

# Inferencia, Causalidad y Políticas Públicas

## ECO-60116

Week 03: Variables Instrumentales

Eduard F. Martinez Gonzalez, Ph.D.

Departamento de Economía, Universidad Icesi

September 25, 2025

# Roadmap

## Recap Week-02

- **Experimentos Aleatorios (RCTs):** el estándar de oro para estimar efectos causales.
  - ▶ Asignación aleatoria asegura comparabilidad entre grupos.
  - ▶ Validez interna fuerte, aunque limitada validez externa.

## Recap Week-02

- **Experimentos Aleatorios (RCTs):** el estándar de oro para estimar efectos causales.
  - ▶ Asignación aleatoria asegura comparabilidad entre grupos.
  - ▶ Validez interna fuerte, aunque limitada validez externa.
- **Validez de los RCTs**
  - ▶ Alta *validez interna*.
  - ▶ Posibles limitaciones en *validez externa*.

# Recap Week-02

- **Experimentos Aleatorios (RCTs):** el estándar de oro para estimar efectos causales.
  - ▶ Asignación aleatoria asegura comparabilidad entre grupos.
  - ▶ Validez interna fuerte, aunque limitada validez externa.
- **Validez de los RCTs**
  - ▶ Alta *validez interna*.
  - ▶ Posibles limitaciones en *validez externa*.
- **Amenazas a la validez:**
  - ▶ Incumplimiento en la asignación.
  - ▶ Attrition (deserción de participantes).
  - ▶ Spillovers o contaminación entre grupos.

## Recap Week-02

- **Experimentos Aleatorios (RCTs):** el estándar de oro para estimar efectos causales.
  - ▶ Asignación aleatoria asegura comparabilidad entre grupos.
  - ▶ Validez interna fuerte, aunque limitada validez externa.
- **Validez de los RCTs**
  - ▶ Alta *validez interna*.
  - ▶ Posibles limitaciones en *validez externa*.
- **Amenazas a la validez:**
  - ▶ Incumplimiento en la asignación.
  - ▶ Attrition (deserción de participantes).
  - ▶ Spillovers o contaminación entre grupos.
- **Ejemplo aplicado:** incentivos para mejorar asistencia de maestros en India (Duflo et al., 2012).

# Roadmap

# Motivación

- Los RCTs son el estándar de oro, pero no siempre son factibles:
  - ▶ Costos altos o barreras éticas.
  - ▶ Incumplimiento en la asignación (non-compliance).
  - ▶ Disponibilidad limitada de programas o tratamientos.



# Motivación

- Los RCTs son el estándar de oro, pero no siempre son factibles:
  - ▶ Costos altos o barreras éticas.
  - ▶ Incumplimiento en la asignación (non-compliance).
  - ▶ Disponibilidad limitada de programas o tratamientos.
- En muchos contextos de política pública debemos trabajar con datos observacionales.

# Motivación

- Los RCTs son el estándar de oro, pero no siempre son factibles:
  - ▶ Costos altos o barreras éticas.
  - ▶ Incumplimiento en la asignación (non-compliance).
  - ▶ Disponibilidad limitada de programas o tratamientos.
- En muchos contextos de política pública debemos trabajar con datos observacionales.
- Problema central: **endogeneidad**
  - ▶ Variables omitidas.
  - ▶ Causalidad inversa.
  - ▶ Error de medición.

# Motivación

- Los RCTs son el estándar de oro, pero no siempre son factibles:
  - ▶ Costos altos o barreras éticas.
  - ▶ Incumplimiento en la asignación (non-compliance).
  - ▶ Disponibilidad limitada de programas o tratamientos.
- En muchos contextos de política pública debemos trabajar con datos observacionales.
- Problema central: **endogeneidad**
  - ▶ Variables omitidas.
  - ▶ Causalidad inversa.
  - ▶ Error de medición.
- **Idea de IV:** aprovechar una fuente de variación exógena que afecte la variable endógena de interés, pero no al resultado directamente.

