Frontier Topics in Empirical Economics: Week 5
Introduction to IV

Zibin Huang [1]

[1]College of Business, Shanghai University of Finance and Economics

November 30, 2023

Endogeneity: Motivating Example

- Consider the effect of schooling on wage
- Assume linear homogeneous (constant) effect
- For individual i:

$$Y_i = \alpha + \rho s_i + \eta_i \tag{1}$$

Y_i: wage; s_i: schooling; η_i: unobserved term

- If s_i is randomly assigned $\Longrightarrow$ ρ is ATT/ATE
- But s_i is usually an endogenous choice of i
- Selection bias: People attending colleges have higher ability

Endogeneity: Motivating Example

- Consider the effect of schooling on wage
- Assume linear homogeneous (constant) effect
- For individual i:

$$Y_i = \alpha + \rho s_i + \eta_i \tag{1}$$

  Y_i: wage; s_i: schooling; η_i: unobserved term
- If s_i is randomly assigned $\Rightarrow$ ρ is ATT/ATE
- But s_i is usually an endogenous choice of i
- Selection bias: People attending colleges have higher ability

Endogeneity: Motivating Example

- Consider the effect of schooling on wage
- Assume linear homogeneous (constant) effect
- For individual i:

$$Y_i = \alpha + \rho s_i + \eta_i \tag{1}$$

  Y_i: wage; s_i: schooling; η_i: unobserved term
- If s_i is randomly assigned $\Rightarrow$ ρ is ATT/ATE
- But s_i is usually an endogenous choice of i
- Selection bias: People attending colleges have higher ability

Endogeneity: Motivating Example

- Consider the effect of schooling on wage
- Assume linear homogeneous (constant) effect
- For individual i:

$$Y_i = \alpha + \rho s_i + \eta_i \tag{1}$$

  Y_i: wage; s_i: schooling; η_i: unobserved term
- If s_i is randomly assigned $\Rightarrow \rho$ is ATT/ATE
- But s_i is usually an endogenous choice of i
- Selection bias: People attending colleges have higher ability

Endogeneity: Motivating Example

- Consider the effect of schooling on wage
- Assume linear homogeneous (constant) effect
- For individual i:

$$Y_i = \alpha + \rho s_i + \eta_i \tag{1}$$

  Y_i: wage; s_i: schooling; η_i: unobserved term
- If s_i is randomly assigned $\Rightarrow$ ρ is ATT/ATE
- But s_i is usually an endogenous choice of i
- Selection bias: People attending colleges have higher ability

Endogeneity: Motivating Example

- Consider the effect of schooling on wage
- Assume linear homogeneous (constant) effect
- For individual i:

$$Y_i = \alpha + \rho s_i + \eta_i \tag{1}$$

  Y_i: wage; s_i: schooling; η_i: unobserved term
- If s_i is randomly assigned $\Rightarrow$ ρ is ATT/ATE
- But s_i is usually an endogenous choice of i
- Selection bias: People attending colleges have higher ability

Endogeneity: Motivating Example

- Consider the effect of schooling on wage
- Assume linear homogeneous (constant) effect
- For individual i:

$$Y_i = \alpha + \rho s_i + \eta_i \tag{1}$$

  Y_i: wage; s_i: schooling; η_i: unobserved term
- If s_i is randomly assigned $\Rightarrow \rho$ is ATT/ATE
- But s_i is usually an endogenous choice of i
- Selection bias: People attending colleges have higher ability

Endogeneity: Motivating Example

- Assume A_i is ability and we have:

$$\eta_i = \gamma A_i + \nu_i \tag{2}$$

- Assume that $s_i \perp\!\!\!\perp \nu_i$, plug (2) to (4), we have:

$$Y_i = \alpha + \rho s_i + \gamma A_i + \nu_i \tag{3}$$

- What to do if A_i is observed? $\Longrightarrow$ Control it
- What if A_i is not observed? $\Longrightarrow$ Omitted Variable Bias

Endogeneity: Motivating Example

- Assume A_i is ability and we have:

$$\eta_i = \gamma A_i + \nu_i \tag{2}$$

- Assume that $s_i \perp\!\!\!\perp \nu_i$, plug (2) to (4), we have:

$$Y_i = \alpha + \rho s_i + \gamma A_i + \nu_i \tag{3}$$

- What to do if A_i is observed? $\Rightarrow$ Control it
- What if A_i is not observed? $\Rightarrow$ Omitted Variable Bias

Endogeneity: Motivating Example

- Assume A_i is ability and we have:

$$\eta_i = \gamma A_i + \nu_i \tag{2}$$

- Assume that $s_i \perp\!\!\!\perp \nu_i$, plug (2) to (4), we have:

$$Y_i = \alpha + \rho s_i + \gamma A_i + \nu_i \tag{3}$$

- What to do if A_i is observed? $\Longrightarrow$ Control it
- What if A_i is not observed? $\Longrightarrow$ Omitted Variable Bias

Endogeneity: Motivating Example

- Assume A_i is ability and we have:

$$\eta_i = \gamma A_i + \nu_i \tag{2}$$

- Assume that $s_i \perp\!\!\!\perp \nu_i$, plug (2) to (4), we have:

$$Y_i = \alpha + \rho s_i + \gamma A_i + \nu_i \tag{3}$$

- What to do if A_i is observed? $\implies$ Control it
- What if A_i is not observed? $\implies$ Omitted Variable Bias

Endogeneity: Motivating Example

- Assume A_i is ability and we have:

$$\eta_i = \gamma A_i + \nu_i \tag{2}$$

- Assume that $s_i \perp\!\!\!\perp \nu_i$, plug (2) to (4), we have:

$$Y_i = \alpha + \rho s_i + \gamma A_i + \nu_i \tag{3}$$

- What to do if A_i is observed? $\Rightarrow$ Control it
- What if A_i is not observed? $\Rightarrow$ Omitted Variable Bias

Simple IV: Definition

Let's focus on the simplest case first:
Single endogenous variable, single instrument, constant treatment effect

- Assume that, there is a variable z_i, such that

$$(1)\ z_i \perp\!\!\!\perp \eta_i \quad \text{(Exclusion Restriction)}$$
$$(2)\ Cov(s_i, z_i) \neq 0 \quad \text{(Existence of First Stage)}$$

We call it an "Instrumental Variable" (IV).

Simple IV: Definition

Let's focus on the simplest case first:
Single endogenous variable, single instrument, constant treatment effect

- Assume that, there is a variable z_i, such that

$$(1)\ z_i \perp\!\!\!\perp \eta_i \quad \text{(Exclusion Restriction)}$$
$$(2)\ Cov(s_i, z_i) \neq 0 \quad \text{(Existence of First Stage)}$$

We call it an "Instrumental Variable" (IV).

Simple IV: Definition

Let's focus on the simplest case first:
Single endogenous variable, single instrument, constant treatment effect

- Assume that, there is a variable z_i, such that

$$(1)\ z_i \perp\!\!\!\perp \eta_i \quad \text{(Exclusion Restriction)}$$
$$(2)\ Cov(s_i, z_i) \neq 0 \quad \text{(Existence of First Stage)}$$

We call it an "Instrumental Variable" (IV).

Simple IV: Identification

- Calculating covariance of z_i and Y_i:

$$Cov(z_i, Y_i) = Cov(z_i, \alpha + \rho s_i + \eta_i) = \rho Cov(z_i, s_i)$$

$$\Rightarrow \rho = \frac{Cov(z_i, Y_i)}{Cov(z_i, s_i)} = \frac{Cov(z_i, Y_i)/Var(z_i)}{Cov(z_i, s_i)/Var(z_i)}$$

  Thus, treatment effect is identified by dividing two correlations.

- When IV z_i is binary:

$$\rho = \frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{E[s_i|z_i = 1] - E[s_i|z_i = 0]}$$

Simple IV: Identification

- Calculating covariance of z_i and Y_i:

$$Cov(z_i, Y_i) = Cov(z_i, \alpha + \rho s_i + \eta_i) = \rho Cov(z_i, s_i)$$
$$\Rightarrow \rho = \frac{Cov(z_i, Y_i)}{Cov(z_i, s_i)} = \frac{Cov(z_i, Y_i)/Var(z_i)}{Cov(z_i, s_i)/Var(z_i)}$$

Thus, treatment effect is identified by dividing two correlations.

- When IV z_i is binary:

$$\rho = \frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{E[s_i|z_i = 1] - E[s_i|z_i = 0]}$$

Simple IV: Identification

- Calculating covariance of z_i and Y_i:

$$Cov(z_i, Y_i) = Cov(z_i, \alpha + \rho s_i + \eta_i) = \rho Cov(z_i, s_i)$$
$$\Rightarrow \rho = \frac{Cov(z_i, Y_i)}{Cov(z_i, s_i)} = \frac{Cov(z_i, Y_i)/Var(z_i)}{Cov(z_i, s_i)/Var(z_i)}$$

  Thus, treatment effect is identified by dividing two correlations.

- When IV z_i is binary:

$$\rho = \frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{E[s_i|z_i = 1] - E[s_i|z_i = 0]}$$

Simple IV: Wald Estimator

- Correlations are regression coefficients (single variable):

$$s_i = \alpha + \pi_1 z_i + \eta_i \quad \text{(First Stage)}$$
$$Y_i = \alpha + \pi_2 z_i + \eta_i \quad \text{(Reduced Form)}$$
$$\rho = \frac{\pi_2}{\pi_1}$$

- Estimation of ρ is simple:

$$\hat{\rho}_{wald} = \frac{\hat{\pi}_2^{ols}}{\hat{\pi}_1^{ols}}$$

- We call this Wald/IV estimator

Simple IV: Wald Estimator

- Correlations are regression coefficients (single variable):

$$s_i = \alpha + \pi_1 z_i + \eta_i \quad \text{(First Stage)}$$
$$Y_i = \alpha + \pi_2 z_i + \eta_i \quad \text{(Reduced Form)}$$
$$\rho = \frac{\pi_2}{\pi_1}$$

- Estimation of ρ is simple:

$$\hat{\rho}_{wald} = \frac{\hat{\pi}_2^{ols}}{\hat{\pi}_1^{ols}}$$

- We call this Wald/IV estimator

Simple IV: Wald Estimator

- Correlations are regression coefficients (single variable):

$$s_i = \alpha + \pi_1 z_i + \eta_i \quad \text{(First Stage)}$$
$$Y_i = \alpha + \pi_2 z_i + \eta_i \quad \text{(Reduced Form)}$$
$$\rho = \frac{\pi_2}{\pi_1}$$

- Estimation of ρ is simple:

$$\hat{\rho}_{wald} = \frac{\hat{\pi}_2^{ols}}{\hat{\pi}_1^{ols}}$$

- We call this Wald/IV estimator

Simple IV: Wald Estimator

- Correlations are regression coefficients (single variable):

$$
\begin{aligned}
s_i &= \alpha + \pi_1 z_i + \eta_i \quad \text{(First Stage)} \\
Y_i &= \alpha + \pi_2 z_i + \eta_i \quad \text{(Reduced Form)} \\
\rho &= \frac{\pi_2}{\pi_1}
\end{aligned}
$$

- Estimation of ρ is simple:

$$\hat{\rho}_{wald} = \frac{\hat{\pi}_2^{ols}}{\hat{\pi}_1^{ols}}$$

- We call this Wald/IV estimator

Simple IV: 2SLS

- Another way of using IV is Two-Stage Least Squares (2SLS)
- Assume that we have the following main and first stage equation:

$$Y_i = X_i'\alpha + \rho s_i + \eta_i \tag{4}$$

$$s_i = X_i'\pi_{10} + \pi_{11} z_i + \xi_{1i} \tag{5}$$

- X_i is a set of control variables.

Simple IV: 2SLS

- Another way of using IV is Two-Stage Least Squares (2SLS)
- Assume that we have the following main and first stage equation:

$$Y_i = X_i^{\prime}\alpha + \rho s_i + \eta_i \tag{4}$$

$$s_i = X_i^{\prime}\pi_{10} + \pi_{11} z_i + \xi_{1i} \tag{5}$$

- X_i is a set of control variables.

Simple IV: 2SLS

- Another way of using IV is Two-Stage Least Squares (2SLS)
- Assume that we have the following main and first stage equation:

$$Y_i = X_i'\alpha + \rho s_i + \eta_i \quad (4)$$

$$s_i = X_i'\pi_{10} + \pi_{11} z_i + \xi_{1i} \quad (5)$$

- X_i is a set of control variables.

Simple IV: 2SLS

- Another way of using IV is Two-Stage Least Squares (2SLS)
- Assume that we have the following main and first stage equation:

$$Y_i = X_i'\alpha + \rho s_i + \eta_i \tag{4}$$

$$s_i = X_i'\pi_{10} + \pi_{11} z_i + \xi_{1i} \tag{5}$$

- X_i is a set of control variables.

Simple IV: 2SLS

- Plug (5) into (4):

$$\begin{aligned} Y_i &= \alpha' X_i + \rho(X_i' \pi_{10} + \pi_{11} z_i + \xi_{1i}) + \eta_i \\ &= \alpha' X_i + \rho(X_i' \pi_{10} + \pi_{11} z_i) + \xi_{2i} \end{aligned} \tag{6}$$

- Because $\xi_{2i} = \rho\xi_{1i} + \eta_i$, we have $z_i \perp \xi_{2i}$
- $(X_i' \pi_{10} + \pi_{11} z_i)$ is the CEF/regression prediction of s_i on z_i given X_i

Simple IV: 2SLS

- Plug (5) into (4):

$$\begin{aligned} Y_i &= \alpha' X_i + \rho(X_i' \pi_{10} + \pi_{11} z_i + \xi_{1i}) + \eta_i \\ &= \alpha' X_i + \rho(X_i' \pi_{10} + \pi_{11} z_i) + \xi_{2i} \end{aligned} \quad (6)$$

- Because $\xi_{2i} = \rho\xi_{1i} + \eta_i$, we have $z_i \perp \xi_{2i}$
- $(X_i'\pi_{10} + \pi_{11} z_i)$ is the CEF/regression prediction of s_i on z_i given X_i

Simple IV: 2SLS

- Plug (5) into (4):

$$
\begin{aligned}
Y_i &= \alpha' X_i + \rho(X_i'\pi_{10} + \pi_{11} z_i + \xi_{1i}) + \eta_i \\
&= \alpha' X_i + \rho(X_i'\pi_{10} + \pi_{11} z_i) + \xi_{2i}
\end{aligned} \tag{6}
$$

- Because $\xi_{2i} = \rho\xi_{1i} + \eta_i$, we have $z_i \perp \xi_{2i}$
- $(X_i'\pi_{10} + \pi_{11} z_i)$ is the CEF/regression prediction of s_i on z_i given X_i

Simple IV: 2SLS

- Plug (5) into (4):

$$\begin{aligned} Y_i &= \alpha' X_i + \rho(X_i' \pi_{10} + \pi_{11} z_i + \xi_{1i}) + \eta_i \\ &= \alpha' X_i + \rho(X_i' \pi_{10} + \pi_{11} z_i) + \xi_{2i} \end{aligned} \tag{6}$$

- Because $\xi_{2i} = \rho\xi_{1i} + \eta_i$, we have $z_i \perp \xi_{2i}$
- $(X_i' \pi_{10} + \pi_{11} z_i)$ is the CEF/regression prediction of s_i on z_i given X_i

