes are independent of treatment assignment: $Y_i^0, Y_i^1 \perp\!\!\!\perp D_i$
2. Exclusion restriction: $Z$ does not has a direct effect on $Y$

3. First-stage relevance

# New IV assumptions

1. Exogeneity: $Z$ is as-if randomly assigned
   - Compare with the previous exogenous regressor assumption, which one do you prefer?
   - Under counterfactual framework, this implies that potential outcomes are independent of treatment assignment: $Y_i^0, Y_i^1 \perp\!\!\!\perp D_i$
2. Exclusion restriction: $Z$ does not has a direct effect on $Y$
   - Under counterfactual framework, once we fix the value of the actual treatment $D$, $Z$ does not impact $Y$
3. First-stage relevance

# Potential Outcome Framework under Non-Compliance

- Previously we know potential outcome for outcome $Y$

$$Y_i = \begin{cases} Y_i^0 : D_i = 0 \\ Y_i^1 : D_i = 1 \end{cases}$$

$$D_i = \begin{cases} D_i^0 : Z_i = 0 \\ D_i^1 : Z_i = 1 \end{cases}$$

# Potential Outcome Framework under Non-Compliance

- Previously we know potential outcome for outcome $Y$

$$Y_i = \begin{cases} Y_i^0 : D_i = 0 \\ Y_i^1 : D_i = 1 \end{cases}$$

- Now we can define an additional potential outcome for actual treatment $D$ (since it's the "outcome" of treatment assignment)

$$D_i = \begin{cases} D_i^0 : Z_i = 0 \\ D_i^1 : Z_i = 1 \end{cases}$$

# Potential Outcome Framework under Non-Compliance

- Previously we know potential outcome for outcome $Y$

$$Y_i = \begin{cases} Y_i^0 : D_i = 0 \\ Y_i^1 : D_i = 1 \end{cases}$$

- Now we can define an additional potential outcome for actual treatment $D$ (since it's the "outcome" of treatment assignment)
- Angrist, Imbens, and Rubin, 1996, JASA

$$D_i = \begin{cases} D_i^0 : Z_i = 0 \\ D_i^1 : Z_i = 1 \end{cases}$$

# Potential Outcome Framework under Non-Compliance

- Previously we know potential outcome for outcome $Y$

$$Y_i = \begin{cases} Y_i^0 : D_i = 0 \\ Y_i^1 : D_i = 1 \end{cases}$$

- Now we can define an additional potential outcome for actual treatment $D$ (since it's the "outcome" of treatment assignment)
- Angrist, Imbens, and Rubin, 1996, JASA

$$D_i = \begin{cases} D_i^0 : Z_i = 0 \\ D_i^1 : Z_i = 1 \end{cases}$$

- $D_i^0$: potential treatment take-up, if $i$ were not assigned to treatment

# Potential Outcome Framework under Non-Compliance

- Previously we know potential outcome for outcome $Y$

$$Y_i = \begin{cases} Y_i^0 : D_i = 0 \\ Y_i^1 : D_i = 1 \end{cases}$$

- Now we can define an additional potential outcome for actual treatment $D$ (since it's the "outcome" of treatment assignment)
- Angrist, Imbens, and Rubin, 1996, JASA

$$D_i = \begin{cases} D_i^0 : Z_i = 0 \\ D_i^1 : Z_i = 1 \end{cases}$$

- $D_i^0$: potential treatment take-up, if $i$ were not assigned to treatment
- $D_i^1$: potential treatment take-up, if $i$ were assigned to treatment

# Potential Outcome Framework under Non-Compliance

- Previously we know potential outcome for outcome $Y$

$$Y_i = \begin{cases} Y_i^0 : D_i = 0 \\ Y_i^1 : D_i = 1 \end{cases}$$

- Now we can define an additional potential outcome for actual treatment $D$ (since it's the "outcome" of treatment assignment)
- Angrist, Imbens, and Rubin, 1996, JASA

$$D_i = \begin{cases} D_i^0 : Z_i = 0 \\ D_i^1 : Z_i = 1 \end{cases}$$

- $D_i^0$: potential treatment take-up, if $i$ were not assigned to treatment
- $D_i^1$: potential treatment take-up, if $i$ were assigned to treatment
- We only observed $D_i$, not $D_i^0$ and $D_i^1$

# Compliance types in the ideal world

| | $D_i^0 = 0$ | $D_i^0 = 1$ |
|---|---|---|
| $D_i^1 = 0$ | never-taker | defier |
| $D_i^1 = 1$ | complier | always-taker |

