Frontier Topics in Empirical Economics: Chapter 5 Introduction to IV and Endogeneity Issue

Zibin Huang [1]

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

October 8, 2026

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 $\Longrightarrow$ 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 $\Longrightarrow$ 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 $\Longrightarrow$ 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 $\Rightarrow$ 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

- 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) into (1), 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
- Here we need an instrumental variable

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) into (1), 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
- Here we need an instrumental variable

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) into (1), 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
- Here we need an instrumental variable

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) into (1), 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
- Here we need an instrumental variable

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) into (1), 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
- Here we need an instrumental variable

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) into (1), 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
- Here we need an instrumental variable

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

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

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

- Exclusion: z_i affects Y_i only through s_i, with no direct effect on Y_i
- Relevance: z_i actually moves s_i, so we have a "lever" on the endogenous variable

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

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

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

- Exclusion: z_i affects Y_i only through s_i, with no direct effect on Y_i
- Relevance: z_i actually moves s_i, so we have a "lever" on the endogenous variable

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

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

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

- Exclusion: z_i affects Y_i only through s_i, with no direct effect on Y_i
- Relevance: z_i actually moves s_i, so we have a "lever" on the endogenous variable

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

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

  We call it an "Instrumental Variable" (IV).
- Exclusion: z_i affects Y_i only through s_i, with no direct effect on Y_i
- Relevance: z_i actually moves s_i, so we have a "lever" on the endogenous variable

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

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

  We call it an "Instrumental Variable" (IV).
- Exclusion: z_i affects Y_i only through s_i, with no direct effect on Y_i
- Relevance: z_i actually moves s_i, so we have a "lever" on the endogenous variable

Simple IV: Identification

- Identification of ρ using IV is simple via covariances
- Calculating covariance of z_i and Y_i:

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

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

- Thus, ATE is identified by the ratio of two covariances
- Equivalently, two regression coefficients, (reg y on z)/(reg s on z)
- When IV z_i is binary:

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

- $E[Y_i|z_i = 1] - E[Y_i|z_i = 0]$ is called Intention-to-treat

Simple IV: Identification

- Identification of ρ using IV is simple via covariances
- Calculating covariance of z_i and Y_i:

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

- Thus, ATE is identified by the ratio of two covariances
- Equivalently, two regression coefficients, (reg y on z)/(reg s on z)
- When IV z_i is binary:

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

- $E[Y_i|z_i = 1] - E[Y_i|z_i = 0]$ is called Intention-to-treat

Simple IV: Identification

- Identification of ρ using IV is simple via covariances
- 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, ATE is identified by the ratio of two covariances
- Equivalently, two regression coefficients, (reg y on z)/(reg s on z)
- When IV z_i is binary:

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

- $E[Y_i|z_i = 1] - E[Y_i|z_i = 0]$ is called Intention-to-treat

Simple IV: Identification

- Identification of ρ using IV is simple via covariances
- 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, ATE is identified by the ratio of two covariances
- Equivalently, two regression coefficients, (reg y on z)/(reg s on z)
- When IV z_i is binary:

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

- $E[Y_i|z_i = 1] - E[Y_i|z_i = 0]$ is called Intention-to-treat

Simple IV: Identification

- Identification of ρ using IV is simple via covariances
- 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, ATE is identified by the ratio of two covariances
- Equivalently, two regression coefficients, (reg y on z)/(reg s on z)
- When IV z_i is binary:

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

- $E[Y_i|z_i = 1] - E[Y_i|z_i = 0]$ is called Intention-to-treat

Simple IV: Identification

- Identification of ρ using IV is simple via covariances
- 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, ATE is identified by the ratio of two covariances
- Equivalently, two regression coefficients, (reg y on z)/(reg s on z)
- When IV z_i is binary:

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

- $E[Y_i|z_i = 1] - E[Y_i|z_i = 0]$ is called Intention-to-treat

Simple IV: Identification

- Identification of ρ using IV is simple via covariances
- 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, ATE is identified by the ratio of two covariances
- Equivalently, two regression coefficients, (reg y on z)/(reg s on z)
- When IV z_i is binary:

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

- $E[Y_i|z_i = 1] - E[Y_i|z_i = 0]$ is called Intention-to-treat

Simple IV: Wald Estimator

- Simple regression coefficients are ratios of covariance to variance:

$$s_i = \pi_{10} + \pi_1 z_i + \xi_{1i} \quad \text{(First Stage)}$$
$$Y_i = \pi_{20} + \pi_2 z_i + \xi_{2i} \quad \text{(Reduced Form)}$$
$$\rho = \frac{\pi_2}{\pi_1}$$

- 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

- Simple regression coefficients are ratios of covariance to variance:

$$
\begin{aligned}
s_i &= \pi_{10} + \pi_1 z_i + \xi_{1i} \quad \text{(First Stage)} \\
Y_i &= \pi_{20} + \pi_2 z_i + \xi_{2i} \quad \text{(Reduced Form)} \\
\rho &= \frac{\pi_2}{\pi_1}
\end{aligned}
$$

- 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: Wald Estimator

- Simple regression coefficients are ratios of covariance to variance:

$$
\begin{aligned}
s_i &= \pi_{10} + \pi_1 z_i + \xi_{1i} \quad \text{(First Stage)} \\
Y_i &= \pi_{20} + \pi_2 z_i + \xi_{2i} \quad \text{(Reduced Form)} \\
\rho &= \frac{\pi_2}{\pi_1}
\end{aligned}
$$

- 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: Wald Estimator

- Simple regression coefficients are ratios of covariance to variance:

$$
\begin{aligned}
s_i &= \pi_{10} + \pi_1 z_i + \xi_{1i} \quad \text{(First Stage)} \\
Y_i &= \pi_{20} + \pi_2 z_i + \xi_{2i} \quad \text{(Reduced Form)} \\
\rho &= \frac{\pi_2}{\pi_1}
\end{aligned}
$$