Simple IV: 2SLS

- Procedure of 2SLS estimation of ρ:
  - Step 1: Running s on both z and X to get the predicted value $\hat{s}$

$$\hat{s}_i = X_i'\hat{\pi}_{10} + \hat{\pi}_{11} z_i$$

  - Step 2: Running Y on predicted value $\hat{s}$ and X_i

$$Y_i = \alpha' X_i + \rho \hat{s}_i + \xi_i$$

Simple IV: 2SLS

- Procedure of 2SLS estimation of ρ:
  - Step 1: Running s on both z and X to get the predicted value $\hat{s}$

$$\hat{s}_i = X_i'\hat{\pi}_{10} + \hat{\pi}_{11} z_i$$

  - Step 2: Running Y on predicted value $\hat{s}$ and X_i

$$Y_i = \alpha' X_i + \rho \hat{s}_i + \xi_{2i}'$$

Simple IV: 2SLS

- Procedure of 2SLS estimation of ρ:
  - Step 1: Running s on both z and X to get the predicted value $\hat{s}$

$$\hat{s}_i = X_i'\hat{\pi}_{10} + \hat{\pi}_{11} z_i$$

  - Step 2: Running Y on predicted value $\hat{s}$ and X_i

$$Y_i = \alpha' X_i + \rho \hat{s}_i + \xi_{2i}'$$

Simple IV: 2SLS

- Procedure of 2SLS estimation of ρ:
  - Step 1: Running s on both z and X to get the predicted value $\hat{s}$

$$\hat{s}_i = X_i'\hat{\pi}_{10} + \hat{\pi}_{11} z_i$$

  - Step 2: Running Y on predicted value $\hat{s}$ and X_i

$$Y_i = \alpha' X_i + \rho \hat{s}_i + \xi_{2i}'$$

Simple IV: Some Tips

- In 2SLS, you need to control the same X_i in both steps
- Never do 2SLS by hand, use packages in Stata
  OLS second stage std err is wrong.
- Do we need causal interpretation for first stage? No!
  You can always run regressions without causal meanings.
- But in practice it is better you have a reason to believe that Z affects X
- Wald estimator is only available when # of endogenous variables equals # of IVs
- When # of endogenous variables equals # of IVs (just-identified)
  2SLS estimator is identical to Wald estimator
- In general, 2SLS is relatively efficient (best under homosk)

Simple IV: Some Tips

- In 2SLS, you need to control the same X_i in both steps
- Never do 2SLS by hand, use packages in Stata
  OLS second stage std err is wrong.
- Do we need causal interpretation for first stage? No!
  You can always run regressions without causal meanings.
- But in practice it is better you have a reason to believe that Z affects X
- Wald estimator is only available when # of endogenous variables equals # of IVs
- When # of endogenous variables equals # of IVs (just-identified)
  2SLS estimator is identical to Wald estimator
- In general, 2SLS is relatively efficient (best under homosk)

Simple IV: Some Tips

- In 2SLS, you need to control the same X_i in both steps
- Never do 2SLS by hand, use packages in Stata
  OLS second stage std err is wrong.
- Do we need causal interpretation for first stage? No!
  You can always run regressions without causal meanings.
- But in practice it is better you have a reason to believe that Z affects X
- Wald estimator is only available when # of endogenous variables equals # of IVs
- When # of endogenous variables equals # of IVs (just-identified)
  2SLS estimator is identical to Wald estimator
- In general, 2SLS is relatively efficient (best under homosk)

Simple IV: Some Tips

- In 2SLS, you need to control the same X_i in both steps
- Never do 2SLS by hand, use packages in Stata
  OLS second stage std err is wrong.
- Do we need causal interpretation for first stage? No!
  You can always run regressions without causal meanings.
- But in practice it is better you have a reason to believe that Z affects X
- Wald estimator is only available when # of endogenous variables equals # of IVs
- When # of endogenous variables equals # of IVs (just-identified)
  2SLS estimator is identical to Wald estimator
- In general, 2SLS is relatively efficient (best under homosk)

Simple IV: Some Tips

- In 2SLS, you need to control the same X_i in both steps
- Never do 2SLS by hand, use packages in Stata
  OLS second stage std err is wrong.
- Do we need causal interpretation for first stage? No!
  You can always run regressions without causal meanings.
- But in practice it is better you have a reason to believe that Z affects X
- Wald estimator is only available when # of endogenous variables equals # of IVs
- When # of endogenous variables equals # of IVs (just-identified)
  2SLS estimator is identical to Wald estimator
- In general, 2SLS is relatively efficient (best under homosk)

Simple IV: Some Tips

- In 2SLS, you need to control the same X_i in both steps
- Never do 2SLS by hand, use packages in Stata
  OLS second stage std err is wrong.
- Do we need causal interpretation for first stage? No!
  You can always run regressions without causal meanings.
- But in practice it is better you have a reason to believe that Z affects X
- Wald estimator is only available when # of endogenous variables equals # of IVs
- When # of endogenous variables equals # of IVs (just-identified)
  2SLS estimator is identical to Wald estimator
- In general, 2SLS is relatively efficient (best under homosk)

Simple IV: Some Tips

- In 2SLS, you need to control the same X_i in both steps
- Never do 2SLS by hand, use packages in Stata
  OLS second stage std err is wrong.
- Do we need causal interpretation for first stage? No!
  You can always run regressions without causal meanings.
- But in practice it is better you have a reason to believe that Z affects X
- Wald estimator is only available when # of endogenous variables equals # of IVs
- When # of endogenous variables equals # of IVs (just-identified)
  2SLS estimator is identical to Wald estimator
- In general, 2SLS is relatively efficient (best under homosk)

Simple IV: Some Tips

- In 2SLS, you need to control the same X_i in both steps
- Never do 2SLS by hand, use packages in Stata
  OLS second stage std err is wrong.
- Do we need causal interpretation for first stage? No!
  You can always run regressions without causal meanings.
- But in practice it is better you have a reason to believe that Z affects X
- Wald estimator is only available when # of endogenous variables equals # of IVs
- When # of endogenous variables equals # of IVs (just-identified)
  2SLS estimator is identical to Wald estimator
- In general, 2SLS is relatively efficient (best under homosk)

IV with Heterogeneous Treatment Effect: Settings

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- Now we relax (3) to have heterogeneous treatment effect
- Motivating example: Military service on earning (Angrist and Krueger 1992)
  Y_i: wage earning; D_i: whether served in the army before; z_i: draft lottery number below cutoff (draft eligible)
- During the Vietnam War, young men in the U.S. were drafted to the army
- A random draft lottery number was assigned to each birthday
- Man with a number below the cutoff is likely to be drafted

IV with Heterogeneous Treatment Effect: Settings

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- Now we relax (3) to have heterogeneous treatment effect
- Motivating example: Military service on earning (Angrist and Krueger 1992)
  Y_i: wage earning; D_i: whether served in the army before; z_i: draft lottery number below cutoff (draft eligible)
- During the Vietnam War, young men in the U.S. were drafted to the army
- A random draft lottery number was assigned to each birthday
- Man with a number below the cutoff is likely to be drafted

IV with Heterogeneous Treatment Effect: Settings

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- Now we relax (3) to have heterogeneous treatment effect
- Motivating example: Military service on earning (Angrist and Krueger 1992)
  Y_i: wage earning; D_i: whether served in the army before; z_i: draft lottery number below cutoff (draft eligible)
- During the Vietnam War, young men in the U.S. were drafted to the army
- A random draft lottery number was assigned to each birthday
- Man with a number below the cutoff is likely to be drafted

IV with Heterogeneous Treatment Effect: Settings

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- Now we relax (3) to have heterogeneous treatment effect
- Motivating example: Military service on earning (Angrist and Krueger 1992)
  Y_i: wage earning; D_i: whether served in the army before; z_i: draft lottery number below cutoff (draft eligible)
- During the Vietnam War, young men in the U.S. were drafted to the army
- A random draft lottery number was assigned to each birthday
- Man with a number below the cutoff is likely to be drafted

IV with Heterogeneous Treatment Effect: Settings

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- Now we relax (3) to have heterogeneous treatment effect
- Motivating example: Military service on earning (Angrist and Krueger 1992)
  Y_i: wage earning; D_i: whether served in the army before; z_i: draft lottery number below cutoff (draft eligible)
- During the Vietnam War, young men in the U.S. were drafted to the army
- A random draft lottery number was assigned to each birthday
- Man with a number below the cutoff is likely to be drafted

IV with Heterogeneous Treatment Effect: Settings

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- Now we relax (3) to have heterogeneous treatment effect
- Motivating example: Military service on earning (Angrist and Krueger 1992)
  Y_i: wage earning; D_i: whether served in the army before; z_i: draft lottery number below cutoff (draft eligible)
- During the Vietnam War, young men in the U.S. were drafted to the army
- A random draft lottery number was assigned to each birthday
- Man with a number below the cutoff is likely to be drafted

IV with Heterogeneous Treatment Effect: Settings

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- Now we relax (3) to have heterogeneous treatment effect
- Motivating example: Military service on earning (Angrist and Krueger 1992)
  Y_i: wage earning; D_i: whether served in the army before; z_i: draft lottery number below cutoff (draft eligible)
- During the Vietnam War, young men in the U.S. were drafted to the army
- A random draft lottery number was assigned to each birthday
- Man with a number below the cutoff is likely to be drafted

IV with Heterogeneous Treatment Effect: Settings

- We define two potential outcomes
- $Y_i(d, z)$: Potential final outcome (wage), given treatment (military service) and instrument (draft number)
- D_{1i}, D_{0i}: Potential treatment outcome (military service), given instrument (draft number)
- Now we introduce four assumptions needed for LATE Theorem
- Assumption 1: Independence

$$\{Y_i(D_{1i}, 1), Y_i(D_{0i}, 0), D_{1i}, D_{0i}\} \perp\!\!\!\perp z_i$$

- Instrument is assigned as good as random $\iff$ instrument is independent of potential outcome and potential treatment (agent type)

IV with Heterogeneous Treatment Effect: Settings

- We define two potential outcomes
- $Y_i(d, z)$: Potential final outcome (wage), given treatment (military service) and instrument (draft number)
- D_{1i}, D_{0i}: Potential treatment outcome (military service), given instrument (draft number)
- Now we introduce four assumptions needed for LATE Theorem
- Assumption 1: Independence

$$\{Y_i(D_{1i}, 1), Y_i(D_{0i}, 0), D_{1i}, D_{0i}\} \perp\!\!\!\perp z_i$$

- Instrument is assigned as good as random $\iff$ instrument is independent of potential outcome and potential treatment (agent type)

IV with Heterogeneous Treatment Effect: Settings

- We define two potential outcomes
- $Y_i(d, z)$: Potential final outcome (wage), given treatment (military service) and instrument (draft number)
- D_{1i}, D_{0i}: Potential treatment outcome (military service), given instrument (draft number)
- Now we introduce four assumptions needed for LATE Theorem
- Assumption 1: Independence

$$\{Y_i(D_{1i}, 1), Y_i(D_{0i}, 0), D_{1i}, D_{0i}\} \perp\!\!\!\perp z_i$$

- Instrument is assigned as good as random $\iff$ instrument is independent of potential outcome and potential treatment (agent type)

IV with Heterogeneous Treatment Effect: Settings

- We define two potential outcomes
- $Y_i(d, z)$: Potential final outcome (wage), given treatment (military service) and instrument (draft number)
- D_{1i}, D_{0i}: Potential treatment outcome (military service), given instrument (draft number)
- Now we introduce four assumptions needed for LATE Theorem
- Assumption 1: Independence

$$\{Y_i(D_{1i}, 1), Y_i(D_{0i}, 0), D_{1i}, D_{0i}\} \perp\!\!\!\perp z_i$$

- Instrument is assigned as good as random $\iff$ instrument is independent of potential outcome and potential treatment (agent type)

IV with Heterogeneous Treatment Effect: Settings

- We define two potential outcomes
- $Y_i(d, z)$: Potential final outcome (wage), given treatment (military service) and instrument (draft number)
- D_{1i}, D_{0i}: Potential treatment outcome (military service), given instrument (draft number)
- Now we introduce four assumptions needed for LATE Theorem
- Assumption 1: Independence

$$\{Y_i(D_{1i}, 1), Y_i(D_{0i}, 0), D_{1i}, D_{0i}\} \perp\!\!\!\perp z_i$$

- Instrument is assigned as good as random $\iff$ instrument is independent of potential outcome and potential treatment (agent type)

IV with Heterogeneous Treatment Effect: Settings

- We define two potential outcomes
- $Y_i(d, z)$: Potential final outcome (wage), given treatment (military service) and instrument (draft number)
- D_{1i}, D_{0i}: Potential treatment outcome (military service), given instrument (draft number)
- Now we introduce four assumptions needed for LATE Theorem
- Assumption 1: Independence

$$\{Y_i(D_{1i}, 1), Y_i(D_{0i}, 0), D_{1i}, D_{0i}\} \perp\!\!\!\perp z_i$$

- Instrument is assigned as good as random $\iff$ instrument is independent of potential outcome and potential treatment (agent type)

IV with Heterogeneous Treatment Effect: Settings

- We define two potential outcomes
- $Y_i(d, z)$: Potential final outcome (wage), given treatment (military service) and instrument (draft number)
- D_{1i}, D_{0i}: Potential treatment outcome (military service), given instrument (draft number)
- Now we introduce four assumptions needed for LATE Theorem
- Assumption 1: Independence

$$\{Y_i(D_{1i}, 1), Y_i(D_{0i}, 0), D_{1i}, D_{0i}\} \perp\!\!\!\perp z_i$$

- Instrument is assigned as good as random $\Longleftrightarrow$ instrument is independent of potential outcome and potential treatment (agent type)

IV with Heterogeneous Treatment Effect: Settings

- Assumption 2: Exclusion

$$Y_i(d,0) = Y_i(d,1) \equiv Y_{di} \quad \text{for d=0,1}$$

- Instrument can only affect final outcome through treatment
- Example: Draft number affects future wages only by changing military service experience, but not other channel (education etc)
- Assumption 3: Existence of first stage

$$E[D_{1i} - D_{0i}] \neq 0$$

IV with Heterogeneous Treatment Effect: Settings

- Assumption 2: Exclusion

$$Y_i(d,0) = Y_i(d,1) \equiv Y_{di} \quad \text{for d=0,1}$$

- Instrument can only affect final outcome through treatment
- Example: Draft number affects future wages only by changing military service experience, but not other channel (education etc)
- Assumption 3: Existence of first stage

$$E[D_{1i} - D_{0i}] \neq 0$$

IV with Heterogeneous Treatment Effect: Settings

- Assumption 2: Exclusion

$$Y_i(d,0) = Y_i(d,1) \equiv Y_{di} \quad \text{for d=0,1}$$

- Instrument can only affect final outcome through treatment
- Example: Draft number affects future wages only by changing military service experience, but not other channel (education etc)
- Assumption 3: Existence of first stage

$$E[D_{1i} - D_{0i}] \neq 0$$

IV with Heterogeneous Treatment Effect: Settings

- Assumption 2: Exclusion

$$Y_i(d,0) = Y_i(d,1) \equiv Y_{di} \quad \text{for d=0,1}$$

- Instrument can only affect final outcome through treatment
- Example: Draft number affects future wages only by changing military service experience, but not other channel (education etc)
- Assumption 3: Existence of first stage

$$E[D_{1i} - D_{0i}] \neq 0$$

IV with Heterogeneous Treatment Effect: Settings

- Assumption 2: Exclusion

$$Y_i(d,0) = Y_i(d,1) \equiv Y_{di} \quad \text{for d=0,1}$$

- Instrument can only affect final outcome through treatment
- Example: Draft number affects future wages only by changing military service experience, but not other channel (education etc)
- Assumption 3: Existence of first stage

$$E[D_{1i} - D_{0i}] \neq 0$$

IV with Heterogeneous Treatment Effect: Settings

- Assumption 4: Monotonicity

$$\forall i, D_{1i} - D_{0i} \geq 0 \quad \text{or vice versa}$$

- For everyone, instrument changes treatment in the same direction (or no change)
- Example: For a person who will serve (voluntarily) even when his number is above the cutoff, he will of course serve if his number is below the cutoff
- Complier: $D_{1i} > D_{0i}$ people who change their choice by instrument
- Always-taker: $D_{1i} = D_{0i} = 1$ people who always take treatment
- Never-taker: $D_{1i} = D_{0i} = 0$ people who always do not take treatment
- No defiers!