- Never-taker: those who never take the treatment regardless of assignment

# Compliance types in the ideal world

| | $D_i^0 = 0$ | $D_i^0 = 1$ |
|---|---|---|
| $D_i^1 = 0$ | never-taker | defier |
| $D_i^1 = 1$ | complier | always-taker |

- Never-taker: those who never take the treatment regardless of assignment
- Always-taker: those who always take the treatment regardless of assignment

# Compliance types in the ideal world

| | $D_i^0 = 0$ | $D_i^0 = 1$ |
|---|---|---|
| $D_i^1 = 0$ | never-taker | defier |
| $D_i^1 = 1$ | complier | always-taker |

- Never-taker: those who never take the treatment regardless of assignment
- Always-taker: those who always take the treatment regardless of assignment
- Complier: those who would always take the treatment if assigned to, and would not if not assigned to treatment

# Compliance types in the ideal world

| | $D_i^0 = 0$ | $D_i^0 = 1$ |
|---|---|---|
| $D_i^1 = 0$ | never-taker | defier |
| $D_i^1 = 1$ | complier | always-taker |

- Never-taker: those who never take the treatment regardless of assignment
- Always-taker: those who always take the treatment regardless of assignment
- Complier: those who would always take the treatment if assigned to, and would not if not assigned to treatment
- Defier: those who would always take the treatment if not assigned to, and vice versa

# Compliance type in observed data

- Each time we only observe one of $D_i^0$ and $D_i^1$, never both

| | $Z_i = 0$ | $Z_i = 1$ |
|---|---|---|
| $D_i = 0$ | never-taker/complier | defier/never-taker |
| $D_i = 1$ | always-taker/defier | always-taker/complier |

# Compliance type in observed data

- Each time we only observe one of $D_i^0$ and $D_i^1$, never both
- Hence, among real (treatment assignment, treatment take-up) pairs

| | $Z_i = 0$ | $Z_i = 1$ |
|---|---|---|
| $D_i = 0$ | never-taker/complier | defier/never-taker |
| $D_i = 1$ | always-taker/defier | always-taker/complier |

# Compliance type in observed data

- Each time we only observe one of $D_i^0$ and $D_i^1$, never both
- Hence, among real (treatment assignment, treatment take-up) pairs
- Each observed subgroup is a mix of two types

| | $Z_i = 0$ | $Z_i = 1$ |
|---|---|---|
| $D_i = 0$ | never-taker/complier | defier/never-taker |
| $D_i = 1$ | always-taker/defier | always-taker/complier |

# Assumption 4: Monotonicity/No-Defiers

- Assumption 4: No-Defiers/Monotonicity

$$D_i^1 \geq D_i^0$$

# Assumption 4: Monotonicity/No-Defiers

- Assumption 4: No-Defiers/Monotonicity
- In math: no $D_i^1 = 0 \& D_i^0 = 1$, or equivalently:

$$D_i^1 \geq D_i^0$$

# Assumption 4: Monotonicity/No-Defiers

- Assumption 4: No-Defiers/Monotonicity
- In math: no $D_i^1 = 0 \& D_i^0 = 1$, or equivalently:

$$D_i^1 \geq D_i^0$$

- That is, potential treatment if $i$ were assigned to treatment should be no less than the potential treatment if $i$ were not assigned to treatment (hence monotonicity)

# Assumption 4: Monotonicity/No-Defiers

- Assumption 4: No-Defiers/Monotonicity
- In math: no $D_i^1 = 0 \& D_i^0 = 1$, or equivalently:

$$D_i^1 \geq D_i^0$$

- That is, potential treatment if $i$ were assigned to treatment should be no less than the potential treatment if $i$ were not assigned to treatment (hence monotonicity)
- This assumption looks very different from the constant effect assumption (Assumption 4) of traditional IV

# Assumption 4: Monotonicity/No-Defiers

- Assumption 4: No-Defiers/Monotonicity
- In math: no $D_i^1 = 0 \& D_i^0 = 1$, or equivalently:

$$D_i^1 \geq D_i^0$$

- That is, potential treatment if $i$ were assigned to treatment should be no less than the potential treatment if $i$ were not assigned to treatment (hence monotonicity)
- This assumption looks very different from the constant effect assumption (Assumption 4) of traditional IV
- And it is easier to understand

# Compliance type in observed data

- With no-defier assumption:

| | $Z_i = 0$ | $Z_i = 1$ |
|---|---|---|
| $D_i = 0$ | never-taker/complier | never-taker |
| $D_i = 1$ | always-taker | always-taker/complier |

# Compliance type in observed data

- With no-defier assumption:

| | $Z_i = 0$ | $Z_i = 1$ |
|---|---|---|
| $D_i = 0$ | never-taker/complier | never-taker |
| $D_i = 1$ | always-taker | always-taker/complier |