- 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 Estimator

- Another way of IV estimation 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 Estimator

- Another way of IV estimation 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 Estimator

- Another way of IV estimation 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 Estimator

- Another way of IV estimation 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 Estimator

- 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 $Cov[z_i \xi_{2i}] = 0$
- $(X_i'\pi_{10} + \pi_{11} z_i)$ is the CEF/regression prediction of s_i on z_i given X_i

Simple IV: 2SLS Estimator

- 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 $Cov[z_i \xi_{2i}] = 0$
- $(X_i'\pi_{10} + \pi_{11} z_i)$ is the CEF/regression prediction of s_i on z_i given X_i

Simple IV: 2SLS Estimator

- 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 $Cov[z_i \xi_{2i}] = 0$
- $(X_i'\pi_{10} + \pi_{11} z_i)$ is the CEF/regression prediction of s_i on z_i given X_i

Simple IV: 2SLS Estimator

- 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 $Cov[z_i \xi_{2i}] = 0$
- $(X_i'\pi_{10} + \pi_{11} z_i)$ is the CEF/regression prediction of s_i on z_i given X_i

Simple IV: 2SLS Estimator

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

$$x = \hat{\pi}_{zx} z + \hat{\pi}_{wx} w + \hat{v}$$

  - Step 2: Running Y on predicted value $\hat{x}$ and w

$$Y = \rho \hat{x} + \beta w + \hat{\varepsilon}$$

Simple IV: 2SLS Estimator

- 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 Estimator

- 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 Estimator

- 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 standard error 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 if you have a reason to believe that z affects s

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 standard error 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 if you have a reason to believe that z affects s

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 standard error 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 if you have a reason to believe that z affects s

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 standard error 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 if you have a reason to believe that z affects s

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 standard error 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 if you have a reason to believe that z affects s

Simple IV: Some Tips

- 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 more efficient
  most efficient under homoskedasticity

Simple IV: Some Tips

- 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 more efficient
  most efficient under homoskedasticity

Simple IV: Some Tips

- 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 more efficient most efficient under homoskedasticity

Simple IV: Some Tips

- 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 more efficient
  most efficient under homoskedasticity

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 earnings (Angrist, 1990)
  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
- A 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 earnings (Angrist, 1990)
  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
- A 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 earnings (Angrist, 1990)
  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
- A 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 earnings (Angrist, 1990)
  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
- A 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 earnings (Angrist, 1990)
  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
- A 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 earnings (Angrist, 1990)
  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
- A 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 earnings (Angrist, 1990)
  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
- A 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 $\iff$ 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
- 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
- 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
- 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
- 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
- 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, the instrument changes treatment in the same direction (or no change)
- 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 never 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, the instrument changes treatment in the same direction (or no change)
- 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 never 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, the instrument changes treatment in the same direction (or no change)
- 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 never 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, the instrument changes treatment in the same direction (or no change)
- 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 never 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, the instrument changes treatment in the same direction (or no change)
- 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 never 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, the instrument changes treatment in the same direction (or no change)
- 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 never 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, the instrument changes treatment in the same direction (or no change)
- 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 never 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, the instrument changes treatment in the same direction (or no change)
- 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 never take treatment
- No defiers!

IV with Heterogeneous Treatment Effect: LATE

- Local Average Treatment Effect (LATE), from the landmark paper Imbens and Angrist (1994)

IV with Heterogeneous Treatment Effect: LATE

- Local Average Treatment Effect (LATE), from the landmark paper Imbens and Angrist (1994)

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

- Local Average Treatment Effect (LATE), from the landmark paper Imbens and Angrist (1994)

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: By monotonicity, there are no defiers, so the population consists of always-takers (A), compliers (C), and never-takers (N). We decompose ITT as follows.

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

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):

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

IV with Heterogeneous Treatment Effect: LATE

Proof: By monotonicity, there are no defiers, so the population consists of always-takers (A), compliers (C), and never-takers (N). We decompose ITT as follows.

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

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):

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

IV with Heterogeneous Treatment Effect: LATE

Proof: By monotonicity, there are no defiers, so the population consists of always-takers (A), compliers (C), and never-takers (N). We decompose ITT as follows.

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

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):

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

IV with Heterogeneous Treatment Effect: LATE

Proof: By monotonicity, there are no defiers, so the population consists of always-takers (A), compliers (C), and never-takers (N). We decompose ITT as follows.

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

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):

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

$$\Rightarrow E[Y_{1i} - Y_{0i}|C_i] = \frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{P[C_i]}, \quad \text{Here we have } P(C_i|z_i = 1) = P(C_i|z_i = 0)$$

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

IV with Heterogeneous Treatment Effect: LATE

Proof: By monotonicity, there are no defiers, so the population consists of always-takers (A), compliers (C), and never-takers (N). We decompose ITT as follows.