IV with Heterogeneous Treatment Effect: Settings

- Assumption 4: Monotonicity

$$\forall i, D_{1i} - D_{0i} \geq 0 \quad \text{or vice versa}$$

- For everyone, instrument changes treatment in the same direction (or no change)
- Example: For a person who will serve (voluntarily) even when his number is above the cutoff, he will of course serve if his number is below the cutoff
- Complier: $D_{1i} > D_{0i}$ people who change their choice by instrument
- Always-taker: $D_{1i} = D_{0i} = 1$ people who always take treatment
- Never-taker: $D_{1i} = D_{0i} = 0$ people who always do not take treatment
- No defiers!

IV with Heterogeneous Treatment Effect: Settings

- Assumption 4: Monotonicity

$$\forall i, D_{1i} - D_{0i} \geq 0 \quad \text{or vice versa}$$

- For everyone, instrument changes treatment in the same direction (or no change)
- Example: For a person who will serve (voluntarily) even when his number is above the cutoff, he will of course serve if his number is below the cutoff
- Complier: $D_{1i} > D_{0i}$ people who change their choice by instrument
- Always-taker: $D_{1i} = D_{0i} = 1$ people who always take treatment
- Never-taker: $D_{1i} = D_{0i} = 0$ people who always do not take treatment
- No defiers!

IV with Heterogeneous Treatment Effect: Settings

- Assumption 4: Monotonicity

$$\forall i, D_{1i} - D_{0i} \geq 0 \quad \text{or vice versa}$$

- For everyone, instrument changes treatment in the same direction (or no change)
- Example: For a person who will serve (voluntarily) even when his number is above the cutoff, he will of course serve if his number is below the cutoff
- Complier: $D_{1i} > D_{0i}$ people who change their choice by instrument
- Always-taker: $D_{1i} = D_{0i} = 1$ people who always take treatment
- Never-taker: $D_{1i} = D_{0i} = 0$ people who always do not take treatment
- No defiers!

IV with Heterogeneous Treatment Effect: Settings

- Assumption 4: Monotonicity

$$\forall i, D_{1i} - D_{0i} \geq 0 \quad \text{or vice versa}$$

- For everyone, instrument changes treatment in the same direction (or no change)
- Example: For a person who will serve (voluntarily) even when his number is above the cutoff, he will of course serve if his number is below the cutoff
- Complier: $D_{1i} > D_{0i}$ people who change their choice by instrument
- Always-taker: $D_{1i} = D_{0i} = 1$ people who always take treatment
- Never-taker: $D_{1i} = D_{0i} = 0$ people who always do not take treatment
- No defiers!

IV with Heterogeneous Treatment Effect: Settings

- Assumption 4: Monotonicity

$$\forall i, D_{1i} - D_{0i} \geq 0 \quad \text{or vice versa}$$

- For everyone, instrument changes treatment in the same direction (or no change)
- Example: For a person who will serve (voluntarily) even when his number is above the cutoff, he will of course serve if his number is below the cutoff
- Complier: $D_{1i} > D_{0i}$ people who change their choice by instrument
- Always-taker: $D_{1i} = D_{0i} = 1$ people who always take treatment
- Never-taker: $D_{1i} = D_{0i} = 0$ people who always do not take treatment
- No defiers!

IV with Heterogeneous Treatment Effect: Settings

- Assumption 4: Monotonicity

$$\forall i, D_{1i} - D_{0i} \geq 0 \quad \text{or vice versa}$$

- For everyone, instrument changes treatment in the same direction (or no change)
- Example: For a person who will serve (voluntarily) even when his number is above the cutoff, he will of course serve if his number is below the cutoff
- Complier: $D_{1i} > D_{0i}$ people who change their choice by instrument
- Always-taker: $D_{1i} = D_{0i} = 1$ people who always take treatment
- Never-taker: $D_{1i} = D_{0i} = 0$ people who always do not take treatment
- No defiers!

IV with Heterogeneous Treatment Effect: Settings

- Assumption 4: Monotonicity

$$\forall i, D_{1i} - D_{0i} \geq 0 \quad \text{or vice versa}$$

- For everyone, instrument changes treatment in the same direction (or no change)
- Example: For a person who will serve (voluntarily) even when his number is above the cutoff, he will of course serve if his number is below the cutoff
- Complier: $D_{1i} > D_{0i}$ people who change their choice by instrument
- Always-taker: $D_{1i} = D_{0i} = 1$ people who always take treatment
- Never-taker: $D_{1i} = D_{0i} = 0$ people who always do not take treatment
- No defiers!

IV with Heterogeneous Treatment Effect: LATE

- Intention-to-treat: $E[Y_i|z_i = 1] - E[Y_i|z_i = 0]$
- Local Average Treatment Effect (LATE)

LATE Theorem 4.4.1 in Angrist and Pischke (2008) MHE

If we have Assumption 1-4, then

$$\frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{E[D_i|z_i = 1] - E[D_i|z_i = 0]} = E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]$$

IV (Wald) identifies the average treatment effect for the complier group.

IV with Heterogeneous Treatment Effect: LATE

- Intention-to-treat: $E[Y_i|z_i = 1] - E[Y_i|z_i = 0]$
- Local Average Treatment Effect (LATE)

LATE Theorem 4.4.1 in Angrist and Pischke (2009) MHE

If we have Assumption 1-4, then

$$\frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{E[D_i|z_i = 1] - E[D_i|z_i = 0]} = E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]$$

IV (Wald) identifies the average treatment effect for the complier group.

IV with Heterogeneous Treatment Effect: LATE

- Intention-to-treat: $E[Y_i|z_i = 1] - E[Y_i|z_i = 0]$
- Local Average Treatment Effect (LATE)

LATE Theorem 4.4.1 in Angrist and Pischke (2009) MHE

If we have Assumption 1-4, then

$$\frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{E[D_i|z_i = 1] - E[D_i|z_i = 0]} = E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]$$

IV (Wald) identifies the average treatment effect for the complier group.

IV with Heterogeneous Treatment Effect: LATE

- Intention-to-treat: $E[Y_i|z_i = 1] - E[Y_i|z_i = 0]$
- Local Average Treatment Effect (LATE)

LATE Theorem 4.4.1 in Angrist and Pischke (2009) MHE

If we have Assumption 1-4, then

$$\frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{E[D_i|z_i = 1] - E[D_i|z_i = 0]} = E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]$$

IV (Wald) identifies the average treatment effect for the complier group.

IV with Heterogeneous Treatment Effect: LATE

Proof: Let's denote A as always-taker, C as complier, N as never-taker. We decompose ITT as follows.

$$E(Y_i|z_i=1)-E(Y_i|z_i=0)=$$
$$P(A_i|z_i=1)E(Y_{1i}|A_i,z_i=1)+P(C_i|z_i=1)E(Y_{1i}|C_i,z_i=1)+P(N_i|z_i=1)E(Y_{0i}|N_i,z_i=1)$$
$$-[P(A_i|z_i=0)E(Y_{1i}|A_i,z_i=0)+P(C_i|z_i=0)E(Y_{0i}|C_i,z_i=0)+P(N_i|z_i=0)E(Y_{0i}|N_i,z_i=0)]$$

As we know z_i is randomly assigned, it is independent of compliance type and potential outcome. Thus, we can cancel out red (A) and green (N) terms and leave only the blue term (C):

$$E(Y_i|z_i=1)-E(Y_i|z_i=0)=P(C_i|z_i=1)E(Y_{1i}|C_i,z_i=1)-P(C_i|z_i=0)E(Y_{0i}|C_i,z_i=0)$$
$$\Rightarrow E[Y_{0i}|C_i]=\frac{E[Y_i|z_i=1]-E[Y_i|z_i=0]}{P[C_i]}$$
$$\Rightarrow E[Y_{1i}-Y_{0i}|\underbrace{D_{1i}>D_{0i}}_{\text{This is complier}}]=\underbrace{\frac{E[Y_i|z_i=1]-E[Y_i|z_i=0]}{E[D_i|z_i=1]-E[D_i|z_i=0]}}_{\text{This is the fraction of complier, } P[C]}$$

IV with Heterogeneous Treatment Effect: LATE

Proof: Let's denote A as always-taker, C as complier, N as never-taker. We decompose ITT as follows.

$$E(Y_i|z_i=1)-E(Y_i|z_i=0)=$$
$$P(A_i|z_i=1)E(Y_{1i}|A_i,z_i=1)+P(C_i|z_i=1)E(Y_{1i}|C_i,z_i=1)+P(N_i|z_i=1)E(Y_{0i}|N_i,z_i=1)$$
$$-[P(A_i|z_i=0)E(Y_{1i}|A_i,z_i=0)+P(C_i|z_i=0)E(Y_{0i}|C_i,z_i=0)+P(N_i|z_i=0)E(Y_{0i}|N_i,z_i=0)]$$

As we know z_i is randomly assigned, it is independent of compliance type and potential outcome. Thus, we can cancel out red (A) and green (N) terms and leave only the blue term (C):

$$E(Y_i|z_i=1)-E(Y_i|z_i=0)=P(C_i|z_i=1)E(Y_{1i}|C_i,z_i=1)-P(C_i|z_i=0)E(Y_{0i}|C_i,z_i=0)$$
$$\Rightarrow E[Y_{0i}|C_i]=\frac{E[Y_i|z_i=1]-E[Y_i|z_i=0]}{P[C_i]}$$
$$\Rightarrow E[Y_{1i}-Y_{0i}|\underbrace{D_{1i}>D_{0i}}_{\text{This is complier}}]=\underbrace{\frac{E[Y_i|z_i=1]-E[Y_i|z_i=0]}{E[D_i|z_i=1]-E[D_i|z_i=0]}}_{\text{This is the fraction of complier, } P[C]}$$

IV with Heterogeneous Treatment Effect: LATE

Proof: Let's denote A as always-taker, C as complier, N as never-taker. We decompose ITT as follows.

$$E(Y_i|z_i = 1) - E(Y_i|z_i = 0) =$$
$$P(A_i|z_i = 1)E(Y_{1i}|A_i, z_i = 1) + P(C_i|z_i = 1)E(Y_{1i}|C_i, z_i = 1) + P(N_i|z_i = 1)E(Y_{0i}|N_i, z_i = 1)$$
$$- [P(A_i|z_i = 0)E(Y_{1i}|A_i, z_i = 0) + P(C_i|z_i = 0)E(Y_{0i}|C_i, z_i = 0) + P(N_i|z_i = 0)E(Y_{0i}|N_i, z_i = 0)]$$

As we know z_i is randomly assigned, it is independent of compliance type and potential outcome. Thus, we can cancel out red (A) and green (N) terms and leave only the blue term (C):

$$E(Y_i|z_i = 1) - E(Y_i|z_i = 0) = P(C_i|z_i = 1)E(Y_{1i}|C_i, z_i = 1) - P(C_i|z_i = 0)E(Y_{0i}|C_i, z_i = 0)$$

$$\Rightarrow E[Y_{0i}|C_i] = \frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{P[C_i]}$$

$$\Rightarrow E[Y_{1i} - Y_{0i}| \underbrace{D_{1i} > D_{0i}}_{\text{This is complier}}] = \underbrace{\frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{E[D_i|z_i = 1] - E[D_i|z_i = 0]}}_{\text{This is the fraction of complier, } P[C]}$$

IV with Heterogeneous Treatment Effect: LATE

Proof: Let's denote A as always-taker, C as complier, N as never-taker. We decompose ITT as follows.

$$E(Y_i|z_i = 1) - E(Y_i|z_i = 0) =$$
$$P(A_i|z_i = 1)E(Y_{1i}|A_i, z_i = 1) + P(C_i|z_i = 1)E(Y_{1i}|C_i, z_i = 1) + P(N_i|z_i = 1)E(Y_{0i}|N_i, z_i = 1)$$
$$-[P(A_i|z_i = 0)E(Y_{1i}|A_i, z_i = 0) + P(C_i|z_i = 0)E(Y_{0i}|C_i, z_i = 0) + P(N_i|z_i = 0)E(Y_{0i}|N_i, z_i = 0)]$$

As we know z_i is randomly assigned, it is independent of compliance type and potential outcome. Thus, we can cancel out red (A) and green (N) terms and leave only the blue term (C):

$$E(Y_i|z_i = 1) - E(Y_i|z_i = 0) = P(C_i|z_i = 1)E(Y_{1i}|C_i, z_i = 1) - P(C_i|z_i = 0)E(Y_{0i}|C_i, z_i = 0)$$

$$\Rightarrow E[Y_{0i}|C_i] = \frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{P[C_i]}$$

$$\Rightarrow E[Y_{1i} - Y_{0i}|\underbrace{D_{1i} > D_{0i}}_{\text{This is complier}}] = \underbrace{\frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{E[D_i|z_i = 1] - E[D_i|z_i = 0]}}_{\text{This is the fraction of complier, } P[C]}$$

IV with Heterogeneous Treatment Effect: LATE

Proof: Let's denote A as always-taker, C as complier, N as never-taker. We decompose ITT as follows.

$$E(Y_i|z_i = 1) - E(Y_i|z_i = 0) =$$
$$P(A_i|z_i = 1)E(Y_{1i}|A_i, z_i = 1) + P(C_i|z_i = 1)E(Y_{1i}|C_i, z_i = 1) + P(N_i|z_i = 1)E(Y_{0i}|N_i, z_i = 1)$$
$$- [P(A_i|z_i = 0)E(Y_{1i}|A_i, z_i = 0) + P(C_i|z_i = 0)E(Y_{0i}|C_i, z_i = 0) + P(N_i|z_i = 0)E(Y_{0i}|N_i, z_i = 0)]$$