- Actual treated units are a mix of always takers and compliers

# Compliance type in observed data

- With no-defier assumption:

| | $Z_i = 0$ | $Z_i = 1$ |
|---|---|---|
| $D_i = 0$ | never-taker/complier | never-taker |
| $D_i = 1$ | always-taker | always-taker/complier |

- Actual treated units are a mix of always takers and compliers
- Actual control units are a mix of never takers and compliers

# LATE Theorem

- Angrist, Imbens, and Rubin, 1996, JASA

$$\frac{\hat{E}(Y|Z=1)-\hat{E}(Y|Z=0)}{\hat{E}(D|Z=1)-\hat{E}(D|Z=0)} = E(Y^1 - Y^0|complier) = \text{LATE} \quad (6)$$

# LATE Theorem

- Angrist, Imbens, and Rubin, 1996, JASA
- With Assumptions 1 - 4, Wald estimator equals to average treatment effect for compliers

$$\frac{\hat{E}(Y|Z=1)-\hat{E}(Y|Z=0)}{\hat{E}(D|Z=1)-\hat{E}(D|Z=0)} = E(Y^1 - Y^0|complier) = \text{LATE} \quad (6)$$

# LATE Theorem

- Angrist, Imbens, and Rubin, 1996, JASA
- With Assumptions 1 - 4, Wald estimator equals to average treatment effect for compliers

$$\frac{\hat{E}(Y|Z=1)-\hat{E}(Y|Z=0)}{\hat{E}(D|Z=1)-\hat{E}(D|Z=0)} = E(Y^1 - Y^0|complier) = \text{LATE} \quad (6)$$

- IV estimates average complier treatment effect (or LATE, local average treatment effect)

# LATE Theorem

- Angrist, Imbens, and Rubin, 1996, JASA
- With Assumptions 1 - 4, Wald estimator equals to average treatment effect for compliers

$$\frac{\hat{E}(Y|Z=1)-\hat{E}(Y|Z=0)}{\hat{E}(D|Z=1)-\hat{E}(D|Z=0)} = E(Y^1 - Y^0|complier) = \text{LATE} \quad (6)$$

- IV estimates average complier treatment effect (or LATE, local average treatment effect)
- Treated units are a mix of compliers and always-takers

# LATE Theorem

- Angrist, Imbens, and Rubin, 1996, JASA
- With Assumptions 1 - 4, Wald estimator equals to average treatment effect for compliers

$$\frac{\hat{E}(Y|Z=1)-\hat{E}(Y|Z=0)}{\hat{E}(D|Z=1)-\hat{E}(D|Z=0)} = E(Y^1 - Y^0|complier) = \text{LATE} \quad (6)$$

- IV estimates average complier treatment effect (or LATE, local average treatment effect)
- Treated units are a mix of compliers and always-takers
- In general, $LATE \neq ATT = E(Y^1 - Y^0|D=1))$

# LATE Theorem

- Angrist, Imbens, and Rubin, 1996, JASA
- With Assumptions 1 - 4, Wald estimator equals to average treatment effect for compliers

$$\frac{\hat{E}(Y|Z=1)-\hat{E}(Y|Z=0)}{\hat{E}(D|Z=1)-\hat{E}(D|Z=0)} = E(Y^1 - Y^0|complier) = \text{LATE} \quad (6)$$

- IV estimates average complier treatment effect (or LATE, local average treatment effect)
- Treated units are a mix of compliers and always-takers
- In general, $LATE \neq ATT = E(Y^1 - Y^0|D=1))$
- Traditional IV "cheats" by assuming constant treatment effect, thus forcing the effect to be the same for compliers and always-takers

# Compilers, always-takers and never-takers in randomized experiments

- Treated units are a mix of compliers and always-takers

# Compilers, always-takers and never-takers in randomized experiments

- Treated units are a mix of compliers and always-takers
- If there is no always-taker, $LATE = ATT$

# Compilers, always-takers and never-takers in randomized experiments

- Treated units are a mix of compliers and always-takers
- If there is no always-taker, $LATE = ATT$
    - If you were to run an experiment, how can you reduce the number of always-takers?

# Compilers, always-takers and never-takers in randomized experiments

- Treated units are a mix of compliers and always-takers
- If there is no always-taker, $LATE = ATT$
  - If you were to run an experiment, how can you reduce the number of always-takers?
  - Think of the the vitamin A experiment as an example.