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

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):

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

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 the 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 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 the 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 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 the 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 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 the 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 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 the 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 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 the 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 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 the 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 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 $g_i(\beta) = x_i(Y_i - x_i'\beta)$
- Assume no perfect co-linearity:
  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 $g_i(\beta) = x_i(Y_i - x_i'\beta)$
- Assume no perfect co-linearity:
  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 $g_i(\beta) = x_i(Y_i - x_i'\beta)$
- Assume no perfect co-linearity:
  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 $g_i(\beta) = x_i(Y_i - x_i'\beta)$
- Assume no perfect co-linearity:
  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 $g_i(\beta) = x_i(Y_i - x_i'\beta)$
- Assume no perfect co-linearity:
  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 $g_i(\beta) = x_i(Y_i - x_i'\beta)$
- Assume no perfect co-linearity:
  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) $\Rightarrow$ 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) $\Rightarrow$ 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

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

$$\hat{\beta}_{gmm} = \underset{\beta}{\operatorname{argmin}}\, J(\beta)$$

- W is some weighting matrix
- J measures the squared 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

$$J(\beta) = n\bar{g}_n(\beta)' W \bar{g}_n(\beta)$$
$$\hat{\beta}_{gmm} = \underset{\beta}{\operatorname{argmin}}\, J(\beta)$$

- W is some weighting matrix
- J measures the squared 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

$$J(\beta) = n\bar{g}_n(\beta)' W \bar{g}_n(\beta)$$
$$\hat{\beta}_{gmm} = \underset{\beta}{\operatorname{argmin}}\, J(\beta)$$

- W is some weighting matrix
- J measures the squared 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

$$J(\beta) = n\bar{g}_n(\beta)' W \bar{g}_n(\beta)$$
$$\hat{\beta}_{gmm} = \underset{\beta}{\operatorname{argmin}}\, J(\beta)$$

- W is some weighting matrix
- J measures the squared 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

$$J(\beta) = n\bar{g}_n(\beta)' W \bar{g}_n(\beta)$$
$$\hat{\beta}_{gmm} = \underset{\beta}{\text{argmin}}\, J(\beta)$$

- W is some weighting matrix
- J measures the squared 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

$$J(\beta) = n\bar{g}_n(\beta)' W \bar{g}_n(\beta)$$
$$\hat{\beta}_{gmm} = \underset{\beta}{\text{argmin}}\, J(\beta)$$

- W is some weighting matrix
- J measures the squared 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

$$J(\beta) = n\bar{g}_n(\beta)' W \bar{g}_n(\beta)$$
$$\hat{\beta}_{gmm} = \underset{\beta}{\operatorname{argmin}}\, J(\beta)$$

- W is some weighting matrix
- J measures the squared 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

$$J(\beta) = n\bar{g}_n(\beta)' W \bar{g}_n(\beta)$$
$$\hat{\beta}_{gmm} = \underset{\beta}{\operatorname{argmin}}\, J(\beta)$$

- W is some weighting matrix
- J measures the squared 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 (sample analogue) to be:

$$\hat{\beta}_{gmm} = \underset{\beta}{\operatorname{argmin}} \underbrace{\frac{1}{n}(Z'Y - Z'X\beta)'W(Z'Y - Z'X\beta)}_{J(\beta)}$$

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 (sample analogue) to be:

$$\hat{\beta}_{gmm} = \underset{\beta}{\operatorname{argmin}} \underbrace{\frac{1}{n}(Z'Y - Z'X\beta)'W(Z'Y - Z'X\beta)}_{J(\beta)}$$

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 (sample analogue) to be:

$$\hat{\beta}_{gmm} = \underset{\beta}{\operatorname{argmin}} \underbrace{\frac{1}{n}(Z'Y - Z'X\beta)'W(Z'Y - Z'X\beta)}_{J(\beta)}$$

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 (sample analogue) to be:

$$\hat{\beta}_{gmm} = \underset{\beta}{\text{argmin}} \underbrace{\frac{1}{n}(Z'Y - Z'X\beta)'W(Z'Y - Z'X\beta)}_{J(\beta)}$$

Multiple IV: Linear GMM

- Solve this minimization problem, we have

For the linear IV model with k (possibly endogenous) regressors and $\ell \geq k$ instruments

$$\beta_{\text{GMM}} = (X'ZWZ'X)^{-1}(X'ZWZ'Y)$$

- 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 linear IV model with k (possibly endogenous) regressors and $l \geq k$ 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 linear IV model with k (possibly endogenous) regressors and $l \geq k$ 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 linear IV model with k (possibly endogenous) regressors and $l \geq k$ 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 linear IV model with k (possibly endogenous) regressors and $l \geq k$ 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 $X = Z, W = \mathbf{I}$, we have:

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

- 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 $X = Z, W = \mathbf{I}$, we have:

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

- 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 $X = Z, W = \mathbf{I}$, we have:

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

- 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 $X = Z, W = \mathbf{I}$, we have:

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

- 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 $l = k$, $W = \mathbf{I}$, we have:

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

- 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 $l = k, W = \mathbf{I}$, we have:

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

- 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 $l = k, W = \mathbf{I}$, we have:

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

- 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 $l = k, W = \mathbf{I}$, we have:

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

- 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 $P_z = Z(Z'Z)^{-1}Z'$ be the projection matrix
- First stage fitted value then becomes $\hat{X} = P_z X$
- P is idempotent: $P \cdot P = P$
- When $W = (Z'Z)^{-1}$, 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.
- This nesting is not just algebra: the GMM asymptotic theory (consistency, normality, efficient weighting) then covers OLS, Wald, and 2SLS all at once

Multiple IV: Linear GMM

- Let $P_z = Z(Z'Z)^{-1}Z'$ be the projection matrix
- First stage fitted value then becomes $\hat{X} = P_z X$
- P is idempotent: $P \cdot P = P$
- When $W = (Z'Z)^{-1}$, 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.
- This nesting is not just algebra: the GMM asymptotic theory (consistency, normality, efficient weighting) then covers OLS, Wald, and 2SLS all at once

Multiple IV: Linear GMM

- Let $P_z = Z(Z'Z)^{-1}Z'$ be the projection matrix
- First stage fitted value then becomes $\hat{X} = P_z X$
- P is idempotent: $P \cdot P = P$
- When $W = (Z'Z)^{-1}$, 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.
- This nesting is not just algebra: the GMM asymptotic theory (consistency, normality, efficient weighting) then covers OLS, Wald, and 2SLS all at once