As we know z_i is randomly assigned, it is independent of compliance type and potential outcome. Thus, we can cancel out red (A) and green (N) terms and leave only the blue term (C):

$$E(Y_i|z_i = 1) - E(Y_i|z_i = 0) = P(C_i|z_i = 1)E(Y_{1i}|C_i, z_i = 1) - P(C_i|z_i = 0)E(Y_{0i}|C_i, z_i = 0)$$
$$\Rightarrow E[Y_{0i}|C_i] = \frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{P[C_i]}$$
$$\Rightarrow E[Y_{1i} - Y_{0i}| \underbrace{D_{1i} > D_{0i}}_{\text{This is complier}}] = \underbrace{\frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{E[D_i|z_i = 1] - E[D_i|z_i = 0]}}_{\text{This is the fraction of complier, } P[C]}$$

IV with Heterogeneous Treatment Effect: LATE

- LATE represents an average TE for a special group: compliers
- Monotonicity is important: No room for defiers
- If there are defiers, effects from compliers could be contaminated by effects from defiers
- LATE is internally valid
- Complier group can be policy relevant: Those whose behaviors CAN be changed by the policy instrument

IV with Heterogeneous Treatment Effect: LATE

- LATE represents an average TE for a special group: compliers
- Monotonicity is important: No room for defiers
- If there are defiers, effects from compliers could be contaminated by effects from defiers
- LATE is internally valid
- Complier group can be policy relevant: Those whose behaviors CAN be changed by the policy instrument

IV with Heterogeneous Treatment Effect: LATE

- LATE represents an average TE for a special group: compliers
- Monotonicity is important: No room for defiers
- If there are defiers, effects from compliers could be contaminated by effects from defiers
- LATE is internally valid
- Complier group can be policy relevant: Those whose behaviors CAN be changed by the policy instrument

IV with Heterogeneous Treatment Effect: LATE

- LATE represents an average TE for a special group: compliers
- Monotonicity is important: No room for defiers
- If there are defiers, effects from compliers could be contaminated by effects from defiers
- LATE is internally valid
- Complier group can be policy relevant: Those whose behaviors CAN be changed by the policy instrument

IV with Heterogeneous Treatment Effect: LATE

- LATE represents an average TE for a special group: compliers
- Monotonicity is important: No room for defiers
- If there are defiers, effects from compliers could be contaminated by effects from defiers
- LATE is internally valid
- Complier group can be policy relevant: Those whose behaviors CAN be changed by the policy instrument

IV with Heterogeneous Treatment Effect: LATE

- LATE represents an average TE for a special group: compliers
- Monotonicity is important: No room for defiers
- If there are defiers, effects from compliers could be contaminated by effects from defiers
- LATE is internally valid
- Complier group can be policy relevant: Those whose behaviors CAN be changed by the policy instrument

IV with Heterogeneous Treatment Effect: LATE

What are the weaknesses of LATE interpretation?

- LATE is not externally valid, since the complier group changes when policy is changed
- When instrument and treatment become multi-valued, interpreting IV in a traditional way becomes very very hard
- Why? The number of types increase exponentially! Much faster than your available equations
- Still remember Pinto (2015)?
- We need new weapons for this: IV + Choice Model (next lecture)

IV with Heterogeneous Treatment Effect: LATE

What are the weaknesses of LATE interpretation?

- LATE is not externally valid, since the complier group changes when policy is changed
- When instrument and treatment become multi-valued, interpreting IV in a traditional way becomes very very hard
- Why? The number of types increase exponentially! Much faster than your available equations
- Still remember Pinto (2015)?
- We need new weapons for this: IV + Choice Model (next lecture)

IV with Heterogeneous Treatment Effect: LATE

What are the weaknesses of LATE interpretation?

- LATE is not externally valid, since the complier group changes when policy is changed
- When instrument and treatment become multi-valued, interpreting IV in a traditional way becomes very very hard
- Why? The number of types increase exponentially! Much faster than your available equations
- Still remember Pinto (2015)?
- We need new weapons for this: IV + Choice Model (next lecture)

IV with Heterogeneous Treatment Effect: LATE

What are the weaknesses of LATE interpretation?

- LATE is not externally valid, since the complier group changes when policy is changed
- When instrument and treatment become multi-valued, interpreting IV in a traditional way becomes very very hard
- Why? The number of types increase exponentially! Much faster than your available equations
- Still remember Pinto (2015)?
- We need new weapons for this: IV + Choice Model (next lecture)

IV with Heterogeneous Treatment Effect: LATE

What are the weaknesses of LATE interpretation?

- LATE is not externally valid, since the complier group changes when policy is changed
- When instrument and treatment become multi-valued, interpreting IV in a traditional way becomes very very hard
- Why? The number of types increase exponentially! Much faster than your available equations
- Still remember Pinto (2015)?
- We need new weapons for this: IV + Choice Model (next lecture)

IV with Heterogeneous Treatment Effect: LATE

What are the weaknesses of LATE interpretation?

- LATE is not externally valid, since the complier group changes when policy is changed
- When instrument and treatment become multi-valued, interpreting IV in a traditional way becomes very very hard
- Why? The number of types increase exponentially! Much faster than your available equations
- Still remember Pinto (2015)?
- We need new weapons for this: IV + Choice Model (next lecture)

IV with Heterogeneous Treatment Effect: LATE

What are the weaknesses of LATE interpretation?

- LATE is not externally valid, since the complier group changes when policy is changed
- When instrument and treatment become multi-valued, interpreting IV in a traditional way becomes very very hard
- Why? The number of types increase exponentially! Much faster than your available equations
- Still remember Pinto (2015)?
- We need new weapons for this: IV + Choice Model (next lecture)

Multiple IV: GMM Framework

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- We just investigated the case when (3) is relaxed
- Now we relax (1) and (2), considering multiple endogenous variables and IV
- We can discuss this general question in the GMM framework
- All common IV related estimators (Wald, 2SLS...) are special cases of GMM estimator

Multiple IV: GMM Framework

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- We just investigated the case when (3) is relaxed
- Now we relax (1) and (2), considering multiple endogenous variables and IV
- We can discuss this general question in the GMM framework
- All common IV related estimators (Wald, 2SLS...) are special cases of GMM estimator

Multiple IV: GMM Framework

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- We just investigated the case when (3) is relaxed
- Now we relax (1) and (2), considering multiple endogenous variables and IV
- We can discuss this general question in the GMM framework
- All common IV related estimators (Wald, 2SLS...) are special cases of GMM estimator

Multiple IV: GMM Framework

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- We just investigated the case when (3) is relaxed
- Now we relax (1) and (2), considering multiple endogenous variables and IV
- We can discuss this general question in the GMM framework
- All common IV related estimators (Wald, 2SLS...) are special cases of GMM estimator

Multiple IV: GMM Framework

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- We just investigated the case when (3) is relaxed
- Now we relax (1) and (2), considering multiple endogenous variables and IV
- We can discuss this general question in the GMM framework
- All common IV related estimators (Wald, 2SLS...) are special cases of GMM estimator

Multiple IV: GMM Framework

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- We just investigated the case when (3) is relaxed
- Now we relax (1) and (2), considering multiple endogenous variables and IV
- We can discuss this general question in the GMM framework
- All common IV related estimators (Wald, 2SLS...) are special cases of GMM estimator

Multiple IV: GMM Definition

- Let $g_i(\beta)$ be a known $l \times 1$ function of $k \times 1$ parameter β
- Definition: A moment equation model is

$$E[g_i(\beta)] = 0$$

- In this system we have l known equations and k unknown parameters
- Example: Linear regression model is a moment equation model with $l = k$ and $g_i(\beta) = x_i(Y_i - x_i'\beta)$
- If $l = k$, just-identified; if $l > k$, over-identified; if $l < k$, under-identified

Multiple IV: GMM Definition

- Let $g_i(\beta)$ be a known $l \times 1$ function of $k \times 1$ parameter β
- Definition: A moment equation model is

$$E[g_i(\beta)] = 0$$

- In this system we have l known equations and k unknown parameters
- Example: Linear regression model is a moment equation model with $l = k$ and $g_i(\beta) = x_i(Y_i - x_i'\beta)$
- If $l = k$, just-identified; if $l > k$, over-identified; if $l < k$, under-identified

Multiple IV: GMM Definition

- Let $g_i(\beta)$ be a known $l \times 1$ function of $k \times 1$ parameter β
- Definition: A moment equation model is

$$E[g_i(\beta)] = 0$$

- In this system we have l known equations and k unknown parameters
- Example: Linear regression model is a moment equation model with $l = k$ and $g_i(\beta) = x_i(Y_i - x_i'\beta)$
- If $l = k$, just-identified; if $l > k$, over-identified; if $l < k$, under-identified

Multiple IV: GMM Definition

- Let $g_i(\beta)$ be a known $l \times 1$ function of $k \times 1$ parameter β
- Definition: A moment equation model is

$$E[g_i(\beta)] = 0$$

- In this system we have l known equations and k unknown parameters
- Example: Linear regression model is a moment equation model with $l = k$ and $g_i(\beta) = x_i(Y_i - x_i^l \beta)$
- If $l = k$, just-identified; if $l > k$, over-identified; if $l < k$, under-identified

Multiple IV: GMM Definition

- Let $g_i(\beta)$ be a known $l \times 1$ function of $k \times 1$ parameter β
- Definition: A moment equation model is

$$E[g_i(\beta)] = 0$$

- In this system we have l known equations and k unknown parameters
- Example: Linear regression model is a moment equation model with $l = k$ and $g_i(\beta) = x_i(Y_i - x_i'\beta)$
- If $l = k$, just-identified; if $l > k$, over-identified; if $l < k$, under-identified

Multiple IV: GMM Definition

- Let $g_i(\beta)$ be a known $l \times 1$ function of $k \times 1$ parameter β
- Definition: A moment equation model is

$$E[g_i(\beta)] = 0$$

- In this system we have l known equations and k unknown parameters
- Example: Linear regression model is a moment equation model with $l = k$ and $g_i(\beta) = x_i(Y_i - x_i^{\prime}\beta)$
- If $l = k$, just-identified; if $l > k$, over-identified; if $l < k$, under-identified

Multiple IV: GMM Definition

- Given $E[g_i(\beta)] = 0$, how to use data to estimate β?
- Simple and straightforward when $l = k$ (just-identified) $\Longrightarrow$ Using sample means
- Method of Moments Estimator (MME):

$$\bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\beta}) = 0$$

- Example: OLS estimator is also a MME

$$\bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} x_i(Y_i - x_i'\hat{\beta}) = 0$$

Multiple IV: GMM Definition

- Given $E[g_i(\beta)] = 0$, how to use data to estimate β?
- Simple and straightforward when $l = k$ (just-identified) $\Longrightarrow$ Using sample means
- Method of Moments Estimator (MME):

$$\bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\beta}) = 0$$

- Example: OLS estimator is also a MME

$$\bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} x_i(Y_i - x_i'\hat{\beta}) = 0$$

Multiple IV: GMM Definition

- Given $E[g_i(\beta)] = 0$, how to use data to estimate β?
- Simple and straightforward when $l = k$ (just-identified) $\Longrightarrow$ Using sample means
- Method of Moments Estimator (MME):

$$\bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\beta}) = 0$$

- Example: OLS estimator is also a MME

$$\bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} x_i(Y_i - x_i'\hat{\beta}) = 0$$

Multiple IV: GMM Definition

- Given $E[g_i(\beta)] = 0$, how to use data to estimate β?
- Simple and straightforward when $l = k$ (just-identified) $\Longrightarrow$ Using sample means
- Method of Moments Estimator (MME):

$$\bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\beta}) = 0$$

- Example: OLS estimator is also a MME

$$\bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} x_i(Y_i - x_i'\hat{\beta}) = 0$$

Multiple IV: GMM Definition

- Given $E[g_i(\beta)] = 0$, how to use data to estimate β?
- Simple and straightforward when $l = k$ (just-identified) $\Longrightarrow$ Using sample means
- Method of Moments Estimator (MME):

$$\bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\beta}) = 0$$

- Example: OLS estimator is also a MME

$$\bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} x_i(Y_i - x_i'\hat{\beta}) = 0$$

Multiple IV: GMM Definition

- What if $l > k$? (over-identified)
- Now we have more equations than unknowns
- We cannot directly equate sample mean to zero and solve for β
- Our target then becomes to minimize the distance between the moment vector and zero

$$J(\beta) = n\bar{g}_n(\beta)' W \bar{g}_n(\beta)$$
$$\hat{\beta}_{gmm} = argmin_\beta J(\beta)$$

- W is some weighting matrix
- J measures the square of weighted euclidean distance between $\bar{g}_n$ and 0
- MME (thus OLS) is a special case of GMM when $l = k$

Multiple IV: GMM Definition

- What if $l > k$? (over-identified)
- Now we have more equations than unknowns
- We cannot directly equate sample mean to zero and solve for β
- Our target then becomes to minimize the distance between the moment vector and zero

$$J(\beta) = n\bar{g}_n(\beta)' W \bar{g}_n(\beta)$$
$$\hat{\beta}_{gmm} = argmin_{\beta} J(\beta)$$

- W is some weighting matrix
- J measures the square of weighted euclidean distance between $\bar{g}_n$ and 0
- MME (thus OLS) is a special case of GMM when $l = k$

Multiple IV: GMM Definition

- What if $l > k$? (over-identified)
- Now we have more equations than unknowns
- We cannot directly equate sample mean to zero and solve for β
- Our target then becomes to minimize the distance between the moment vector and zero

$$J(\beta) = n\bar{g}_n(\beta)' W \bar{g}_n(\beta)$$
$$\hat{\beta}_{gmm} = argmin_{\beta} J(\beta)$$

- W is some weighting matrix
- J measures the square of weighted euclidean distance between $\bar{g}_n$ and 0
- MME (thus OLS) is a special case of GMM when $l = k$

Multiple IV: GMM Definition

- What if $l > k$? (over-identified)
- Now we have more equations than unknowns
- We cannot directly equate sample mean to zero and solve for β
- Our target then becomes to minimize the distance between the moment vector and zero

$$J(\beta) = n\bar{g}_n(\beta)' W \bar{g}_n(\beta)$$
$$\hat{\beta}_{gmm} = argmin_{\beta} J(\beta)$$

- W is some weighting matrix
- J measures the square of weighted euclidean distance between $\bar{g}_n$ and 0
- MME (thus OLS) is a special case of GMM when $l = k$

Multiple IV: GMM Definition

- What if $l > k$? (over-identified)
- Now we have more equations than unknowns
- We cannot directly equate sample mean to zero and solve for β
- Our target then becomes to minimize the distance between the moment vector and zero

$$J(\beta) = n\bar{g}_n(\beta)' W \bar{g}_n(\beta)$$
$$\hat{\beta}_{gmm} = argmin_{\beta} J(\beta)$$

- W is some weighting matrix
- J measures the square of weighted euclidean distance between $\bar{g}_n$ and 0
- MME (thus OLS) is a special case of GMM when $l = k$

Multiple IV: GMM Definition

- What if $l > k$? (over-identified)
- Now we have more equations than unknowns
- We cannot directly equate sample mean to zero and solve for β
- Our target then becomes to minimize the distance between the moment vector and zero

$$J(\beta) = n\bar{g}_n(\beta)' W \bar{g}_n(\beta)$$

$$\hat{\beta}_{gmm} = argmin_{\beta} J(\beta)$$

- W is some weighting matrix
- J measures the square of weighted euclidean distance between $\bar{g}_n$ and 0
- MME (thus OLS) is a special case of GMM when $l = k$