# Compilers, always-takers and never-takers in randomized experiments

- Treated units are a mix of compliers and always-takers
- If there is no always-taker, $LATE = ATT$
  - If you were to run an experiment, how can you reduce the number of always-takers?
  - Think of the the vitamin A experiment as an example.
- Further, if there also no never-takers, $LATE = ATE$

# Compilers, always-takers and never-takers in randomized experiments

- Treated units are a mix of compliers and always-takers
- If there is no always-taker, $LATE = ATT$
  - If you were to run an experiment, how can you reduce the number of always-takers?
  - Think of the the vitamin A experiment as an example.
- Further, if there also no never-takers, $LATE = ATE$
  - How can you reduce the number of never-takers?

# Compilers, always-takers and never-takers in randomized experiments

- In medical trials for new drugs, often there are no always-taker

| | $Z_i = 0$ | $Z_i = 1$ |
|---|---|---|
| $D_i = 0$ | never-taker/complier | never-taker |
| $D_i = 1$ | | complier |

# Compilers, always-takers and never-takers in randomized experiments

- In medical trials for new drugs, often there are no always-taker
  - Only people who were assigned to treatment were given new drugs

| | $Z_i = 0$ | $Z_i = 1$ |
|---|---|---|
| $D_i = 0$ | never-taker/complier | never-taker |
| $D_i = 1$ | | complier |

# Compilers, always-takers and never-takers in randomized experiments

- In medical trials for new drugs, often there are no always-taker
  - Only people who were assigned to treatment were given new drugs
  - But never-takers are more likely to exist: patients refuse to take new drugs

| | $Z_i = 0$ | $Z_i = 1$ |
|---|---|---|
| $D_i = 0$ | never-taker/complier | never-taker |
| $D_i = 1$ | | complier |

# Compilers, always-takers and never-takers in randomized experiments

- In medical trials for new drugs, often there are no always-taker
  - Only people who were assigned to treatment were given new drugs
  - But never-takers are more likely to exist: patients refuse to take new drugs
  - The chart further simplifies to:

| | $Z_i = 0$ | $Z_i = 1$ |
| --- | --- | --- |
| $D_i = 0$ | never-taker/complier | never-taker |
| $D_i = 1$ | | complier |

# Compilers, always-takers and never-takers in randomized experiments

- In medical trials for new drugs, often there are no always-taker
  - Only people who were assigned to treatment were given new drugs
  - But never-takers are more likely to exist: patients refuse to take new drugs
  - The chart further simplifies to:

| | $Z_i = 0$ | $Z_i = 1$ |
|---|---|---|
| $D_i = 0$ | never-taker/complier | never-taker |
| $D_i = 1$ | | complier |

- $LATE = ATT$ when there is no always-taker

# Compliers, always-takers and never-takers in observational studies

- In observational studies, you have much less control over the data generating process

# Compliers, always-takers and never-takers in observational studies

- In observational studies, you have much less control over the data generating process
- Compliers may be only a small subset of the population

# Compliers, always-takers and never-takers in observational studies

- In observational studies, you have much less control over the data generating process
- Compliers may be only a small subset of the population
- Back to Angrist and Krueger (1991); instrument is quarter of birth, and actual treatment is school length

# Compliers, always-takers and never-takers in observational studies

- In observational studies, you have much less control over the data generating process
- Compliers may be only a small subset of the population
- Back to Angrist and Krueger (1991); instrument is quarter of birth, and actual treatment is school length
- Let us have a thought exercise

# Compliers, always-takers and never-takers in observational studies

- In observational studies, you have much less control over the data generating process
- Compliers may be only a small subset of the population
- Back to Angrist and Krueger (1991); instrument is quarter of birth, and actual treatment is school length
- Let us have a thought exercise
- Who are compliers?

# Compliers, always-takers and never-takers in observational studies

- In observational studies, you have much less control over the data generating process
- Compliers may be only a small subset of the population
- Back to Angrist and Krueger (1991); instrument is quarter of birth, and actual treatment is school length
- Let us have a thought exercise
- Who are compliers?
  - people would be actually utilize the "advantage" that they were born in the 1st quarter to dropout after 16

# ITT vs LATE

- To understand why LATE may not generalize to the whole population, there is an alternative view

$$\begin{aligned} \text{LATE} &= \frac{\hat{E}(Y|Z=1) - \hat{E}(Y|Z=0)}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)} \\ &= \frac{ITT}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)} \end{aligned} \tag{7}$$

# ITT vs LATE

- To understand why LATE may not generalize to the whole population, there is an alternative view
- ITT (intent-to-treat effect; effect on treatment assignment; it's the reduced-form using traditional IV's language)

$$\begin{aligned} \text{LATE} &= \frac{\hat{E}(Y|Z=1) - \hat{E}(Y|Z=0)}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)} \\ &= \frac{ITT}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)} \end{aligned} \tag{7}$$