Multiple IV: Linear GMM

- Let $P_z = Z(Z'Z)^{-1}Z'$ be the projection matrix
- First stage fitted value then becomes $\hat{X} = P_z X$
- P is idempotent: $P \cdot P = P$
- When $W = (Z'Z)^{-1}$, 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.
- This nesting is not just algebra: the GMM asymptotic theory (consistency, normality, efficient weighting) then covers OLS, Wald, and 2SLS all at once

Multiple IV: Linear GMM

- Let $P_z = Z(Z'Z)^{-1}Z'$ be the projection matrix
- First stage fitted value then becomes $\hat{X} = P_z X$
- P is idempotent: $P \cdot P = P$
- When $W = (Z'Z)^{-1}$, 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.
- This nesting is not just algebra: the GMM asymptotic theory (consistency, normality, efficient weighting) then covers OLS, Wald, and 2SLS all at once

Multiple IV: Linear GMM

- Let $P_z = Z(Z'Z)^{-1}Z'$ be the projection matrix
- First stage fitted value then becomes $\hat{X} = P_z X$
- P is idempotent: $P \cdot P = P$
- When $W = (Z'Z)^{-1}$, 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.
- This nesting is not just algebra: the GMM asymptotic theory (consistency, normality, efficient weighting) then covers OLS, Wald, and 2SLS all at once

Multiple IV: Linear GMM

- Let $P_z = Z(Z'Z)^{-1}Z'$ be the projection matrix
- First stage fitted value then becomes $\hat{X} = P_z X$
- P is idempotent: $P \cdot P = P$
- When $W = (Z'Z)^{-1}$, 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.
- This nesting is not just algebra: the GMM asymptotic theory (consistency, normality, efficient weighting) then covers OLS, Wald, and 2SLS all at once

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

When J is evaluated with the efficient weight matrix $W = \{\hat{E}[g_i g_i']\}^{-1}$ (or a consistent estimate of it), as $n \to \infty$,

$$J = n\hat{g}_n(\hat{\beta}_{\text{GMM}})' \hat{S}^{-1} \hat{g}_n(\hat{\beta}_{\text{GMM}}) \xrightarrow{d} \chi^2_{r-k}$$

For a prespecified α, let $c = \hat{G}_{\chi^2_{r-k}}(1-\alpha)$, where $\hat{G}_{\chi^2_{r-k}}$ is the χ^2_{r-k} distribution function. $P[J > c | H_0] \to \alpha$ 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)

When J is evaluated with the efficient weight matrix $W = \left(E[g_i g_i']\right)^{-1}$ (or a consistent estimate of it), as $n \to \infty$,

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

For c satisfying $\alpha = 1 - G_{l-k}(c)$, where G_{l-k} is the χ^2_{l-k} distribution function, $P[J > c|H_0] \to \alpha$ 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)

When J is evaluated with the efficient weight matrix $W = \left(E[g_i g_i']\right)^{-1}$ (or a consistent estimate of it), as $n \to \infty$,

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

For c satisfying $\alpha = 1 - G_{l-k}(c)$, where G_{l-k} is the χ^2_{l-k} distribution function, $P[J > c | H_0] \to \alpha$ 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)

When J is evaluated with the efficient weight matrix $W = \left(E[g_i g_i']\right)^{-1}$ (or a consistent estimate of it), as $n \to \infty$,

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

For c satisfying $\alpha = 1 - G_{l-k}(c)$, where G_{l-k} is the χ^2_{l-k} distribution function, $P[J > c | H_0] \to \alpha$ 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 and heterogeneous treatment effect
- $E(g_i) \neq 0 \not\Rightarrow$ 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 $\not\Rightarrow E(g_i) \neq 0$, since this is a specification test
  There can be other reasons why the null is rejected, such as non-linearity and heterogeneous treatment effect
- $E(g_i) \neq 0 \not\Rightarrow$ 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 and heterogeneous treatment effect
- $E(g_i) \neq 0 \nRightarrow$ 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 and heterogeneous treatment effect
- $E(g_i) \neq 0 \nRightarrow$ 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 and heterogeneous treatment effect
- $E(g_i) \neq 0 \not\Rightarrow$ 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 $\not\Rightarrow E(g_i) \neq 0$, since this is a specification test
  There can be other reasons why the null is rejected, such as non-linearity and heterogeneous treatment effect
- $E(g_i) \neq 0 \not\Rightarrow$ 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 $\not\Rightarrow E(g_i) \neq 0$, since this is a specification test
  There can be other reasons why the null is rejected, such as non-linearity and heterogeneous treatment effect
- $E(g_i) \neq 0 \not\Rightarrow$ J-test rejects null
- Actual size in finite-sample is too large (too many rejections)

Next Step Work

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- We have investigated what will happen if we relax only (3), and only (1)+(2)
- In the next class, we will try to relax all three conditions
- We will consider heterogenous treatment effect beyond binary variables and binary IV

Next Step Work

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- We have investigated what will happen if we relax only (3), and only (1)+(2)
- In the next class, we will try to relax all three conditions
- We will consider heterogenous treatment effect beyond binary variables and binary IV

Next Step Work

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- We have investigated what will happen if we relax only (3), and only (1)+(2)
- In the next class, we will try to relax all three conditions
- We will consider heterogenous treatment effect beyond binary variables and binary IV

Next Step Work

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- We have investigated what will happen if we relax only (3), and only (1)+(2)
- In the next class, we will try to relax all three conditions
- We will consider heterogenous treatment effect beyond binary variables and binary IV

Next Step Work

- In the simple IV case, we consider:
  (1) single endogenous variable; (2) single IV; (3) constant treatment effect
- We have investigated what will happen if we relax only (3), and only (1)+(2)
- In the next class, we will try to relax all three conditions
- We will consider heterogenous treatment effect beyond binary variables and binary IV