Multiple IV: GMM Definition

- What if $l > k$? (over-identified)
- Now we have more equations than unknowns
- We cannot directly equate sample mean to zero and solve for β
- Our target then becomes to minimize the distance between the moment vector and zero

$$J(\beta) = n\bar{g}_n(\beta)' W \bar{g}_n(\beta)$$
$$\hat{\beta}_{gmm} = argmin_\beta J(\beta)$$

- W is some weighting matrix
- J measures the square of weighted euclidean distance between $\bar{g}_n$ and 0
- MME (thus OLS) is a special case of GMM when $l = k$

Multiple IV: GMM Definition

- What if $l > k$? (over-identified)
- Now we have more equations than unknowns
- We cannot directly equate sample mean to zero and solve for β
- Our target then becomes to minimize the distance between the moment vector and zero

$$J(\beta) = n\bar{g}_n(\beta)' W \bar{g}_n(\beta)$$
$$\hat{\beta}_{gmm} = argmin_{\beta} J(\beta)$$

- W is some weighting matrix
- J measures the square of weighted euclidean distance between $\bar{g}_n$ and 0
- MME (thus OLS) is a special case of GMM when $l = k$

Multiple IV: Linear GMM

- Let X_i be the endogenous variables, Z_i be the instruments
- Instruments are not correlated with the error, so we have the linear moment equations:

$$E[g_i(\beta)] = E[Z_i(Y_i - X_i'\beta)] = 0 \tag{7}$$

- Stack over the sample, we have GMM estimator to be:

$$\hat{\beta}_{gmm} = argmin_{\beta} \underbrace{n(Z'Y - Z'X\beta)'W(Z'Y - Z'X\beta)}_{J(\beta)}$$

Multiple IV: Linear GMM

- Let X_i be the endogenous variables, Z_i be the instruments
- Instruments are not correlated with the error, so we have the linear moment equations:

$$E[g_i(\beta)] = E[Z_i(Y_i - X_i'\beta)] = 0 \tag{7}$$

- Stack over the sample, we have GMM estimator to be:

$$\hat{\beta}_{gmm} = argmin_{\beta} \underbrace{n(Z'Y - Z'X\beta)'W(Z'Y - Z'X\beta)}_{J(\beta)}$$

Multiple IV: Linear GMM

- Let X_i be the endogenous variables, Z_i be the instruments
- Instruments are not correlated with the error, so we have the linear moment equations:

$$E[g_i(\beta)] = E[Z_i(Y_i - X_i'\beta)] = 0 \tag{7}$$

- Stack over the sample, we have GMM estimator to be:

$$\hat{\beta}_{gmm} = argmin_{\beta} \underbrace{n(Z'Y - Z'X\beta)'W(Z'Y - Z'X\beta)}_{J(\beta)}$$

Multiple IV: Linear GMM

- Let X_i be the endogenous variables, Z_i be the instruments
- Instruments are not correlated with the error, so we have the linear moment equations:

$$E[g_i(\beta)] = E[Z_i(Y_i - X_i'\beta)] = 0 \tag{7}$$

- Stack over the sample, we have GMM estimator to be:

$$\hat{\beta}_{gmm} = argmin_\beta \underbrace{n(Z'Y - Z'X\beta)'W(Z'Y - Z'X\beta)}_{J(\beta)}$$

Multiple IV: Linear GMM

- Solve this minimization problem, we have

Theorem 13.1 in Hansen (2022)

For the over-identified linear IV model with l endogenous variables and k instruments

$$\hat{\beta}_{gmm} = (X'ZWZ'X)^{-1}(X'ZWZ'Y)$$

- GMM is really general
- Many estimators are special cases of GMM estimator

Multiple IV: Linear GMM

- Solve this minimization problem, we have

Theorem 13.1 in Hansen (2022)

For the over-identified linear IV model with l endogenous variables and k instruments

$$\hat{\beta}_{gmm} = (X'ZWZ'X)^{-1}(X'ZWZ'Y)$$

- GMM is really general
- Many estimators are special cases of GMM estimator

Multiple IV: Linear GMM

- Solve this minimization problem, we have

Theorem 13.1 in Hansen (2022)

For the over-identified linear IV model with l endogenous variables and k instruments

$$\hat{\beta}_{gmm} = (X'ZWZ'X)^{-1}(X'ZWZ'Y)$$

- GMM is really general
- Many estimators are special cases of GMM estimator

Multiple IV: Linear GMM

- Solve this minimization problem, we have

Theorem 13.1 in Hansen (2022)

For the over-identified linear IV model with l endogenous variables and k instruments

$$\hat{\beta}_{gmm} = (X'ZWZ'X)^{-1}(X'ZWZ'Y)$$

- GMM is really general
- Many estimators are special cases of GMM estimator

Multiple IV: Linear GMM

- Solve this minimization problem, we have

Theorem 13.1 in Hansen (2022)

For the over-identified linear IV model with l endogenous variables and k instruments

$$\hat{\beta}_{gmm} = (X'ZWZ'X)^{-1}(X'ZWZ'Y)$$

- GMM is really general
- Many estimators are special cases of GMM estimator

Multiple IV: Linear GMM

- When $X = Z$, $W = I$, we have:

$$\begin{aligned}\hat{\beta}_{gmm} &= (X'XIX'X)^{-1}(X'XIX'Y) \\ &= (X'X)^{-1}(X'X)^{-1}(X'X)Y \\ &= (X'X)^{-1}X'Y = \hat{\beta}_{ols}\end{aligned}$$

- We have the second line since $X = Z$, $X'X$ is a square matrix
- When we do not have endogenous variables, and use identity weighting matrix, GMM is OLS.

Multiple IV: Linear GMM

- When $X = Z, W = I$, we have:

$$\begin{aligned}\hat{\beta}_{gmm} &= (X'XIX'X)^{-1}(X'XIX'Y) \\ &= (X'X)^{-1}(X'X)^{-1}(X'X)Y \\ &= (X'X)^{-1}X'Y = \hat{\beta}_{ols}\end{aligned}$$

- We have the second line since $X = Z$, $X'X$ is a square matrix
- When we do not have endogenous variables, and use identity weighting matrix, GMM is OLS.

Multiple IV: Linear GMM

- When $X = Z, W = I$, we have:

$$\begin{aligned}\hat{\beta}_{gmm} &= (X'XIX'X)^{-1}(X'XIX'Y) \\ &= (X'X)^{-1}(X'X)^{-1}(X'X)Y \\ &= (X'X)^{-1}X'Y = \hat{\beta}_{ols}\end{aligned}$$

- We have the second line since $X = Z$, $X'X$ is a square matrix
- When we do not have endogenous variables, and use identity weighting matrix, GMM is OLS.

Multiple IV: Linear GMM

- When $X = Z, W = I$, we have:

$$\begin{aligned}\hat{\beta}_{gmm} &= (X'XIX'X)^{-1}(X'XIX'Y) \\ &= (X'X)^{-1}(X'X)^{-1}(X'X)Y \\ &= (X'X)^{-1}X'Y = \hat{\beta}_{ols}\end{aligned}$$

- We have the second line since $X = Z$, $X'X$ is a square matrix
- When we do not have endogenous variables, and use identity weighting matrix, GMM is OLS.

Multiple IV: Linear GMM

- When $l = k$, $W = I$, we have:

$$\begin{aligned}\hat{\beta}_{gmm} &= (X'ZIZ'X)^{-1}(X'ZIZ'Y)\\ &= (Z'X)^{-1}(X'Z)^{-1}(X'Z)(Z'Y)\\ &= (Z'X)^{-1}Z'Y\end{aligned}$$

- We have the second line since $l = k$, $X'Z$ is a square matrix
- This is the Wald/IV estimator.

Multiple IV: Linear GMM

- When $l = k$, $W = I$, we have:

$$\begin{aligned}\hat{\beta}_{gmm} &= (X'ZIZ'X)^{-1}(X'ZIZ'Y) \\ &= (Z'X)^{-1}(X'Z)^{-1}(X'Z)(Z'Y) \\ &= (Z'X)^{-1}Z'Y\end{aligned}$$

- We have the second line since $l = k$, $X'Z$ is a square matrix
- This is the Wald/IV estimator.

Multiple IV: Linear GMM

- When $l = k$, $W = I$, we have:

$$\begin{aligned}\hat{\beta}_{gmm} &= (X'ZIZ'X)^{-1}(X'ZIZ'Y) \\ &= (Z'X)^{-1}(X'Z)^{-1}(X'Z)(Z'Y) \\ &= (Z'X)^{-1}Z'Y\end{aligned}$$

- We have the second line since $l = k$, $X'Z$ is a square matrix
- This is the Wald/IV estimator.

Multiple IV: Linear GMM

- When $l = k$, $W = I$, we have:

$$\begin{aligned}\hat{\beta}_{gmm} &= (X'ZIZ'X)^{-1}(X'ZIZ'Y) \\ &= (Z'X)^{-1}(X'Z)^{-1}(X'Z)(Z'Y) \\ &= (Z'X)^{-1}Z'Y\end{aligned}$$

- We have the second line since $l = k$, $X'Z$ is a square matrix
- This is the Wald/IV estimator.

Multiple IV: Linear GMM

- Let $P_z = Z(Z'Z)^{-1}Z'$ to be the projection matrix
- First stage fitted value then becomes $\hat{X} = P_z X$
- P is idempotent: $P \cdot P = P$
- When $\hat{W} = (Z'Z)^{-1}$, we have:

$$\begin{aligned}\hat{\beta}_{gmm} &= (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z'Y) \\ &= (X'P_zX)^{-1}X'P_zY \\ &= (\hat{X}'\hat{X})^{-1}\hat{X}'Y\end{aligned}$$

- This is the 2SLS estimator.

Multiple IV: Linear GMM

- Let $P_z = Z(Z'Z)^{-1}Z'$ to be the projection matrix
- First stage fitted value then becomes $\hat{X} = P_z X$
- P is idempotent: $P \cdot P = P$
- When $\hat{W} = (Z'Z)^{-1}$, we have:

$$\begin{aligned}\hat{\beta}_{gmm} &= (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z'Y) \\ &= (X'P_zX)^{-1}X'P_zY \\ &= (\hat{X}'\hat{X})^{-1}\hat{X}'Y\end{aligned}$$

- This is the 2SLS estimator.

Multiple IV: Linear GMM

- Let $P_z = Z(Z'Z)^{-1}Z'$ to be the projection matrix
- First stage fitted value then becomes $\hat{X} = P_z X$
- P is idempotent: $P \cdot P = P$
- When $\hat{W} = (Z'Z)^{-1}$, we have:

$$\begin{aligned}\hat{\beta}_{gmm} &= (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z'Y) \\ &= (X'P_zX)^{-1}X'P_zY \\ &= (\hat{X}'\hat{X})^{-1}\hat{X}'Y\end{aligned}$$

- This is the 2SLS estimator.

Multiple IV: Linear GMM

- Let $P_z = Z(Z'Z)^{-1}Z'$ to be the projection matrix
- First stage fitted value then becomes $\hat{X} = P_z X$
- P is idempotent: $P \cdot P = P$
- When $\hat{W} = (Z'Z)^{-1}$, we have:

$$\begin{aligned}\hat{\beta}_{gmm} &= (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z'Y) \\ &= (X'P_zX)^{-1}X'P_zY \\ &= (\hat{X}'\hat{X})^{-1}\hat{X}'Y\end{aligned}$$

- This is the 2SLS estimator.

Multiple IV: Linear GMM

- Let $P_z = Z(Z'Z)^{-1}Z'$ to be the projection matrix
- First stage fitted value then becomes $\hat{X} = P_z X$
- P is idempotent: $P \cdot P = P$
- When $\hat{W} = (Z'Z)^{-1}$, we have:

$$\begin{aligned}\hat{\beta}_{gmm} &= (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z'Y) \\ &= (X'P_z X)^{-1}X'P_z Y \\ &= (\hat{X}'\hat{X})^{-1}\hat{X}'Y\end{aligned}$$

- This is the 2SLS estimator.

Multiple IV: Linear GMM

- Let $P_z = Z(Z'Z)^{-1}Z'$ to be the projection matrix
- First stage fitted value then becomes $\hat{X} = P_z X$
- P is idempotent: $P \cdot P = P$
- When $\hat{W} = (Z'Z)^{-1}$, we have:

$$
\begin{aligned}
\hat{\beta}_{gmm} &= (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z'Y) \\
&= (X'P_z X)^{-1}X'P_z Y \\
&= (\hat{X}'\hat{X})^{-1}\hat{X}'Y
\end{aligned}
$$

- This is the 2SLS estimator.

Multiple IV: Over-identification Test

- We can test whether moment conditions hold (IV is valid)
- Basic idea: If IV is valid, our calculated distance J should be close enough to zero

Hansen's test: Theorem 13.14 in Hansen (2022)

Under some mild assumptions, as $n \to \infty$,

$$J = J(\hat{\beta}_{gmm}) \xrightarrow{d} \chi^2_{l-k}$$

For c satisfying $\alpha = 1 - G_{l-k}(c)$, $P[J > c | H_0] \to \alpha$ so the test "Reject H_0 if $J > c$" has asymptotic size α.

Multiple IV: Over-identification Test

- We can test whether moment conditions hold (IV is valid)
- Basic idea: If IV is valid, our calculated distance J should be close enough to zero

Hansen's test Theorem 13.14 in Hansen (2022)

Under some mild assumptions, as $n \to \infty$,

$$J = J(\hat{\beta}_{gmm}) \xrightarrow{d} \chi^2_{l-k}$$

For c satisfying $\alpha = 1 - G_{l-k}(c)$, $P[J > c | H_0] \to \alpha$ so the test "Reject H_0 if $J > c$" has asymptotic size α.

Multiple IV: Over-identification Test

- We can test whether moment conditions hold (IV is valid)
- Basic idea: If IV is valid, our calculated distance J should be close enough to zero

Hansen's test Theorem 13.14 in Hansen (2022)

Under some mild assumptions, as $n \to \infty$,

$$J = J(\hat{\beta}_{gmm}) \xrightarrow{d} \chi^2_{l-k}$$

For c satisfying $\alpha = 1 - G_{l-k}(c)$, $P[J > c | H_0] \to \alpha$ so the test "Reject H_0 if $J > c$" has asymptotic size α.

Multiple IV: Over-identification Test

- We can test whether moment conditions hold (IV is valid)
- Basic idea: If IV is valid, our calculated distance J should be close enough to zero

Hansen's test Theorem 13.14 in Hansen (2022)

Under some mild assumptions, as $n \to \infty$,

$$J = J(\hat{\beta}_{gmm}) \xrightarrow{d} \chi^2_{l-k}$$

For c satisfying $\alpha = 1 - G_{l-k}(c)$, $P[J > c | H_0] \to \alpha$ so the test "Reject H_0 if $J > c$" has asymptotic size α.

Multiple IV: Over-identification Test

Be careful using this test!

- If you want to have a valid IV, you should hope J-statistic to be NOT significant
- This is feasible only when you have more instruments than endogenous variables
- J-test rejects null $\not\Rightarrow E(g_i) \neq 0$, since this is a specification test
  There can be other reasons why the null is rejected, such as non-linearity
- $E(g_i) \neq 0 \not\Rightarrow$ J-test rejects null
- Actual size in finite-sample is too large (too many rejections)

Multiple IV: Over-identification Test

Be careful using this test!