# Roadmap

# Problema de Endogeneidad: Educación y Salarios

- **Contexto:** Queremos estimar el efecto de la educación ( $X_i$ ) sobre los ingresos ( $Y_i$ ):

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

# Problema de Endogeneidad: Educación y Salarios

- **Contexto:** Queremos estimar el efecto de la educación ( $X_i$ ) sobre los ingresos ( $Y_i$ ):

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- **Dificultad:**  $X_i$  (años de escolaridad) puede estar correlacionado con factores no observados ( $\varepsilon_i$ ).

# Problema de Endogeneidad: Educación y Salarios

- **Contexto:** Queremos estimar el efecto de la educación ( $X_i$ ) sobre los ingresos ( $Y_i$ ):

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- **Dificultad:**  $X_i$  (años de escolaridad) puede estar correlacionado con factores no observados ( $\varepsilon_i$ ).
  - ▶ Habilidad o motivación influyen en educación y salarios.
  - ▶ Expectativas de ingresos afectan la decisión de estudiar.

# Problema de Endogeneidad: Educación y Salarios

- **Contexto:** Queremos estimar el efecto de la educación ( $X_i$ ) sobre los ingresos ( $Y_i$ ):

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- **Dificultad:**  $X_i$  (años de escolaridad) puede estar correlacionado con factores no observados ( $\varepsilon_i$ ).
  - ▶ Habilidad o motivación influyen en educación y salarios.
  - ▶ Expectativas de ingresos afectan la decisión de estudiar.
- **Consecuencia:** El estimador OLS de  $\beta$  es sesgado e inconsistente.



# Problema de Endogeneidad: Educación y Salarios

- **Contexto:** Queremos estimar el efecto de la educación ( $X_i$ ) sobre los ingresos ( $Y_i$ ):

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- **Dificultad:**  $X_i$  (años de escolaridad) puede estar correlacionado con factores no observados ( $\varepsilon_i$ ).
  - ▶ Habilidad o motivación influyen en educación y salarios.
  - ▶ Expectativas de ingresos afectan la decisión de estudiar.
- **Consecuencia:** El estimador OLS de  $\beta$  es sesgado e inconsistente.
- **¿Cómo responder a la endogeneidad?**

# Problema de Endogeneidad: Educación y Salarios

- **Contexto:** Queremos estimar el efecto de la educación ( $X_i$ ) sobre los ingresos ( $Y_i$ ):

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- **Dificultad:**  $X_i$  (años de escolaridad) puede estar correlacionado con factores no observados ( $\varepsilon_i$ ).
  - ▶ Habilidad o motivación influyen en educación y salarios.
  - ▶ Expectativas de ingresos afectan la decisión de estudiar.
- **Consecuencia:** El estimador OLS de  $\beta$  es sesgado e inconsistente.
- **¿Cómo responder a la endogeneidad?**
  - ▶ ¿Agregar más variables explicativas?

# Problema de Endogeneidad: Educación y Salarios

- **Contexto:** Queremos estimar el efecto de la educación ( $X_i$ ) sobre los ingresos ( $Y_i$ ):

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- **Dificultad:**  $X_i$  (años de escolaridad) puede estar correlacionado con factores no observados ( $\varepsilon_i$ ).
  - ▶ Habilidad o motivación influyen en educación y salarios.
  - ▶ Expectativas de ingresos afectan la decisión de estudiar.
- **Consecuencia:** El estimador OLS de  $\beta$  es sesgado e inconsistente.
- **¿Cómo responder a la endogeneidad?**
  - ▶ ¿Agregar más variables explicativas?
  - ▶ Diseños cuasi-experimentales: aprovechar variación externa o reglas institucionales. **Hablaremos de esto después...**

# Problema de Endogeneidad: Educación y Salarios

- **Contexto:** Queremos estimar el efecto de la educación ( $X_i$ ) sobre los ingresos ( $Y_i$ ):