Oster Bound: Endogeneity without IV

- Coming up with a good IV is super hard
- Unfortunately, we sometimes cannot find a valid instrument
- How to deal with endogeneity without a valid instrument?
- We are going to introduce one of the most popular 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 sometimes cannot find a valid instrument
- How to deal with endogeneity without a valid instrument?
- We are going to introduce one of the most popular 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 sometimes cannot find a valid instrument
- How to deal with endogeneity without a valid instrument?
- We are going to introduce one of the most popular 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 sometimes cannot find a valid instrument
- How to deal with endogeneity without a valid instrument?
- We are going to introduce one of the most popular 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 sometimes cannot find a valid instrument
- How to deal with endogeneity without a valid instrument?
- We are going to introduce one of the most popular 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 sometimes cannot find a valid instrument
- How to deal with endogeneity without a valid instrument?
- We are going to introduce one of the most popular 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 $\Longrightarrow$ Set identification/bound estimation (in a loose way)
- Remember, Oster Bound can only give you suggestive evidence when you have nothing else can do

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 (in a loose way)
- Remember, Oster Bound can only give you suggestive evidence when you have nothing else can do

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 (in a loose way)
- Remember, Oster Bound can only give you suggestive evidence when you have nothing else can do

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 (in a loose way)
- Remember, Oster Bound can only give you suggestive evidence when you have nothing else can do

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 (in a loose way)
- Remember, Oster Bound can only give you suggestive evidence when you have nothing else can do

Oster Bound: Endogeneity without IV

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 observationally 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 observationally 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 observationally 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 observationally 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 observationally 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 observationally 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 observationally 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 observationally 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 observationally 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 observationally 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 observationally 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 observationally 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 observationally 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 observationally 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 observationally 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 observationally 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 observationally 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 observationally equivalent

Oster Bound: Endogeneity without IV

- The intuition of Oster bound is very simple
- We can use observed variables to evaluate how large is the omitted variable bias
- 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 is the omitted variable bias
- 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 is the omitted variable bias
- 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 is the omitted variable bias
- 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:
  a. How stable is the coefficient? (stability)
  b. How much of y is explained by this control? (informative)
- If the coefficient estimation is changed only a little, by a strong, informative 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, informative control
  $\Rightarrow$ 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, informative control

  $\Rightarrow$ 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, informative control

  $\Rightarrow$ 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, informative control

  $\Rightarrow$ 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, informative control

  $\Rightarrow$ 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:

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

- 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:

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

- 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

- 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 less important than the unobserved one
- When $\delta = 1$, the unobservable 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 less important than the unobserved one
- When $\delta = 1$, the unobservable 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 less important than the unobserved one
- When $\delta = 1$, the unobservable 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 less important than the unobserved one
- When $\delta = 1$, the unobservable 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 less important than the unobserved one
- When $\delta = 1$, the unobservable 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 δ and β (just do the regs)
  - We can infer how much variation we explain using observed variables
  - Thus, knowing R_{max} means knowing the portion of variance we can explain by the additional unobserved control W_2

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 $\tilde{R}$ and $\bar{R}$ (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 unobserved control W_2

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 $\tilde{R}$ and $\bar{R}$ (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 unobserved control W_2

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 unobserved control W_2

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 unobserved control W_2

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 unobserved control W_2

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 unobserved control W_2

Oster Bound: Theory

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

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

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

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 targeted 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 targeted 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 targeted 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 theoretical
- We never know what δ and R_{max} are, 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 theoretical
- We never know what δ and R_{max} are, 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 theoretical
- We never know what δ and R_{max} are, 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 theoretical
- We never know what δ and R_{max} are, 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 theoretical
- We never know what δ and R_{max} are, 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 theoretical
- We never know what δ and R_{max} are, 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
  $\Rightarrow$ Our results are robust

Oster Bound: Theory

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

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

Where $\delta = 1$, the delbossed 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

- Propositions 2 and 3 give two equations connecting δ, R_{max} and bias
- They are very complicated (cubic equation)
- 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

- Propositions 2 and 3 give two equations connecting δ, R_{max} and bias
- They are very complicated (cubic equation)
- 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

- Propositions 2 and 3 give two equations connecting δ, R_{max} and bias
- They are very complicated (cubic equation)
- 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

- Propositions 2 and 3 give two equations connecting δ, R_{max} and bias
- They are very complicated (cubic equation)
- 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

- Propositions 2 and 3 give two equations connecting δ, R_{max} and bias
- They are very complicated (cubic equation)
- 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 $R_{max} = \min\{1.3\tilde{R}^2, 1\}$
    - 1.3 is defined to let 90% of the RCT studies in top journals pass this test
  - Set $\delta = \delta_0$ and the corresponding λ
  - If $\beta > 1$, 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
- Stata package: `psacalc`

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
- Stata package: `psacalc`

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
- Stata package: `psacalc`

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
- Stata package: `psacalc`

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
- Stata package: `psacalc`

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
- Stata package: `psacalc`

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
- Stata package: `psacalc`

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
- Stata package: `psacalc`

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
- Stata package: `psacalc`

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
- Stata package: `psacalc`

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
- Stata package: `psacalc`

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 $R_{max} = \min(1.3 \cdot 3\tilde{R}^2, 1)$
  - 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: Problem and Extension

- However, Masten and Poirier (2026) shows further severe issues in this method
- The mapping from δ to β can be discontinuous and multi-valued
- The point of δ to get $\beta = 0 \neq$ The point of δ to get the sign of β flipped!!!
- Explain away breakdown point δ^0 vs Sign change breakdown point δ^{sign}

Oster Bound: Problem and Extension

- However, Masten and Poirier (2026) shows further severe issues in this method
- The mapping from δ to β can be discontinuous and multi-valued
- The point of δ to get $\beta = 0 \neq$ The point of δ to get the sign of β flipped!!!
- Explain away breakdown point δ^0 vs Sign change breakdown point δ^{sign}