- If you want to have a valid IV, you should hope J-statistic to be NOT significant
- This is feasible only when you have more instruments than endogenous variables
- J-test rejects null $\nRightarrow E(g_i) \neq 0$, since this is a specification test
  There can be other reasons why the null is rejected, such as non-linearity
- $E(g_i) \neq 0 \nRightarrow$ J-test rejects null
- Actual size in finite-sample is too large (too many rejections)

Multiple IV: Over-identification Test

Be careful using this test!

- If you want to have a valid IV, you should hope J-statistic to be NOT significant
- This is feasible only when you have more instruments than endogenous variables
- J-test rejects null $\nRightarrow E(g_i) \neq 0$, since this is a specification test
  There can be other reasons why the null is rejected, such as non-linearity
- $E(g_i) \neq 0 \nRightarrow$ J-test rejects null
- Actual size in finite-sample is too large (too many rejections)

Multiple IV: Over-identification Test

Be careful using this test!

- If you want to have a valid IV, you should hope J-statistic to be NOT significant
- This is feasible only when you have more instruments than endogenous variables
- J-test rejects null $\nRightarrow E(g_i) \neq 0$, since this is a specification test
  There can be other reasons why the null is rejected, such as non-linearity
- $E(g_i) \neq 0 \nRightarrow$ J-test rejects null
- Actual size in finite-sample is too large (too many rejections)

Multiple IV: Over-identification Test

Be careful using this test!

- If you want to have a valid IV, you should hope J-statistic to be NOT significant
- This is feasible only when you have more instruments than endogenous variables
- J-test rejects null $\not\Rightarrow E(g_i) \neq 0$, since this is a specification test
  There can be other reasons why the null is rejected, such as non-linearity
- $E(g_i) \neq 0 \not\Rightarrow$ J-test rejects null
- Actual size in finite-sample is too large (too many rejections)

Multiple IV: Over-identification Test

Be careful using this test!

- If you want to have a valid IV, you should hope J-statistic to be NOT significant
- This is feasible only when you have more instruments than endogenous variables
- J-test rejects null $\nRightarrow E(g_i) \neq 0$, since this is a specification test
  There can be other reasons why the null is rejected, such as non-linearity
- $E(g_i) \neq 0 \nRightarrow$ J-test rejects null
- Actual size in finite-sample is too large (too many rejections)

Multiple IV: Over-identification Test

Be careful using this test!

- If you want to have a valid IV, you should hope J-statistic to be NOT significant
- This is feasible only when you have more instruments than endogenous variables
- J-test rejects null $\not\Rightarrow E(g_i) \neq 0$, since this is a specification test
  There can be other reasons why the null is rejected, such as non-linearity
- $E(g_i) \neq 0 \not\Rightarrow$ J-test rejects null
- Actual size in finite-sample is too large (too many rejections)

Oster Bound: Endogeneity without IV

- Coming up with a good IV is super hard
- Unfortunately, we often cannot find a valid instrument
- How to deal with endogeneity without a valid instrument?
- We are going to introduce one of the methods: Oster Bound
- Oster (2019) Unobservable Selection and Coefficient Stability: Theory and Evidence

Oster Bound: Endogeneity without IV

- Coming up with a good IV is super hard
- Unfortunately, we often cannot find a valid instrument
- How to deal with endogeneity without a valid instrument?
- We are going to introduce one of the methods: Oster Bound
- Oster (2019) Unobservable Selection and Coefficient Stability: Theory and Evidence

Oster Bound: Endogeneity without IV

- Coming up with a good IV is super hard
- Unfortunately, we often cannot find a valid instrument
- How to deal with endogeneity without a valid instrument?
- We are going to introduce one of the methods: Oster Bound
- Oster (2019) Unobservable Selection and Coefficient Stability: Theory and Evidence

Oster Bound: Endogeneity without IV

- Coming up with a good IV is super hard
- Unfortunately, we often cannot find a valid instrument
- How to deal with endogeneity without a valid instrument?
- We are going to introduce one of the methods: Oster Bound
- Oster (2019) Unobservable Selection and Coefficient Stability: Theory and Evidence

Oster Bound: Endogeneity without IV

- Coming up with a good IV is super hard
- Unfortunately, we often cannot find a valid instrument
- How to deal with endogeneity without a valid instrument?
- We are going to introduce one of the methods: Oster Bound
- Oster (2019) Unobservable Selection and Coefficient Stability: Theory and Evidence

Oster Bound: Endogeneity without IV

- Coming up with a good IV is super hard
- Unfortunately, we often cannot find a valid instrument
- How to deal with endogeneity without a valid instrument?
- We are going to introduce one of the methods: Oster Bound
- Oster (2019) Unobservable Selection and Coefficient Stability: Theory and Evidence

Oster Bound: Endogeneity without IV

- In general, point identification in this case is impossible
- This is not a method to help you in point identification/estimation
- But to help you bound your results ⟹ Set identification/bound estimation

Oster Bound: Endogeneity without IV

- In general, point identification in this case is impossible
- This is not a method to help you in point identification/estimation
- But to help you bound your results $\Longrightarrow$ Set identification/bound estimation

Oster Bound: Endogeneity without IV

- In general, point identification in this case is impossible
- This is not a method to help you in point identification/estimation
- But to help you bound your results $\Longrightarrow$ Set identification/bound estimation

Oster Bound: Endogeneity without IV

- In general, point identification in this case is impossible
- This is not a method to help you in point identification/estimation
- But to help you bound your results $\Rightarrow$ Set identification/bound estimation

Oster Bound: Endogeneity without IV

- Point identification means
  - You can recover the exact point of the parameter from the data
  - 1-1 mapping between data and parameter value
  - No other parameter values can generate the same data
  - You cannot find another parameter value that is observational equivalent
- Set identification means
  - You can recover a set of the parameter from the data
  - No other parameter values outside this identified set can generate the same data
  - You cannot find another parameter value outside the identified set that is observational equivalent

Oster Bound: Endogeneity without IV

- Point identification means
  - You can recover the exact point of the parameter from the data
  - 1-1 mapping between data and parameter value
  - No other parameter values can generate the same data
  - You cannot find another parameter value that is observational equivalent
- Set identification means
  - You can recover a set of the parameter from the data
  - No other parameter values outside this identified set can generate the same data
  - You cannot find another parameter value outside this identified set that is observational equivalent

Oster Bound: Endogeneity without IV

- Point identification means
  - You can recover the exact point of the parameter from the data
  - 1-1 mapping between data and parameter value
  - No other parameter values can generate the same data
  - You cannot find another parameter value that is observational equivalent
- Set identification means
  - You can recover a set of the parameter from the data
  - No other parameter values outside this identified set can generate the same data
  - You cannot find another parameter value outside this identified set that is observational equivalent

Oster Bound: Endogeneity without IV

- Point identification means
  - You can recover the exact point of the parameter from the data
  - 1-1 mapping between data and parameter value
  - No other parameter values can generate the same data
  - You cannot find another parameter value that is observational equivalent
- Set identification means
  - You can recover a set of the parameter from the data
  - No other parameter values outside this identified set can generate the same data
  - You cannot find another parameter value outside this identified set that is observational equivalent

Oster Bound: Endogeneity without IV

- Point identification means
  - You can recover the exact point of the parameter from the data
  - 1-1 mapping between data and parameter value
  - No other parameter values can generate the same data
  - You cannot find another parameter value that is observational equivalent
- Set identification means
  - You can recover a set of the parameter from the data
  - No other parameter values outside this identified set can generate the same data
  - You cannot find another parameter value outside this identified set that is observational equivalent

Oster Bound: Endogeneity without IV

- Point identification means
  - You can recover the exact point of the parameter from the data
  - 1-1 mapping between data and parameter value
  - No other parameter values can generate the same data
  - You cannot find another parameter value that is observational equivalent
- Set identification means
  - You can recover a set of the parameter from the data
  - No other parameter values outside this identified set can generate the same data
  - You cannot find another parameter value outside this identified set that is observational equivalent

Oster Bound: Endogeneity without IV

- Point identification means
  - You can recover the exact point of the parameter from the data
  - 1-1 mapping between data and parameter value
  - No other parameter values can generate the same data
  - You cannot find another parameter value that is observational equivalent
- Set identification means
  - You can recover a set of the parameter from the data
  - No other parameter values outside this identified set can generate the same data
  - You cannot find another parameter value outside this identified set that is observational equivalent

Oster Bound: Endogeneity without IV

- Point identification means
  - You can recover the exact point of the parameter from the data
  - 1-1 mapping between data and parameter value
  - No other parameter values can generate the same data
  - You cannot find another parameter value that is observational equivalent
- Set identification means
  - You can recover a set of the parameter from the data
  - No other parameter values outside this identified set can generate the same data
  - You cannot find another parameter value outside this identified set that is observational equivalent

Oster Bound: Endogeneity without IV

- Point identification means
  - You can recover the exact point of the parameter from the data
  - 1-1 mapping between data and parameter value
  - No other parameter values can generate the same data
  - You cannot find another parameter value that is observational equivalent
- Set identification means
  - You can recover a set of the parameter from the data
  - No other parameter values outside this identified set can generate the same data
  - You cannot find another parameter value outside this identified set that is observational equivalent

Oster Bound: Endogeneity without IV

- Point identification means
  - You can recover the exact point of the parameter from the data
  - 1-1 mapping between data and parameter value
  - No other parameter values can generate the same data
  - You cannot find another parameter value that is observational equivalent
- Set identification means
  - You can recover a set of the parameter from the data
  - No other parameter values outside this identified set can generate the same data
  - You cannot find another parameter value outside this identified set that is observational equivalent

Oster Bound: Endogeneity without IV

- The intuition of Oster bound is very simple
- We can use observed variables to evaluate how large the omitted bias can be
- Relation between treatment and unobservables can be partially recovered from relation between treatment and observables

Oster Bound: Endogeneity without IV

- The intuition of Oster bound is very simple
- We can use observed variables to evaluate how large the omitted bias can be
- Relation between treatment and unobservables can be partially recovered from relation between treatment and observables

Oster Bound: Endogeneity without IV

- The intuition of Oster bound is very simple
- We can use observed variables to evaluate how large the omitted bias can be
- Relation between treatment and unobservables can be partially recovered from relation between treatment and observables

Oster Bound: Endogeneity without IV

- The intuition of Oster bound is very simple
- We can use observed variables to evaluate how large the omitted bias can be
- Relation between treatment and unobservables can be partially recovered from relation between treatment and observables

Oster Bound: Endogeneity without IV

- If there is large omitted variable bias, inclusion of omitted variables will change the coefficient estimation a lot
- When we additionally include one more control variable:
  - How stable is the coefficient? (stability)
  - How much of y is explained by this control? (informative)
- If the coefficient estimation is changed only a little, by a strong control
  $\Longrightarrow$ We are safe

Oster Bound: Endogeneity without IV

- If there is large omitted variable bias, inclusion of omitted variables will change the coefficient estimation a lot
- When we additionally include one more control variable:
  - How stable is the coefficient? (stability)
  - How much of y is explained by this control? (informative)
- If the coefficient estimation is changed only a little, by a strong control
  $\Longrightarrow$ We are safe

Oster Bound: Endogeneity without IV

- If there is large omitted variable bias, inclusion of omitted variables will change the coefficient estimation a lot
- When we additionally include one more control variable:
  - How stable is the coefficient? (stability)
  - How much of y is explained by this control? (informative)
- If the coefficient estimation is changed only a little, by a strong control
  $\Longrightarrow$ We are safe

Oster Bound: Endogeneity without IV

- If there is large omitted variable bias, inclusion of omitted variables will change the coefficient estimation a lot
- When we additionally include one more control variable:
  - How stable is the coefficient? (stability)
  - How much of y is explained by this control? (informative)
- If the coefficient estimation is changed only a little, by a strong control $\Longrightarrow$ We are safe

Oster Bound: Endogeneity without IV

- If there is large omitted variable bias, inclusion of omitted variables will change the coefficient estimation a lot
- When we additionally include one more control variable:
  - How stable is the coefficient? (stability)
  - How much of y is explained by this control? (informative)
- If the coefficient estimation is changed only a little, by a strong control
  $\Longrightarrow$ We are safe

Oster Bound: Endogeneity without IV

- If there is large omitted variable bias, inclusion of omitted variables will change the coefficient estimation a lot
- When we additionally include one more control variable:
  - How stable is the coefficient? (stability)
  - How much of y is explained by this control? (informative)
- If the coefficient estimation is changed only a little, by a strong control
  $\Longrightarrow$ We are safe

Oster Bound: Theory

- Assume that we are interested in the effect of X on Y
- We have two sets of other variables W_1, W_2, correlated with both X and Y
- W_1 can be represented by some observed proxies, W_2 is unobservable
- Consider the following model:

$$\begin{aligned} Y &= \beta X + \Psi\omega + W_2 + \epsilon \\ W_1 &= \Psi\omega \end{aligned}$$

- Assume that W_1 and W_2 are orthogonal

Oster Bound: Theory

- Assume that we are interested in the effect of X on Y
- We have two sets of other variables W_1, W_2, correlated with both X and Y
- W_1 can be represented by some observed proxies, W_2 is unobservable
- Consider the following model:

$$Y = \beta X + \Psi\omega + W_2 + \epsilon$$
$$W_1 = \Psi\omega$$

- Assume that W_1 and W_2 are orthogonal

Oster Bound: Theory

- Assume that we are interested in the effect of X on Y
- We have two sets of other variables W_1, W_2, correlated with both X and Y
- W_1 can be represented by some observed proxies, W_2 is unobservable
- Consider the following model:

$$Y = \beta X + \Psi\omega + W_2 + \epsilon$$
$$W_1 = \Psi\omega$$

- Assume that W_1 and W_2 are orthogonal

Oster Bound: Theory

- Assume that we are interested in the effect of X on Y
- We have two sets of other variables W_1, W_2, correlated with both X and Y
- W_1 can be represented by some observed proxies, W_2 is unobservable
- Consider the following model:

$$Y = \beta X + \Psi\omega + W_2 + \epsilon$$
$$W_1 = \Psi\omega$$

- Assume that W_1 and W_2 are orthogonal

Oster Bound: Theory

- Assume that we are interested in the effect of X on Y
- We have two sets of other variables W_1, W_2, correlated with both X and Y
- W_1 can be represented by some observed proxies, W_2 is unobservable
- Consider the following model:

$$
\begin{aligned}
Y &= \beta X + \Psi\omega + W_2 + \epsilon \\
W_1 &= \Psi\omega
\end{aligned}
$$

- Assume that W_1 and W_2 are orthogonal

Oster Bound: Theory

- Assume that we are interested in the effect of X on Y
- We have two sets of other variables W_1, W_2, correlated with both X and Y
- W_1 can be represented by some observed proxies, W_2 is unobservable
- Consider the following model:

$$
\begin{aligned}
Y &= \beta X + \Psi\omega + W_2 + \epsilon \\
W_1 &= \Psi\omega
\end{aligned}
$$