# ITT vs LATE

- To understand why LATE may not generalize to the whole population, there is an alternative view
- ITT (intent-to-treat effect; effect on treatment assignment; it's the reduced-form using traditional IV's language)

$$
\begin{aligned}
\text{LATE} &= \frac{\hat{E}(Y|Z=1) - \hat{E}(Y|Z=0)}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)} \\
&= \frac{ITT}{\hat{E}(D|Z=1) - \hat{E}(D|Z=0)}
\end{aligned} \tag{7}
$$

- $\hat{E}(D|Z=1) - \hat{E}(D|Z=0) =$ (% treated among those assigned are assigned to treatment) - (% treated among those assigned to the control group)

# ITT vs LATE

- To understand why LATE may not generalize to the whole population, there is an alternative view
- ITT (intent-to-treat effect; effect on treatment assignment; it's the reduced-form using traditional IV's language)

$$\begin{aligned} \text{LATE} &= \frac{\hat{E}(Y|Z=1)-\hat{E}(Y|Z=0)}{\hat{E}(D|Z=1)-\hat{E}(D|Z=0)} \\ &= \frac{ITT}{\hat{E}(D|Z=1)-\hat{E}(D|Z=0)} \end{aligned} \tag{7}$$

- $\hat{E}(D|Z=1)-\hat{E}(D|Z=0) =$ (% treated among those assigned are assigned to treatment) - (% treated among those assigned to the control group)
- LATE is almost always larger than ITT

# ITT vs LATE

- If there is no always-taker (e.g., the new drug example), $E(D|Z=0)=0$

$$ATT = LATE = \frac{ITT}{\hat{E}(D|Z=1)} = \frac{ITT}{\text{compliance rate}} \tag{8}$$

# ITT vs LATE

- If there is no always-taker (e.g., the new drug example), $E(D|Z=0)=0$
- This is known as one-sided non-compliance

$$ATT = LATE = \frac{ITT}{\hat{E}(D|Z=1)} = \frac{ITT}{\text{compliance rate}} \tag{8}$$

# ITT vs LATE

- If there is no always-taker (e.g., the new drug example), $E(D|Z=0)=0$
- This is known as one-sided non-compliance
- Theorem 4.4.2 in MHE:

$$ATT = LATE = \frac{ITT}{\hat{E}(D|Z=1)} = \frac{ITT}{\text{compliance rate}} \tag{8}$$

# ITT vs LATE

- If there is no always-taker (e.g., the new drug example), $E(D|Z=0)=0$
- This is known as one-sided non-compliance
- Theorem 4.4.2 in MHE:

$$ATT = LATE = \frac{ITT}{\hat{E}(D|Z=1)} = \frac{ITT}{\text{compliance rate}} \tag{8}$$

- Estimate of ATT is almost always larger than ITT

# ITT vs LATE

- If there is no always-taker (e.g., the new drug example), $E(D|Z=0)=0$
- This is known as one-sided non-compliance
- Theorem 4.4.2 in MHE:

$$ATT = LATE = \frac{ITT}{\hat{E}(D|Z=1)} = \frac{ITT}{\text{compliance rate}} \tag{8}$$

- Estimate of ATT is almost always larger than ITT
- And if compliance rate is low, ATT will be a lot bigger than ITT

# ITT vs LATE

- If there is no always-taker (e.g., the new drug example), $E(D|Z=0)=0$
- This is known as one-sided non-compliance
- Theorem 4.4.2 in MHE:

$$ATT = LATE = \frac{ITT}{\hat{E}(D|Z=1)} = \frac{ITT}{\text{compliance rate}} \tag{8}$$

- Estimate of ATT is almost always larger than ITT
- And if compliance rate is low, ATT will be a lot bigger than ITT
  - That is, even among the group that were assigned to treatment, due to self-selection into taking the treatment, ATT will be a lot bigger than ITT

# ITT vs LATE

- For some applications, especially policy evaluation for real field experiments

# ITT vs LATE

- For some applications, especially policy evaluation for real field experiments
- ITT might be more important than LATE

# ITT vs LATE

- For some applications, especially policy evaluation for real field experiments
- ITT might be more important than LATE
- E.g., the Vitamin A experiment

# ITT vs LATE

- For some applications, especially policy evaluation for real field experiments
- ITT might be more important than LATE
- E.g., the Vitamin A experiment
- LATE: whether there is an effect for those who want to take it. A pure scientific quantity