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- **Dificultad:**  $X_i$  (años de escolaridad) puede estar correlacionado con factores no observados ( $\varepsilon_i$ ).
  - ▶ Habilidad o motivación influyen en educación y salarios.
  - ▶ Expectativas de ingresos afectan la decisión de estudiar.
- **Consecuencia:** El estimador OLS de  $\beta$  es sesgado e inconsistente.
- **¿Cómo responder a la endogeneidad?**
  - ▶ ¿Agregar más variables explicativas?
  - ▶ Diseños cuasi-experimentales: aprovechar variación externa o reglas institucionales. **Hablaremos de esto después...**

## Siguiente paso

Hoy introduciremos una estrategia clave: **Variables Instrumentales (IV)** → una fuente de variación en  $X$  exógena respecto a  $\varepsilon$ .

# Idea de Variables Instrumentales (IV)

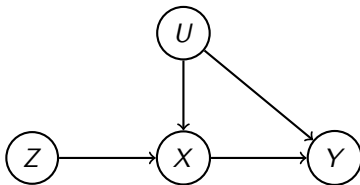
Permite identificar el efecto causal de  $X$  sobre  $Y$  a pesar de la endogeneidad.

# Idea de Variables Instrumentales (IV)

Permite identificar el efecto causal de  $X$  sobre  $Y$  a pesar de la endogeneidad.

## Definición

Una **variable instrumental**  $Z$  es una fuente de variación en  $X$  que cumple con la condición de **relevancia** ( $Z$  debe estar correlacionada con  $X$ ) y **exclusión** ( $Z$  no debe afectar  $Y$  excepto a través de  $X$ ).

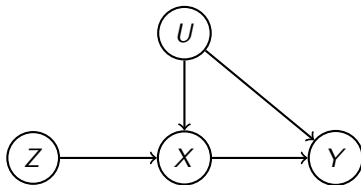


# Idea de Variables Instrumentales (IV)

Permite identificar el efecto causal de  $X$  sobre  $Y$  a pesar de la endogeneidad.

## Definición

Una **variable instrumental**  $Z$  es una fuente de variación en  $X$  que cumple con la condición de **relevancia** ( $Z$  debe estar correlacionada con  $X$ ) y **exclusión** ( $Z$  no debe afectar  $Y$  excepto a través de  $X$ ).



## Intuición

El instrumento genera variación “como si fuera aleatoria” en  $X$ , y nos permite estimar  $\beta$  sin que la correlación  $X \leftrightarrow \varepsilon$  sesgue el resultado.

# Supuestos: Condición de Relevancia

## Definición conceptual

El instrumento  $Z$  debe estar **fuertemente correlacionado** con la variable endógena  $X$ :

$$\text{Cov}(Z_i, X_i) \neq 0$$



# Supuestos: Condición de Relevancia

## Definición conceptual

El instrumento  $Z$  debe estar **fuertemente correlacionado** con la variable endógena  $X$ :

$$\text{Cov}(Z_i, X_i) \neq 0$$

## ¿Por qué importa?

- Si  $Z$  no explica suficiente variación en  $X$ , el instrumento es **débil**.
- Con instrumentos débiles:
  - ▶ Los estimadores IV/2SLS son imprecisos y pueden estar sesgados hacia OLS.
  - ▶ Los intervalos de confianza se vuelven poco informativos.
- En la práctica, se requieren instrumentos que muevan  $X$  de manera sustancial.