- Assume that W_1 and W_2 are orthogonal

Oster Bound: Theory

- Denote δ as the proportional selection relationship:

$$\delta\frac{\sigma_{1X}}{\sigma_1^2}=\frac{\sigma_{2X}}{\sigma_2^2},\text{ where }\sigma_{iX}=cov(W_i,X),\sigma_i^2=Var(W_i)$$

- δ means the relative degree of W_1 and W_2's relation to treatment X
- When δ is large, it means the observed control is relatively not important as the unobserved one
- When $\delta=1$, the unobservale and observable are equally related to the treatment

Oster Bound: Theory

- Denote δ as the proportional selection relationship:

$$\delta\frac{\sigma_{1X}}{\sigma_1^2} = \frac{\sigma_{2X}}{\sigma_2^2}, \text{ where } \sigma_{iX} = cov(W_i, X), \sigma_i^2 = Var(W_i)$$

- δ means the relative degree of W_1 and W_2's relation to treatment X
- When δ is large, it means the observed control is relatively not important as the unobserved one
- When $\delta = 1$, the unobservale and observable are equally related to the treatment

Oster Bound: Theory

- Denote δ as the proportional selection relationship:

$$\delta \frac{\sigma_{1X}}{\sigma_1^2} = \frac{\sigma_{2X}}{\sigma_2^2}, \text{ where } \sigma_{iX} = cov(W_i, X), \sigma_i^2 = Var(W_i)$$

- δ means the relative degree of W_1 and W_2's relation to treatment X
- When δ is large, it means the observed control is relatively not important as the unobserved one
- When $\delta = 1$, the unobservale and observable are equally related to the treatment

Oster Bound: Theory

- Denote δ as the proportional selection relationship:

$$\delta \frac{\sigma_{1X}}{\sigma_1^2} = \frac{\sigma_{2X}}{\sigma_2^2}, \text{ where } \sigma_{iX} = cov(W_i, X), \sigma_i^2 = Var(W_i)$$

- δ means the relative degree of W_1 and W_2's relation to treatment X
- When δ is large, it means the observed control is relatively not important as the unobserved one
- When $\delta = 1$, the unobservale and observable are equally related to the treatment

Oster Bound: Theory

- Denote δ as the proportional selection relationship:

$$\delta\frac{\sigma_{1X}}{\sigma_1^2} = \frac{\sigma_{2X}}{\sigma_2^2}, \text{ where } \sigma_{iX} = cov(W_i, X), \sigma_i^2 = Var(W_i)$$

- δ means the relative degree of W_1 and W_2's relation to treatment X
- When δ is large, it means the observed control is relatively not important as the unobserved one
- When $\delta = 1$, the unobservale and observable are equally related to the treatment

Oster Bound: Theory

- We further denote β and R-square for three regressions
- Short regression: reg Y on X $\Longrightarrow \mathring{\beta}, \mathring{R}$
- Intermediate regression: reg Y on X, $\omega \Longrightarrow \tilde{\beta}, \tilde{R}$
- Full regression: reg Y on X, ω, $W_2 \Longrightarrow R_{max}$

Oster Bound: Theory

- We further denote β and R-square for three regressions
- Short regression: reg Y on X $\Rightarrow \mathring{\beta}, \mathring{R}$
- Intermediate regression: reg Y on X, $\omega \Rightarrow \tilde{\beta}, \tilde{R}$
- Full regression: reg Y on X, ω, $W_2 \Rightarrow R_{max}$

Oster Bound: Theory

- We further denote β and R-square for three regressions
- Short regression: reg Y on X $\Rightarrow \mathring{\beta}, \mathring{R}$
- Intermediate regression: reg Y on X, $\omega \Rightarrow \tilde{\beta}, \tilde{R}$
- Full regression: reg Y on X, ω, $W_2 \Rightarrow R_{max}$

Oster Bound: Theory

- We further denote β and R-square for three regressions
- Short regression: reg Y on X $\Rightarrow \mathring{\beta}, \mathring{R}$
- Intermediate regression: reg Y on X, $\omega \Rightarrow \tilde{\beta}, \tilde{R}$
- Full regression: reg Y on X, ω, $W_2 \Rightarrow R_{max}$

Oster Bound: Theory

- We further denote β and R-square for three regressions
- Short regression: reg Y on X $\Rightarrow \mathring{\beta}, \mathring{R}$
- Intermediate regression: reg Y on X, $\omega \Rightarrow \tilde{\beta}, \tilde{R}$
- Full regression: reg Y on X, ω, $W_2 \Rightarrow R_{max}$

Oster Bound: Theory

- There are two important pieces in this issue
- δ: relative correlation of observed vs. unobserved variable with X
- R_{max}: total variation you can explain
  - Given we know $\mathring{R}$ and $\tilde{R}$ (just do the regs)
  - We can infer how much variation we explain using observed variables
  - Thus, knowing R_{max} means knowing the portion of variations we can explain by the additional observed control W_i

Oster Bound: Theory

- There are two important pieces in this issue
- δ: relative correlation of observed vs. unobserved variable with X
- R_{max}: total variation you can explain
  - Given we know $\mathring{R}$ and $\tilde{R}$ (just do the regs)
  - We can infer how much variation we explain using observed variables
  - Thus, knowing R_{max} means knowing the portion of variations we can explain by the additional observed control W_1

Oster Bound: Theory

- There are two important pieces in this issue
- δ: relative correlation of observed vs. unobserved variable with X
- R_{max}: total variation you can explain
  - Given we know $\mathring{R}$ and $\tilde{R}$ (just do the regs)
  - We can infer how much variation we explain using observed variables
  - Thus, knowing R_{max} means knowing the portion of variations we can explain by the additional observed control W_1

Oster Bound: Theory

- There are two important pieces in this issue
- δ: relative correlation of observed vs. unobserved variable with X
- R_{max}: total variation you can explain
  - Given we know $\mathring{R}$ and $\tilde{R}$ (just do the regs)
  - We can infer how much variation we explain using observed variables
  - Thus, knowing R_{max} means knowing the portion of variations we can explain by the additional observed control W_1

Oster Bound: Theory

- There are two important pieces in this issue
- δ: relative correlation of observed vs. unobserved variable with X
- R_{max}: total variation you can explain
  - Given we know $\mathring{R}$ and $\tilde{R}$ (just do the regs)
  - We can infer how much variation we explain using observed variables
  - Thus, knowing R_{max} means knowing the portion of variations we can explain by the additional observed control W_1

Oster Bound: Theory

- There are two important pieces in this issue
- δ: relative correlation of observed vs. unobserved variable with X
- R_{max}: total variation you can explain
  - Given we know $\mathring{R}$ and $\tilde{R}$ (just do the regs)
  - We can infer how much variation we explain using observed variables
  - Thus, knowing R_{max} means knowing the portion of variations we can explain by the additional observed control W_1

Oster Bound: Theory

- There are two important pieces in this issue
- δ: relative correlation of observed vs. unobserved variable with X
- R_{max}: total variation you can explain
  - Given we know $\mathring{R}$ and $\tilde{R}$ (just do the regs)
  - We can infer how much variation we explain using observed variables
  - Thus, knowing R_{max} means knowing the portion of variations we can explain by the additional observed control W_1

Oster Bound: Theory

- We have two propositions connecting δ, R_{max} and bias:

Proposition 1 in Oster (2019)

Given δ and R_{max}, we can calculate the bias and find a debiased estimator. But in some cases, there will be multiple solutions and we need to implement solution selection. $\delta, R_{max} \rightarrow bias, \beta$

Proposition 2 in Oster (2019)

Given R_{max} and any value of treatment effect β, we can find a δ to make bias zero. $R_{max}, \beta, \text{bias} = 0 \rightarrow \delta$

Oster Bound: Theory

- We have two propositions connecting δ, R_{max} and bias:

Proposition 2 in Oster (2019)

Given δ and R_{max}, we can calculate the bias and find a debiased estimator. But in some cases, there will be multiple solutions and we need to implement solution selection. $\delta, R_{max} \rightarrow bias, \beta$

Proposition 3 in Oster (2019)

Given R_{max} and any value of treatment effect β, we can find a δ to make bias zero. $R_{max}, \beta, bias = 0 \rightarrow \delta$

Oster Bound: Theory

- We have two propositions connecting δ, R_{max} and bias:

Proposition 2 in Oster (2019)

Given δ and R_{max}, we can calculate the bias and find a debiased estimator. But in some cases, there will be multiple solutions and we need to implement solution selection. $\delta, R_{max} \rightarrow bias, \beta$

Proposition 3 in Oster (2019)

Given R_{max} and any value of treatment effect β, we can find a δ to make bias zero. $R_{max}, \beta, bias = 0 \rightarrow \delta$

Oster Bound: Theory

- We have two propositions connecting δ, R_{max} and bias:

Proposition 2 in Oster (2019)

Given δ and R_{max}, we can calculate the bias and find a debiased estimator. But in some cases, there will be multiple solutions and we need to implement solution selection. $\delta, R_{max} \rightarrow bias, \beta$

Proposition 3 in Oster (2019)

Given R_{max} and any value of treatment effect β, we can find a δ to make bias zero. $R_{max}, \beta, bias = 0 \rightarrow \delta$

Oster Bound: Theory

- Proposition 2 is simple, showing the way to calculate the bias
- Therefore, we can have a debiased estimator
- However this is only theoretically
- We never know what are δ and R_{max} since we do not observe W_2
- But it still gives us a chance to calculate a bound on β, when we assume some bounds on δ and R_{max}

Oster Bound: Theory

- Proposition 2 is simple, showing the way to calculate the bias
- Therefore, we can have a debiased estimator
- However this is only theoretically
- We never know what are δ and R_{max} since we do not observe W_2
- But it still gives us a chance to calculate a bound on β, when we assume some bounds on δ and R_{max}

Oster Bound: Theory

- Proposition 2 is simple, showing the way to calculate the bias
- Therefore, we can have a debiased estimator
- However this is only theoretically
- We never know what are δ and R_{max} since we do not observe W_2
- But it still gives us a chance to calculate a bound on β, when we assume some bounds on δ and R_{max}

Oster Bound: Theory

- Proposition 2 is simple, showing the way to calculate the bias
- Therefore, we can have a debiased estimator
- However this is only theoretically
- We never know what are δ and R_{max} since we do not observe W_2
- But it still gives us a chance to calculate a bound on β, when we assume some bounds on δ and R_{max}

Oster Bound: Theory

- Proposition 2 is simple, showing the way to calculate the bias
- Therefore, we can have a debiased estimator
- However this is only theoretically
- We never know what are δ and R_{max} since we do not observe W_2
- But it still gives us a chance to calculate a bound on β, when we assume some bounds on δ and R_{max}

Oster Bound: Theory

- Proposition 2 is simple, showing the way to calculate the bias
- Therefore, we can have a debiased estimator
- However this is only theoretically
- We never know what are δ and R_{max} since we do not observe W_2
- But it still gives us a chance to calculate a bound on β, when we assume some bounds on δ and R_{max}

Oster Bound: Theory

- Proposition 3 has an important implication: we can assume the "true effect" $\beta = 0$ and find the corresponding δ
- It means how large δ has to be to erase our result to zero
- How important should unobservables be (related to X) to make the true effect zero
- If this threshold of δ is large, zero true effect is unlikely to happen
  $\Longrightarrow$ Our results are robust

Oster Bound: Theory

- Proposition 3 has an important implication: we can assume the "true effect" $\beta = 0$ and find the corresponding δ
- It means how large δ has to be to erase our result to zero
- How important should unobservables be (related to X) to make the true effect zero
- If this threshold of δ is large, zero true effect is unlikely to happen
  $\Longrightarrow$ Our results are robust

Oster Bound: Theory

- Proposition 3 has an important implication: we can assume the "true effect" $\beta = 0$ and find the corresponding δ
- It means how large δ has to be to erase our result to zero
- How important should unobservables be (related to X) to make the true effect zero
- If this threshold of δ is large, zero true effect is unlikely to happen
  $\Longrightarrow$ Our results are robust

Oster Bound: Theory

- Proposition 3 has an important implication: we can assume the "true effect" $\beta = 0$ and find the corresponding δ
- It means how large δ has to be to erase our result to zero
- How important should unobservables be (related to X) to make the true effect zero
- If this threshold of δ is large, zero true effect is unlikely to happen
  $\Longrightarrow$ Our results are robust

Oster Bound: Theory

- Proposition 3 has an important implication: we can assume the "true effect" $\beta = 0$ and find the corresponding δ
- It means how large δ has to be to erase our result to zero
- How important should unobservables be (related to X) to make the true effect zero
- If this threshold of δ is large, zero true effect is unlikely to happen $\Longrightarrow$ Our results are robust

Oster Bound: Theory

- Proposition 2 and 3 gives two equations connecting δ, R_{max} and bias
- They are very complicated
- However, if we assume $\delta = 1$, the equation can be reduced to:

$$\beta^* = \tilde{\beta} - \underbrace{[\mathring{\beta} - \tilde{\beta}] \frac{R_{max} - \tilde{R}}{\tilde{R} - \mathring{R}}}_{bias}$$

Proposition 2 in Oster (2019)

When $\delta = 1$, the debiased estimator is asymptotically consistent: $\beta \xrightarrow{p} \beta$

- When we add controls, bias is positively related with coefficient change $\mathring{\beta} - \tilde{\beta}$, negatively related with R-square change $\tilde{R} - \mathring{R}$

Oster Bound: Theory

- Proposition 2 and 3 gives two equations connecting δ, R_{max} and bias
- They are very complicated
- However, if we assume $\delta = 1$, the equation can be reduced to:

$$\beta^* = \tilde{\beta} - \underbrace{[\mathring{\beta} - \tilde{\beta}]\frac{R_{max} - \tilde{R}}{\tilde{R} - \mathring{R}}}_{bias}$$

Proposition 1 in Oster (2019)

When $\delta = 1$, the debiased estimator is asymptotically consistent. $\beta^* \xrightarrow{p} \beta$

- When we add controls, bias is positively related with coefficient change $\mathring{\beta} - \tilde{\beta}$, negatively related with R-square change $\tilde{R} - \mathring{R}$

Oster Bound: Theory

- Proposition 2 and 3 gives two equations connecting δ, R_{max} and bias
- They are very complicated
- However, if we assume $\delta = 1$, the equation can be reduced to:

$$\beta^* = \tilde{\beta} - \underbrace{[\mathring{\beta} - \tilde{\beta}]\frac{R_{max} - \tilde{R}}{\tilde{R} - \mathring{R}}}_{bias}$$

Proposition 1 in Oster (2019)

When $\delta = 1$, the debiased estimator is asymptotically consistent. $\beta^* \xrightarrow{p} \beta$

- When we add controls, bias is positively related with coefficient change $\mathring{\beta} - \tilde{\beta}$, negatively related with R-square change $\tilde{R} - \mathring{R}$

Oster Bound: Theory

- Proposition 2 and 3 gives two equations connecting δ, R_{max} and bias
- They are very complicated
- However, if we assume $\delta = 1$, the equation can be reduced to:

$$\beta^* = \tilde{\beta} - \underbrace{[\mathring{\beta} - \tilde{\beta}]\frac{R_{max} - \tilde{R}}{\tilde{R} - \mathring{R}}}_{bias}$$

Proposition 1 in Oster (2019)

When $\delta = 1$, the debiased estimator is asymptotically consistent. $\beta \xrightarrow{p} \beta$