Oster Bound: Problem and Extension

- However, Masten and Poirier (2026) shows further severe issues in this method
- The mapping from δ to β can be discontinuous and multi-valued
- The point of δ to get $\beta = 0 \neq$ The point of δ to get the sign of β flipped!!!
- Explain away breakdown point δ^0 vs Sign change breakdown point δ^{sign}

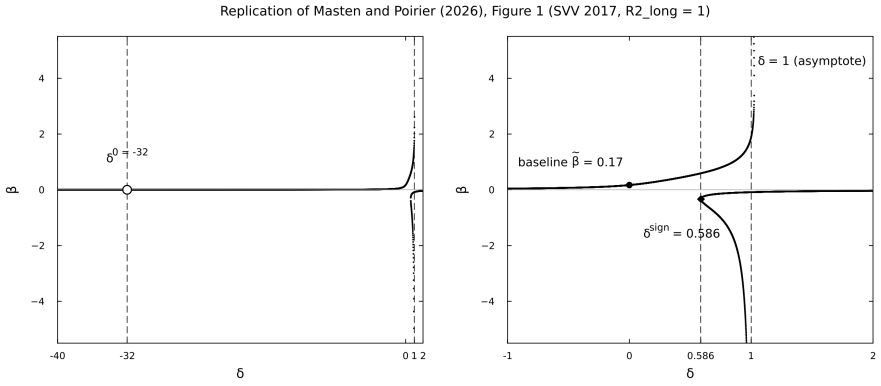
Oster Bound: Problem and Extension

- However, Masten and Poirier (2026) shows further severe issues in this method
- The mapping from δ to β can be discontinuous and multi-valued
- The point of δ to get $\beta = 0 \neq$ The point of δ to get the sign of β flipped!!!
- Explain away breakdown point δ^0 vs Sign change breakdown point δ^{sign}

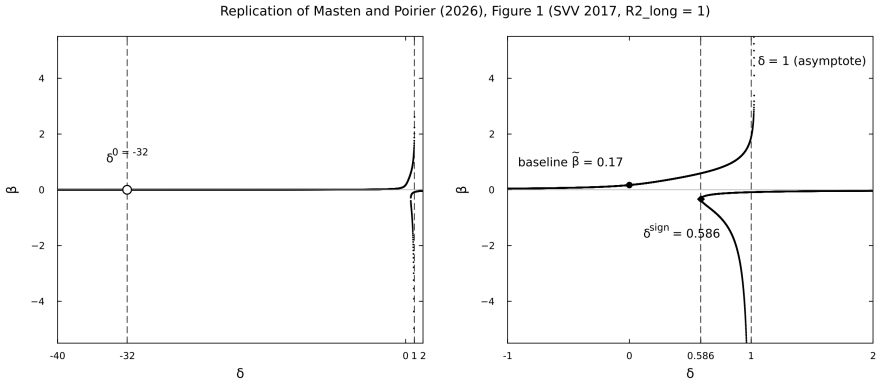
Oster Bound: Problem and Extension

- However, Masten and Poirier (2026) shows further severe issues in this method
- The mapping from δ to β can be discontinuous and multi-valued
- The point of δ to get $\beta = 0 \neq$ The point of δ to get the sign of β flipped!!!
- Explain away breakdown point δ^0 vs Sign change breakdown point δ^{sign}

Oster Bound: Problem and Extension



Oster Bound: Problem and Extension



Oster Bound: Problem and Extension

Oster Bound: Problem and Extension

- Two important issues here
- 1. Sign change breakdown point $|\delta^{sign}| \ll$ Explain away breakdown point $|\delta^{0}|$
  - Hard to erase the effect, but simple to flip the sign!
  - It can be proved that $\delta^{sign} \leq 1$: nothing is robust for rule-of-thumb test!
- 2. The change in β is very very sensitive to δ around $\delta = 1$
  - The bias adjusted point estimates from Prop 1 are fragile
  - A little deviation from our assumption of $\delta = 1$ leads to huge change in β^{*}
  - Why at $\delta = 1$? Colinearity between treatment and covariates could explode the estimation

Oster Bound: Problem and Extension

- Two important issues here
- 1. Sign change breakdown point $|\delta^{sign}| \ll$ Explain away breakdown point $|\delta^{0}|$
  - Hard to erase the effect, but simple to flip the sign!
  - It can be proved that $\delta^{sign} \leq 1$: nothing is robust for rule-of-thumb test!
- 2. The change in β is very very sensitive to δ around $\delta = 1$
  - The bias adjusted point estimates from Prop 1 are fragile
  - A little deviation from our assumption of $\delta = 1$ leads to huge change in δ^{*}
  - Why at $\delta = 1$? Colinearity between treatment and covariates could explode the estimation

Oster Bound: Problem and Extension

- Two important issues here
- 1. Sign change breakdown point $|\delta^{sign}| \ll$ Explain away breakdown point $|\delta^0|$
  - Hard to erase the effect, but simple to flip the sign!
  - It can be proved that $\delta^{sign} \leq 1$: nothing is robust for rule-of-thumb test!
- 2. The change in β is very very sensitive to δ around $\delta = 1$
  - The bias adjusted point estimates from Prop 1 are fragile
  - A little deviation from our assumption of $\delta = 1$ leads to huge change in δ^*
  - Why at $\delta = 1$? Colinearity between treatment and covariates could explode the estimation