# Supuestos: Condición de Relevancia

## Definición conceptual

El instrumento  $Z$  debe estar **fuertemente correlacionado** con la variable endógena  $X$ :

$$\text{Cov}(Z_i, X_i) \neq 0$$

## ¿Por qué importa?

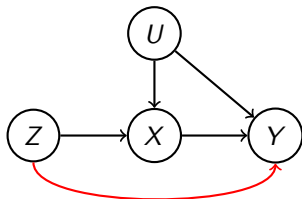
- Si  $Z$  no explica suficiente variación en  $X$ , el instrumento es **débil**.
- Con instrumentos débiles:
  - ▶ Los estimadores IV/2SLS son imprecisos y pueden estar sesgados hacia OLS.
  - ▶ Los intervalos de confianza se vuelven poco informativos.
- En la práctica, se requieren instrumentos que muevan  $X$  de manera sustancial.

## Ejemplo intuitivo

Un instrumento que apenas cambia un poco la educación (por ejemplo, un incentivo mínimo) difícilmente servirá para estimar su efecto en los ingresos.

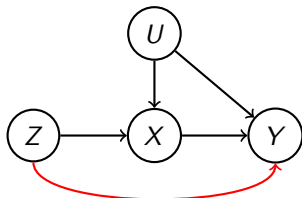
## Supuestos: Condición de Exclusión

**Definición:** El instrumento  $Z$  no debe afectar el resultado  $Y$  por vías distintas a  $X$ :  $\text{Cov}(Z_i, \varepsilon_i) = 0$



# Supuestos: Condición de Exclusión

**Definición:** El instrumento  $Z$  no debe afectar el resultado  $Y$  por vías distintas a  $X$ :  $\text{Cov}(Z_i, \varepsilon_i) = 0$



## Implicaciones prácticas:

- Si  $Z$  afecta  $Y$  por otra vía distinta de  $X$ , el estimador IV identifica un efecto contaminado (violación de exclusión).
- No hay un test concluyente que pruebe la condición de exclusión (no se considera que Sargan/Hansen prueben exogenidad).
- Se debe argumentar que no existe un camino causal directo  $Z \rightarrow Y$  (ni a través de factores omitidos).

# Roadmap

## 2SLS: idea general

**Objetivo:** Identificar el efecto causal de  $X$  sobre  $Y$  usando sólo la variación de  $X$  inducida por  $Z$ .

$$\underbrace{Y_i}_{\text{outcome}} = \beta_0 + \beta_1 \underbrace{X_i}_{\text{endógena}} + W_i' \beta + u_i, \quad \underbrace{X_i}_{\text{endógena}} = \alpha_0 + \alpha_1 \underbrace{Z_i}_{\text{instrumento}} + W_i' \alpha + \nu_i$$

## 2SLS: idea general

**Objetivo:** Identificar el efecto causal de  $X$  sobre  $Y$  usando sólo la variación de  $X$  inducida por  $Z$ .

$$\underbrace{Y_i}_{\text{outcome}} = \beta_0 + \beta_1 \underbrace{X_i}_{\text{endógena}} + W_i' \beta + u_i, \quad \underbrace{X_i}_{\text{endógena}} = \alpha_0 + \alpha_1 \underbrace{Z_i}_{\text{instrumento}} + W_i' \alpha + \nu_i$$

- **Primera etapa:** aísla la parte de  $X$  explicada por  $Z$  (y  $W$ )  $\Rightarrow \hat{X}_i$ .
- **Segunda etapa:** regresa  $Y$  sobre  $\hat{X}$  (y  $W$ )  $\Rightarrow$  estima  $\beta_1$  con variación exógena.
- **Claves:** (i) misma muestra en ambas etapas, (ii) mismos controles  $W$ , (iii) usar un estimador IV (no OLS con  $\hat{X}$  "a mano").