- When we add controls, bias is positively related with coefficient change $\mathring{\beta} - \tilde{\beta}$, negatively related with R-square change $\tilde{R} - \mathring{R}$

Oster Bound: Theory

- Proposition 2 and 3 gives two equations connecting δ, R_{max} and bias
- They are very complicated
- However, if we assume $\delta = 1$, the equation can be reduced to:

$$\beta^* = \tilde{\beta} - \underbrace{[\mathring{\beta} - \tilde{\beta}]\frac{R_{max} - \tilde{R}}{\tilde{R} - \mathring{R}}}_{bias}$$

Proposition 1 in Oster (2019)

When $\delta = 1$, the debiased estimator is asymptotically consistent. $\beta \xrightarrow{p} \beta$

- When we add controls, bias is positively related with coefficient change $\mathring{\beta} - \tilde{\beta}$, negatively related with R-square change $\tilde{R} - \mathring{R}$

Oster Bound: Theory

- Proposition 2 and 3 gives two equations connecting δ, R_{max} and bias
- They are very complicated
- However, if we assume $\delta = 1$, the equation can be reduced to:

$$\beta^* = \tilde{\beta} - \underbrace{[\mathring{\beta} - \tilde{\beta}]\frac{R_{max} - \tilde{R}}{\tilde{R} - \mathring{R}}}_{bias}$$

Proposition 1 in Oster (2019)

When $\delta = 1$, the debiased estimator is asymptotically consistent. $\beta \xrightarrow{p} \beta$

- When we add controls, bias is positively related with coefficient change $\mathring{\beta} - \tilde{\beta}$, negatively related with R-square change $\tilde{R} - \mathring{R}$

Oster Bound: Implementation

- How to implement Oster's method in practice?
- Two methods based on Propositions 2 and 3
- Method 1: Assume a value for R_{max} and calculate the value of δ for which $\beta = 0$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}$
  - 1.3 is derived to let 90% of the RCT studies in top journals pass this test
  - Set $\beta = 0$, find the corresponding δ
  - If $\delta > 1$, we are safe
- If unobservables need to be very important to erase our results, we are OK
- If a relatively small unobservable can erase our results, it is not robust

Oster Bound: Implementation

- How to implement Oster's method in practice?
- Two methods based on Propositions 2 and 3
- Method 1: Assume a value for R_{max} and calculate the value of δ for which $\beta = 0$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}$
  - 1.3 is derived to let 90% of the RCT studies in top journals pass this test
  - Set $\beta = 0$, find the corresponding δ
  - If $\delta > 1$, we are safe
- If unobservables need to be very important to erase our results, we are OK
- If a relatively small unobservable can erase our results, it is not robust

Oster Bound: Implementation

- How to implement Oster's method in practice?
- Two methods based on Propositions 2 and 3
- Method 1: Assume a value for R_{max} and calculate the value of δ for which $\beta = 0$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}$
  - 1.3 is derived to let 90% of the RCT studies in top journals pass this test
  - Set $\beta = 0$, find the corresponding δ
  - If $\delta > 1$, we are safe
- If unobservables need to be very important to erase our results, we are OK
- If a relatively small unobservable can erase our results, it is not robust

Oster Bound: Implementation

- How to implement Oster's method in practice?
- Two methods based on Propositions 2 and 3
- Method 1: Assume a value for R_{max} and calculate the value of δ for which $\beta = 0$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}$
  - 1.3 is derived to let 90% of the RCT studies in top journals pass this test
  - Set $\beta = 0$, find the corresponding δ
  - If $\delta > 1$, we are safe
- If unobservables need to be very important to erase our results, we are OK
- If a relatively small unobservable can erase our results, it is not robust

Oster Bound: Implementation

- How to implement Oster's method in practice?
- Two methods based on Propositions 2 and 3
- Method 1: Assume a value for R_{max} and calculate the value of δ for which $\beta = 0$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}$
  - 1.3 is derived to let 90% of the RCT studies in top journals pass this test
  - Set $\beta = 0$, find the corresponding δ
  - If $\delta > 1$, we are safe
- If unobservables need to be very important to erase our results, we are OK
- If a relatively small unobservable can erase our results, it is not robust

Oster Bound: Implementation

- How to implement Oster's method in practice?
- Two methods based on Propositions 2 and 3
- Method 1: Assume a value for R_{max} and calculate the value of δ for which $\beta = 0$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}$
  - 1.3 is derived to let 90% of the RCT studies in top journals pass this test
  - Set $\beta = 0$, find the corresponding δ
  - If $\delta > 1$, we are safe
- If unobservables need to be very important to erase our results, we are OK
- If a relatively small unobservable can erase our results, it is not robust

Oster Bound: Implementation

- How to implement Oster's method in practice?
- Two methods based on Propositions 2 and 3
- Method 1: Assume a value for R_{max} and calculate the value of δ for which $\beta = 0$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}$
  - 1.3 is derived to let 90% of the RCT studies in top journals pass this test
  - Set $\beta = 0$, find the corresponding δ
  - If $\delta > 1$, we are safe
- If unobservables need to be very important to erase our results, we are OK
- If a relatively small unobservable can erase our results, it is not robust

Oster Bound: Implementation

- How to implement Oster's method in practice?
- Two methods based on Propositions 2 and 3
- Method 1: Assume a value for R_{max} and calculate the value of δ for which $\beta = 0$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}$
  - 1.3 is derived to let 90% of the RCT studies in top journals pass this test
  - Set $\beta = 0$, find the corresponding δ
  - If $\delta > 1$, we are safe
- If unobservables need to be very important to erase our results, we are OK
- If a relatively small unobservable can erase our results, it is not robust

Oster Bound: Implementation

- How to implement Oster's method in practice?
- Two methods based on Propositions 2 and 3
- Method 1: Assume a value for R_{max} and calculate the value of δ for which $\beta = 0$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}$
  - 1.3 is derived to let 90% of the RCT studies in top journals pass this test
  - Set $\beta = 0$, find the corresponding δ
  - If $\delta > 1$, we are safe
- If unobservables need to be very important to erase our results, we are OK
- If a relatively small unobservable can erase our results, it is not robust

Oster Bound: Implementation

- How to implement Oster's method in practice?
- Two methods based on Propositions 2 and 3
- Method 1: Assume a value for R_{max} and calculate the value of δ for which $\beta = 0$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}$
  - 1.3 is derived to let 90% of the RCT studies in top journals pass this test
  - Set $\beta = 0$, find the corresponding δ
  - If $\delta > 1$, we are safe
- If unobservables need to be very important to erase our results, we are OK
- If a relatively small unobservable can erase our results, it is not robust

Oster Bound: Implementation

- Method 2: Assume a conservative value for R_{max} and δ, calculate the debiased estimation β^*, which gives you a bound $[\tilde{\beta}, \beta^*]$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}, \delta = 1$
  - Calculate a debiased $\beta^*(R_{max}, \delta = 1)$
  - A conservative bound of the estimation is $[\tilde{\beta}, \beta^*]$
- If the interval does not contain zero, we are OK
- If the interval contains zero, it is not robust

Oster Bound: Implementation

- Method 2: Assume a conservative value for R_{max} and δ, calculate the debiased estimation β^*, which gives you a bound $[\tilde{\beta}, \beta^*]$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}, \delta = 1$
  - Calculate a debiased $\beta^*(R_{max}, \delta = 1)$
  - A conservative bound of the estimation is $[\tilde{\beta}, \beta^*]$
- If the interval does not contain zero, we are OK
- If the interval contains zero, it is not robust

Oster Bound: Implementation

- Method 2: Assume a conservative value for R_{max} and δ, calculate the debiased estimation β^*, which gives you a bound $[\tilde{\beta}, \beta^*]$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}, \delta = 1$
  - Calculate a debiased $\beta^*(R_{max}, \delta = 1)$
  - A conservative bound of the estimation is $[\tilde{\beta}, \beta^*]$
- If the interval does not contain zero, we are OK
- If the interval contains zero, it is not robust

Oster Bound: Implementation

- Method 2: Assume a conservative value for R_{max} and δ, calculate the debiased estimation β^*, which gives you a bound $[\tilde{\beta}, \beta^*]$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}, \delta = 1$
  - Calculate a debiased $\beta^*(R_{max}, \delta = 1)$
  - A conservative bound of the estimation is $[\tilde{\beta}, \beta^*]$
- If the interval does not contain zero, we are OK
- If the interval contains zero, it is not robust

Oster Bound: Implementation

- Method 2: Assume a conservative value for R_{max} and δ, calculate the debiased estimation β^*, which gives you a bound $[\tilde{\beta}, \beta^*]$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}, \delta = 1$
  - Calculate a debiased $\beta^*(R_{max}, \delta = 1)$
  - A conservative bound of the estimation is $[\tilde{\beta}, \beta^*]$
- If the interval does not contain zero, we are OK
- If the interval contains zero, it is not robust

Oster Bound: Implementation

- Method 2: Assume a conservative value for R_{max} and δ, calculate the debiased estimation β^*, which gives you a bound $[\tilde{\beta}, \beta^*]$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}, \delta = 1$
  - Calculate a debiased $\beta^*(R_{max}, \delta = 1)$
  - A conservative bound of the estimation is $[\tilde{\beta}, \beta^*]$
- If the interval does not contain zero, we are OK
- If the interval contains zero, it is not robust

Oster Bound: Implementation

- Method 2: Assume a conservative value for R_{max} and δ, calculate the debiased estimation β^*, which gives you a bound $[\tilde{\beta}, \beta^*]$
  - As a rule of thumb, choose $R_{max} = \min\{1, 1.3\tilde{R}\}, \delta = 1$
  - Calculate a debiased $\beta^*(R_{max}, \delta = 1)$
  - A conservative bound of the estimation is $[\tilde{\beta}, \beta^*]$
- If the interval does not contain zero, we are OK
- If the interval contains zero, it is not robust

Oster Bound: Conclusion

- Oster bound is the last weapon you can use when nothing else works
- It can also be utilized as a robustness check
- But it has some intrinsic disadvantages
  - The choice of parameters are arbitrary
  - It can only give you a sense of the robustness of your results
- An Application in Economics: Clark et al. (2021)

Oster Bound: Conclusion

- Oster bound is the last weapon you can use when nothing else works
- It can also be utilized as a robustness check
- But it has some intrinsic disadvantages
  - The choice of parameters are arbitrary
  - It can only give you a sense of the robustness of your results
- An Application in Economics: Clark et al. (2021)

Oster Bound: Conclusion

- Oster bound is the last weapon you can use when nothing else works
- It can also be utilized as a robustness check
- But it has some intrinsic disadvantages
  - The choice of parameters are arbitrary
  - It can only give you a sense of the robustness of your results
- An Application in Economics: Clark et al. (2021)

Oster Bound: Conclusion

- Oster bound is the last weapon you can use when nothing else works
- It can also be utilized as a robustness check
- But it has some intrinsic disadvantages
  - The choice of parameters are arbitrary
  - It can only give you a sense of the robustness of your results
- An Application in Economics: Clark et al. (2021)

Oster Bound: Conclusion

- Oster bound is the last weapon you can use when nothing else works
- It can also be utilized as a robustness check
- But it has some intrinsic disadvantages
  - The choice of parameters are arbitrary
  - It can only give you a sense of the robustness of your results
- An Application in Economics: Clark et al. (2021)

Oster Bound: Conclusion

- Oster bound is the last weapon you can use when nothing else works
- It can also be utilized as a robustness check
- But it has some intrinsic disadvantages
  - The choice of parameters are arbitrary
  - It can only give you a sense of the robustness of your results
- An Application in Economics: Clark et al. (2021)

Oster Bound: Conclusion

- Oster bound is the last weapon you can use when nothing else works
- It can also be utilized as a robustness check
- But it has some intrinsic disadvantages
  - The choice of parameters are arbitrary
  - It can only give you a sense of the robustness of your results
- An Application in Economics: Clark et al. (2021)

Final Conclusion

- IV is the main strategy we can use to deal with endogeneity
- At least two assumptions: Exclusion restriction, Existence of first stage
- In heterogeneity TE, IV estimator gives us LATE
- GMM is the general framework for IV
- OLS, Wald, and 2SLS are all special cases of GMM estimator

Final Conclusion

- IV is the main strategy we can use to deal with endogeneity
- At least two assumptions: Exclusion restriction, Existence of first stage
- In heterogeneity TE, IV estimator gives us LATE
- GMM is the general framework for IV
- OLS, Wald, and 2SLS are all special cases of GMM estimator

Final Conclusion

- IV is the main strategy we can use to deal with endogeneity
- At least two assumptions: Exclusion restriction, Existence of first stage
- In heterogeneity TE, IV estimator gives us LATE
- GMM is the general framework for IV
- OLS, Wald, and 2SLS are all special cases of GMM estimator

Final Conclusion

- IV is the main strategy we can use to deal with endogeneity
- At least two assumptions: Exclusion restriction, Existence of first stage
- In heterogeneity TE, IV estimator gives us LATE
- GMM is the general framework for IV
- OLS, Wald, and 2SLS are all special cases of GMM estimator

Final Conclusion

- IV is the main strategy we can use to deal with endogeneity
- At least two assumptions: Exclusion restriction, Existence of first stage
- In heterogeneity TE, IV estimator gives us LATE
- GMM is the general framework for IV
- OLS, Wald, and 2SLS are all special cases of GMM estimator

Final Conclusion

- IV is the main strategy we can use to deal with endogeneity
- At least two assumptions: Exclusion restriction, Existence of first stage
- In heterogeneity TE, IV estimator gives us LATE
- GMM is the general framework for IV
- OLS, Wald, and 2SLS are all special cases of GMM estimator

Final Conclusion

- When valid IV is not available, Oster bound can help us
- The basic idea relates to the stability of the point estimation when a strong control is included
- It gives you a bounding result
- But it is not almighty, since some parameter choices are pretty arbitrary

Final Conclusion

- When valid IV is not available, Oster bound can help us
- The basic idea relates to the stability of the point estimation when a strong control is included
- It gives you a bounding result
- But it is not almighty, since some parameter choices are pretty arbitrary

Final Conclusion

- When valid IV is not available, Oster bound can help us
- The basic idea relates to the stability of the point estimation when a strong control is included
- It gives you a bounding result
- But it is not almighty, since some parameter choices are pretty arbitrary

Final Conclusion

- When valid IV is not available, Oster bound can help us
- The basic idea relates to the stability of the point estimation when a strong control is included
- It gives you a bounding result
- But it is not almighty, since some parameter choices are pretty arbitrary

Final Conclusion

- When valid IV is not available, Oster bound can help us
- The basic idea relates to the stability of the point estimation when a strong control is included
- It gives you a bounding result
- But it is not almighty, since some parameter choices are pretty arbitrary

References

Angrist, Joshua D and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Clark, Andrew E, Huifu Nong, Hongjia Zhu, and Rong Zhu. 2021. "Compensating for Academic Loss: Online Learning and Student Performance during the COVID-19 Pandemic." *China Economic Review* 68:101629.

Hansen, Bruce. 2022. *Econometrics*. Princeton University Press.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37 (2):187–204.

Pinto, Rodrigo. 2015. "Selection Bias in a Controlled Experiment: The Case of Moving to Opportunity." *Unpublished Ph. D. Thesis, University of Chicago, Department of Economics* .