# ITT vs LATE

- For some applications, especially policy evaluation for real field experiments
- ITT might be more important than LATE
- E.g., the Vitamin A experiment
- LATE: whether there is an effect for those who want to take it. A pure scientific quantity
- ITT: if gov want to implement the policy (e.g., providing Vitamin A for everyone), they must take into consideration that some people may not follow the instruction. So ITT provides a more faithful evaluation of the effect on the population affected by the policy

# ITT vs LATE

- For some applications, especially policy evaluation for real field experiments
- ITT might be more important than LATE
- E.g., the Vitamin A experiment
- LATE: whether there is an effect for those who want to take it. A pure scientific quantity
- ITT: if gov want to implement the policy (e.g., providing Vitamin A for everyone), they must take into consideration that some people may not follow the instruction. So ITT provides a more faithful evaluation of the effect on the population affected by the policy
- So be careful how compliers differ from the general population if you care more about ITT, instead of LATE

# Continuous treatment

- The counterfactual framework of IV is built upon binary treatment

# Continuous treatment

- The counterfactual framework of IV is built upon binary treatment
- It can be generalized to continuous treatment case

## Continuous treatment

- The counterfactual framework of IV is built upon binary treatment
- It can be generalized to continuous treatment case
- Continuous treatment example: school length

# Continuous treatment

- The counterfactual framework of IV is built upon binary treatment
- It can be generalized to continuous treatment case
- Continuous treatment example: school length
  - LATE: the effect for compliers at point $s$ (those driven by instrument from just below $s$ to at least $s$)

# Continuous treatment

- The counterfactual framework of IV is built upon binary treatment
- It can be generalized to continuous treatment case
- Continuous treatment example: school length
  - LATE: the effect for compliers at point $s$ (those driven by instrument from just below $s$ to at least $s$)
  - And 2SLS estimate is a weighted mean of all LATE at different $s$ values

# Continuous treatment

- The counterfactual framework of IV is built upon binary treatment
- It can be generalized to continuous treatment case
- Continuous treatment example: school length
  - LATE: the effect for compliers at point $s$ (those driven by instrument from just below $s$ to at least $s$)
  - And 2SLS estimate is a weighted mean of all LATE at different $s$ values
- Essentially what you get from 2SLS estimate is some extension of LATE

# Continuous treatment

- The counterfactual framework of IV is built upon binary treatment
- It can be generalized to continuous treatment case
- Continuous treatment example: school length
  - LATE: the effect for compliers at point $s$ (those driven by instrument from just below $s$ to at least $s$)
  - And 2SLS estimate is a weighted mean of all LATE at different $s$ values
- Essentially what you get from 2SLS estimate is some extension of LATE
- Read MHE 4.5 for more extensions

# Continuous treatment

- The counterfactual framework of IV is built upon binary treatment
- It can be generalized to continuous treatment case
- Continuous treatment example: school length
  - LATE: the effect for compliers at point $s$ (those driven by instrument from just below $s$ to at least $s$)
  - And 2SLS estimate is a weighted mean of all LATE at different $s$ values
- Essentially what you get from 2SLS estimate is some extension of LATE
- Read MHE 4.5 for more extensions
  - Also continuous instrument/multiple instruments/adding covariates/HTE with IV

# Pitfalls in implementing IV

Apoorva Lal, Mackenzie William Lockhart, Yiqing Xu, and Ziwen Zu, *How Much Should We Trust Instrumental Variable Estimates in Political Science? Practical Advice based on Over 60 Replicated Studies*, SSRN Scholarly Paper 3905329, Social Science Research Network, Rochester, NY, 2021

- Replicated 64 articles in APSR, AJPS, and JOP that used IV

# Pitfalls in implementing IV

Apoorva Lal, Mackenzie William Lockhart, Yiqing Xu, and Ziwen Zu, *How Much Should We Trust Instrumental Variable Estimates in Political Science? Practical Advice based on Over 60 Replicated Studies*, SSRN Scholarly Paper 3905329, Social Science Research Network, Rochester, NY, 2021

- Replicated 64 articles in APSR, AJPS, and JOP that used IV
- Major problems:

# Pitfalls in implementing IV

Apoorva Lal, Mackenzie William Lockhart, Yiqing Xu, and Ziwen Zu, *How Much Should We Trust Instrumental Variable Estimates in Political Science? Practical Advice based on Over 60 Replicated Studies*, SSRN Scholarly Paper 3905329, Social Science Research Network, Rochester, NY, 2021

- Replicated 64 articles in APSR, AJPS, and JOP that used IV
- Major problems:
- Weak instrument

# Pitfalls in implementing IV

Apoorva Lal, Mackenzie William Lockhart, Yiqing Xu, and Ziwen Zu, *How Much Should We Trust Instrumental Variable Estimates in Political Science? Practical Advice based on Over 60 Replicated Studies*, SSRN Scholarly Paper 3905329, Social Science Research Network, Rochester, NY, 2021