Oster Bound: Problem and Extension

- Two important issues here
- 1. Sign change breakdown point $|\delta^{sign}| \ll$ Explain away breakdown point $|\delta^0|$
  - Hard to erase the effect, but simple to flip the sign!
  - It can be proved that $\delta^{sign} \leq 1$: nothing is robust for rule-of-thumb test!
- 2. The change in β is very very sensitive to δ around $\delta = 1$
  - The bias adjusted point estimates from Prop 1 are fragile
  - A little deviation from our assumption of $\delta = 1$ leads to huge change in β^*
  - Why at $\delta = 1$? Colinearity between treatment and covariates could explode the estimation

Oster Bound: Problem and Extension

- Two important issues here
- 1. Sign change breakdown point $|\delta^{sign}| \ll$ Explain away breakdown point $|\delta^{0}|$
  - Hard to erase the effect, but simple to flip the sign!
  - It can be proved that $\delta^{sign} \leq 1$: nothing is robust for rule-of-thumb test!
- 2. The change in β is very very sensitive to δ around $\delta = 1$
  - The bias adjusted point estimates from Prop 1 are fragile
  - A little deviation from our assumption of $\delta = 1$ leads to huge change in β^{*}
  - Why at $\delta = 1$? Colinearity between treatment and covariates could explode the estimation

Oster Bound: Problem and Extension

- Two important issues here
- 1. Sign change breakdown point $|\delta^{sign}| \ll$ Explain away breakdown point $|\delta^{0}|$
  - Hard to erase the effect, but simple to flip the sign!
  - It can be proved that $\delta^{sign} \leq 1$: nothing is robust for rule-of-thumb test!
- 2. The change in β is very very sensitive to δ around $\delta = 1$
  - The bias adjusted point estimates from Prop 1 are fragile
  - A little deviation from our assumption of $\delta = 1$ leads to huge change in β^{*}
  - Why at $\delta = 1$? Colinearity between treatment and covariates could explode the estimation

Oster Bound: Problem and Extension

- Two important issues here
- 1. Sign change breakdown point $|\delta^{sign}| \ll$ Explain away breakdown point $|\delta^{0}|$
  - Hard to erase the effect, but simple to flip the sign!
  - It can be proved that $\delta^{sign} \leq 1$: nothing is robust for rule-of-thumb test!
- 2. The change in β is very very sensitive to δ around $\delta = 1$
  - The bias adjusted point estimates from Prop 1 are fragile
  - A little deviation from our assumption of $\delta = 1$ leads to huge change in β^{*}
  - Why at $\delta = 1$? Colinearity between treatment and covariates could explode the estimation

Oster Bound: Problem and Extension

- Two important issues here
- 1. Sign change breakdown point $|\delta^{sign}| \ll$ Explain away breakdown point $|\delta^{0}|$
  - Hard to erase the effect, but simple to flip the sign!
  - It can be proved that $\delta^{sign} \leq 1$: nothing is robust for rule-of-thumb test!
- 2. The change in β is very very sensitive to δ around $\delta = 1$
  - The bias adjusted point estimates from Prop 1 are fragile
  - A little deviation from our assumption of $\delta = 1$ leads to huge change in β^{*}
  - Why at $\delta = 1$? Colinearity between treatment and covariates could explode the estimation

Oster Bound: Problem and Extension

- Proposed fix: Consider both δ^{sign} and δ^{0}
- (1) For test based on Prop 2, report both δ^{sign} and δ^{0}, add magnitude restrictions: $|\beta - \tilde{\beta}| \leq M$
  - Bound the bias by M
  - Now β^{sign} can exceed 1 (see Prop 2 test feasible)
- (2) For bias-corrected estimates, report sets, not points
  - Choose a $\hat{\beta}$, plot the cumulative identified set of β over $|\beta| \leq \beta$
  - Draw the M restriction as horizontal cutoffs
- A new Stata package for this: `regsensitivity`

Oster Bound: Problem and Extension

- Proposed fix: Consider both δ^{sign} and δ^0
- (1) For test based on Prop 2, report both δ^{sign} and δ^0, add magnitude restrictions: $|\beta - \tilde{\beta}| \leq M$
  - Bound the bias by M
  - Now δ^{sign} can exceed 1 $\Longrightarrow$ Prop 2 test feasible
- (2) For bias-corrected estimates, report sets, not points
  - Choose a $\bar{\delta}$, plot the cumulative identified set of β over $|\delta| \leq \bar{\delta}$
  - Draw the M restriction as horizontal cutoffs
- A new Stata package for this: `regsensitivity`

Oster Bound: Problem and Extension

- Proposed fix: Consider both δ^{sign} and δ^0
- (1) For test based on Prop 2, report both δ^{sign} and δ^0, add magnitude restrictions: $|\beta - \tilde{\beta}| \leq M$
  - Bound the bias by M
  - Now δ^{sign} can exceed 1 $\Rightarrow$ Prop 2 test feasible
- (2) For bias-corrected estimates, report sets, not points
  - Choose a $\bar{\delta}$, plot the cumulative identified set of β over $|\delta| \leq \bar{\delta}$
  - Draw the M restriction as horizontal cutoffs
- A new Stata package for this: `regsensitivity`

Oster Bound: Problem and Extension

- Proposed fix: Consider both δ^{sign} and δ^0
- (1) For test based on Prop 2, report both δ^{sign} and δ^0, add magnitude restrictions: $|\beta - \tilde{\beta}| \leq M$
  - Bound the bias by M
  - Now δ^{sign} can exceed 1 $\Rightarrow$ Prop 2 test feasible
- (2) For bias-corrected estimates, report sets, not points
  - Choose a $\bar{\delta}$, plot the cumulative identified set of β over $|\delta| \leq \bar{\delta}$
  - Draw the M restriction as horizontal cutoffs
- A new Stata package for this: `regsensitivity`

Oster Bound: Problem and Extension

- Proposed fix: Consider both δ^{sign} and δ^0
- (1) For test based on Prop 2, report both δ^{sign} and δ^0, add magnitude restrictions: $|\beta - \tilde{\beta}| \leq M$
  - Bound the bias by M
  - Now δ^{sign} can exceed 1 $\Rightarrow$ Prop 2 test feasible
- (2) For bias-corrected estimates, report sets, not points
  - Choose a $\bar{\delta}$, plot the cumulative identified set of β over $|\delta| \leq \bar{\delta}$
  - Draw the M restriction as horizontal cutoffs
- A new Stata package for this: `regsensitivity`

Oster Bound: Problem and Extension

- Proposed fix: Consider both δ^{sign} and δ^0
- (1) For test based on Prop 2, report both δ^{sign} and δ^0, add magnitude restrictions: $|\beta - \tilde{\beta}| \leq M$
  - Bound the bias by M
  - Now δ^{sign} can exceed 1 $\Rightarrow$ Prop 2 test feasible
- (2) For bias-corrected estimates, report sets, not points
  - Choose a $\bar{\delta}$, plot the cumulative identified set of β over $|\delta| \leq \bar{\delta}$
  - Draw the M restriction as horizontal cutoffs
- A new Stata package for this: `regsensitivity`

Oster Bound: Problem and Extension

- Proposed fix: Consider both δ^{sign} and δ^0
- (1) For test based on Prop 2, report both δ^{sign} and δ^0, add magnitude restrictions: $|\beta - \tilde{\beta}| \leq M$
  - Bound the bias by M
  - Now δ^{sign} can exceed 1 $\Rightarrow$ Prop 2 test feasible
- (2) For bias-corrected estimates, report sets, not points
  - Choose a $\bar{\delta}$, plot the cumulative identified set of β over $|\delta| \leq \bar{\delta}$
  - Draw the M restriction as horizontal cutoffs
- A new Stata package for this: `regsensitivity`

Oster Bound: Problem and Extension

- Proposed fix: Consider both δ^{sign} and δ^0
- (1) For test based on Prop 2, report both δ^{sign} and δ^0, add magnitude restrictions: $|\beta - \tilde{\beta}| \leq M$
  - Bound the bias by M
  - Now δ^{sign} can exceed 1 $\Rightarrow$ Prop 2 test feasible
- (2) For bias-corrected estimates, report sets, not points
  - Choose a $\bar{\delta}$, plot the cumulative identified set of β over $|\delta| \leq \bar{\delta}$
  - Draw the M restriction as horizontal cutoffs
- A new Stata package for this: `regsensitivity`

Oster Bound: Problem and Extension

- Proposed fix: Consider both δ^{sign} and δ^{0}
- (1) For test based on Prop 2, report both δ^{sign} and δ^{0}, add magnitude restrictions: $|\beta - \tilde{\beta}| \leq M$
  - Bound the bias by M
  - Now δ^{sign} can exceed 1 $\Rightarrow$ Prop 2 test feasible
- (2) For bias-corrected estimates, report sets, not points
  - Choose a $\bar{\delta}$, plot the cumulative identified set of β over $|\delta| \leq \bar{\delta}$
  - Draw the M restriction as horizontal cutoffs
- A new Stata package for this: `regsensitivity`

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
  a. The choice of parameters are arbitrary
  b. The results may be sensitive to the parameters
  c. It can only give you a sense of the robustness of your results
- Applications in Economics: Clark et al. (2021); Satyanath, Voigtländer, and Voth (2017)

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
  - The results may be sensitive to the parameters
  - It can only give you a sense of the robustness of your results
- Applications in Economics: Clark et al. (2021); Satyanath, Voigtländer, and Voth (2017)

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
  - The results may be sensitive to the parameters
  - It can only give you a sense of the robustness of your results
- Applications in Economics: Clark et al. (2021); Satyanath, Voigtländer, and Voth (2017)

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
  - The results may be sensitive to the parameters
  - It can only give you a sense of the robustness of your results
- Applications in Economics: Clark et al. (2021); Satyanath, Voigtländer, and Voth (2017)

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
  - The results may be sensitive to the parameters
  - It can only give you a sense of the robustness of your results
- Applications in Economics: Clark et al. (2021); Satyanath, Voigtländer, and Voth (2017)

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
  - The results may be sensitive to the parameters
  - It can only give you a sense of the robustness of your results
- Applications in Economics: Clark et al. (2021); Satyanath, Voigtländer, and Voth (2017)

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
  - The results may be sensitive to the parameters
  - It can only give you a sense of the robustness of your results
- Applications in Economics: Clark et al. (2021); Satyanath, Voigtländer, and Voth (2017)

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
  - The results may be sensitive to the parameters
  - It can only give you a sense of the robustness of your results
- Applications in Economics: Clark et al. (2021); Satyanath, Voigtländer, and Voth (2017)

Final Conclusion

- IV is the main strategy we can use to deal with endogeneity
- At least two assumptions: Exclusion restriction, Existence of first stage
- With heterogeneous treatment effects, the 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
- With heterogeneous treatment effects, the 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
- With heterogeneous treatment effects, the 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
- With heterogeneous treatment effects, the 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
- With heterogeneous treatment effects, the 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
- With heterogeneous treatment effects, the 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 accurate, 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 accurate, 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 accurate, 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 accurate, 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 accurate, since some parameter choices are pretty arbitrary

References

Angrist, Joshua D. 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records." *American Economic Review* 80 (3):313–336.

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.

Imbens, Guido W and Joshua D Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2):467–475.

Masten, Matthew A and Alexandre Poirier. 2026. "The Effect of Omitted Variables on the Sign of Regression Coefficients." *American Economic Review* 116 (7):2685–2710.

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* .

Satyanath, Shanker, Nico Voigtländer, and Hans-Joachim Voth. 2017. "Bowling for Fascism: Social Capital and the Rise of the Nazi Party." *Journal of Political Economy* 125 (2):478–526.