- Replicated 64 articles in APSR, AJPS, and JOP that used IV
- Major problems:
- Weak instrument
- Violation of Exogeneity (or endogenous IV)

# Pitfalls in implementing IV

Apoorva Lal, Mackenzie William Lockhart, Yiqing Xu, and Ziwen Zu, *How Much Should We Trust Instrumental Variable Estimates in Political Science? Practical Advice based on Over 60 Replicated Studies*, SSRN Scholarly Paper 3905329, Social Science Research Network, Rochester, NY, 2021

- Replicated 64 articles in APSR, AJPS, and JOP that used IV
- Major problems:
- Weak instrument
- Violation of Exogeneity (or endogenous IV)
    - "researchers usually spend a great amount of effort verbally arguing for its plausibility based on theories and contextual information"

# Pitfalls in implementing IV

Apoorva Lal, Mackenzie William Lockhart, Yiqing Xu, and Ziwen Zu, *How Much Should We Trust Instrumental Variable Estimates in Political Science? Practical Advice based on Over 60 Replicated Studies*, SSRN Scholarly Paper 3905329, Social Science Research Network, Rochester, NY, 2021

- Replicated 64 articles in APSR, AJPS, and JOP that used IV
- Major problems:
- Weak instrument
- Violation of Exogeneity (or endogenous IV)
    - "researchers usually spend a great amount of effort verbally arguing for its plausibility based on theories and contextual information"
- And the combination of the two exacerbates each other

# Types of IV

Table 2. Types of IVs

| IV Type | #Papers | Percentage% |
|---|---|---|
| **Theory** | 40 | 62.5 |
| Geography/climate/weather | 10.5 | 16.4 |
| History | 10 | 15.6 |
| Treatment diffusion | 2.5 | 3.9 |
| Others | 17 | 26.6 |
| **Experiment** | 12 | 18.8 |
| **Rules & policy changes** | 5 | 7.8 |
| Change in exposure | 3 | 4.7 |
| Fuzzy RD | 2 | 3.1 |
| **Econometrics** | 7 | 10.9 |
| Interactions/"Bartik" | 5 | 7.8 |
| Lagged treatment | 1 | 1.6 |
| Empirical test | 1 | 1.6 |
| **Total** | 64 | 100.0 |

***Note:*** One paper uses both geography-based instruments and an instrument based on treatment diffusion from neighbors. We count 0.5 for each category.

## Weak instrument

- The usual rule of thumb is that first-stage $F$ statistics is larger than 10

# Weak instrument

- The usual rule of thumb is that first-stage $F$ statistics is larger than 10
- David S. Lee, Justin McCrary, Marcelo J. Moreira, and Jack R. Porter, *Valid t-ratio Inference for IV*, Working Paper 29124, National Bureau of Economic Research, 2021 suggests that $F$ should be larger than 104.7

# Weak instrument

- The usual rule of thumb is that first-stage $F$ statistics is larger than 10
- David S. Lee, Justin McCrary, Marcelo J. Moreira, and Jack R. Porter, *Valid t-ratio Inference for IV*, Working Paper 29124, National Bureau of Economic Research, 2021 suggests that $F$ should be larger than 104.7
- Among the 65 articles in top journals of political science

# Weak instrument

- The usual rule of thumb is that first-stage $F$ statistics is larger than 10
- David S. Lee, Justin McCrary, Marcelo J. Moreira, and Jack R. Porter, *Valid t-ratio Inference for IV*, Working Paper 29124, National Bureau of Economic Research, 2021 suggests that $F$ should be larger than 104.7
- Among the 65 articles in top journals of political science
- 81% of the studies have $F > 10$

# Weak instrument

- The usual rule of thumb is that first-stage $F$ statistics is larger than 10
- David S. Lee, Justin McCrary, Marcelo J. Moreira, and Jack R. Porter, *Valid t-ratio Inference for IV*, Working Paper 29124, National Bureau of Economic Research, 2021 suggests that $F$ should be larger than 104.7
- Among the 65 articles in top journals of political science
- 81% of the studies have $F > 10$
- 31% have $F > 104.7$

# Weak instrument

- The usual rule of thumb is that first-stage $F$ statistics is larger than 10
- David S. Lee, Justin McCrary, Marcelo J. Moreira, and Jack R. Porter, *Valid t-ratio Inference for IV*, Working Paper 29124, National Bureau of Economic Research, 2021 suggests that $F$ should be larger than 104.7
- Among the 65 articles in top journals of political science
- 81% of the studies have $F > 10$
- 31% have $F > 104.7$
- Median $F$ for IV in experimental studies is 122.5

# Weak instrument

- The usual rule of thumb is that first-stage $F$ statistics is larger than 10
- David S. Lee, Justin McCrary, Marcelo J. Moreira, and Jack R. Porter, *Valid t-ratio Inference for IV*, Working Paper 29124, National Bureau of Economic Research, 2021 suggests that $F$ should be larger than 104.7
- Among the 65 articles in top journals of political science
- 81% of the studies have $F > 10$
- 31% have $F > 104.7$
- Median $F$ for IV in experimental studies is 122.5
- Median $F$ for IV in observational studies is 41.2

# Weak instrument

- The $F$ statistics depends on how you calculate standard errors in the first-stage regression

# Weak instrument

- The $F$ statistics depends on how you calculate standard errors in the first-stage regression
  - classic SE (the default one from regression)

# Weak instrument

- The $F$ statistics depends on how you calculate standard errors in the first-stage regression
  - classic SE (the default one from regression)
  - Robust standard error

# Weak instrument

- The $F$ statistics depends on how you calculate standard errors in the first-stage regression
  - classic SE (the default one from regression)
  - Robust standard error
  - clustered standard error

# Weak instrument

- The $F$ statistics depends on how you calculate standard errors in the first-stage regression
  - classic SE (the default one from regression)
  - Robust standard error
  - clustered standard error
  - Or bootstrapped standard errors

# Weak instrument

- The $F$ statistics depends on how you calculate standard errors in the first-stage regression
  - classic SE (the default one from regression)
  - Robust standard error
  - clustered standard error
  - Or bootstrapped standard errors
- 70% of times, if we were to use clustered or bootstrapped standard errors, they will make $F$ statistics smaller, hence suggesting that the original studies had a weaker instrument than they had thought

# Weak instrument

- The $F$ statistics depends on how you calculate standard errors in the first-stage regression
  - classic SE (the default one from regression)
  - Robust standard error
  - clustered standard error
  - Or bootstrapped standard errors
- 70% of times, if we were to use clustered or bootstrapped standard errors, they will make $F$ statistics smaller, hence suggesting that the original studies had a weaker instrument than they had thought
- In practice, you should use a more conservation standard error, which also lead to smaller $F$ statistics calculation

# IV vs OLS estimates

- 59 of 64 articles, IV estimate is larger than OLS estimate

# IV vs OLS estimates

- 59 of 64 articles, IV estimate is larger than OLS estimate
- Mean ratio of IV/OLS is 14.3

# IV vs OLS estimates

- 59 of 64 articles, IV estimate is larger than OLS estimate
- Mean ratio of IV/OLS is 14.3
- Median ratio of IV/OLS is 2.6

# IV vs OLS estimates

- 59 of 64 articles, IV estimate is larger than OLS estimate
- Mean ratio of IV/OLS is 14.3
- Median ratio of IV/OLS is 2.6
- This is mainly driven by combination of

# IV vs OLS estimates

- 59 of 64 articles, IV estimate is larger than OLS estimate
- Mean ratio of IV/OLS is 14.3
- Median ratio of IV/OLS is 2.6
- This is mainly driven by combination of
  - weak instrument

# IV vs OLS estimates

- 59 of 64 articles, IV estimate is larger than OLS estimate
- Mean ratio of IV/OLS is 14.3
- Median ratio of IV/OLS is 2.6
- This is mainly driven by combination of
  - weak instrument
  - violation of exogeneity

# IV vs OLS estimates

- 59 of 64 articles, IV estimate is larger than OLS estimate
- Mean ratio of IV/OLS is 14.3
- Median ratio of IV/OLS is 2.6
- This is mainly driven by combination of
  - weak instrument
  - violation of exogeneity
- Note that it's easier to address weak instrument issue, but it's much harder to see if there is violation of exogeneity

# Zero-first-stage tests

- Especially in observational studies where choices of instrument is guided by theory

## Zero-first-stage tests

- Especially in observational studies where choices of instrument is guided by theory
- Intuition: for never-takers, the instrument should have no impact on outcomes.

# Zero-first-stage tests

- Especially in observational studies where choices of instrument is guided by theory
- Intuition: for never-takers, the instrument should have no impact on outcomes.
- So we can use this as a placebo test: run reduced-form regression on the subset of never-takers

## Zero-first-stage tests

- Especially in observational studies where choices of instrument is guided by theory
- Intuition: for never-takers, the instrument should have no impact on outcomes.
- So we can use this as a placebo test: run reduced-form regression on the subset of never-takers
  - note: you also need to use theory to know who are likely to be never-takers