# Primera etapa



variación exógena de  $X$

## Regresión de $X$ sobre $Z$ y controles

$$X_i = \alpha_0 + \alpha_1 Z_i + W_i' \alpha + \nu_i \implies \hat{X}_i = \hat{\alpha}_0 + \hat{\alpha}_1 Z_i + W_i' \hat{\alpha}$$



# Primera etapa



variación exógena de  $X$

## Regresión de $X$ sobre $Z$ y controles

$$X_i = \alpha_0 + \alpha_1 Z_i + W_i' \alpha + \nu_i \implies \hat{X}_i = \hat{\alpha}_0 + \hat{\alpha}_1 Z_i + W_i' \hat{\alpha}$$

## Interpretación

- $\hat{X}_i$  = parte de  $X_i$  explicada por el instrumento  $Z$  (y los controles  $W$ ).
- Equivalente a “filtrar”  $X$ : quitamos la variación de  $X$  que está contaminada por  $\varepsilon$ .
- Nos quedamos sólo con la variación exógena de  $X$  inducida por  $Z$ .

# Primera etapa



variación exógena de  $X$

## Regresión de $X$ sobre $Z$ y controles

$$X_i = \alpha_0 + \alpha_1 Z_i + W_i' \alpha + \nu_i \implies \hat{X}_i = \hat{\alpha}_0 + \hat{\alpha}_1 Z_i + W_i' \hat{\alpha}$$

## Interpretación

- $\hat{X}_i$  = parte de  $X_i$  explicada por el instrumento  $Z$  (y los controles  $W$ ).
- Equivalente a “filtrar”  $X$ : quitamos la variación de  $X$  que está contaminada por  $\varepsilon$ .
- Nos quedamos sólo con la variación exógena de  $X$  inducida por  $Z$ .

## Diagnósticos de fuerza del instrumento

- $\hat{\alpha}_1$  significativo y con el signo esperado.
- **$F$  de instrumentos excluidos** (regla rápida:  $F > 10$ ).
- (Opc.)  $R^2$  *parcial* de  $Z$  dado  $W$ .

## Segunda etapa



Regresión de  $Y$  sobre  $\hat{X}$  y controles

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + W_i' \beta + u_i$$

### Interpretación

- $\hat{\beta}_1$  mide el efecto causal de  $X$  sobre  $Y$ , usando sólo la variación en  $X$  inducida por  $Z$ .

## Segunda etapa



### Regresión de $Y$ sobre $\hat{X}$ y controles

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + W_i' \beta + u_i$$

### Interpretación

- $\hat{\beta}_1$  mide el efecto causal de  $X$  sobre  $Y$ , usando sólo la variación en  $X$  inducida por  $Z$ .

### Buenas prácticas

- Estimar con un **comando IV/2SLS** (no OLS “a mano” con  $\hat{X}$ ).
- Usar **los mismos controles**  $W$  y la misma muestra que en la 1ª etapa.
- Reportar *primera etapa*, *forma reducida* y *2SLS* (consistencia de signos).
- Utilizar errores estándar robustos / *clustered* acorde al diseño.

# Forma reducida



Regresión de  $Y$  sobre  $Z$  y controles

$$Y_i = \gamma_0 + \gamma_1 Z_i + W_i' \gamma + \eta_i$$

# Forma reducida



## Regresión de $Y$ sobre $Z$ y controles

$$Y_i = \gamma_0 + \gamma_1 Z_i + W_i' \gamma + \eta_i$$

### Interpretación

- $\hat{\gamma}_1$  captura el efecto total de  $Z$  sobre  $Y$ .
- Es la relación directa  $Z \rightarrow Y$  (sin pasar por  $X$  explícitamente).
- Permite tener una visión más transparente, en una sola regresión (si esta regresión corta no da significativa, es poco probable que el estimador de primera etapa salga significativo).
- A veces la forma reducida tiene una interpretación que es interesante en sí misma.

# Roadmap

# El Estimador de Wald

Cuando  $Z$  es binario:

$$\hat{\beta}^{Wald} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[X|Z=1] - \mathbb{E}[X|Z=0]}$$

- **Numerador:** efecto total de  $Z$  sobre  $Y$  (*forma reducida*).
- **Denominador:** efecto de  $Z$  sobre  $X$  (*primera etapa*).
- **Cociente:** efecto de  $X$  sobre  $Y$  usando sólo la variación inducida por  $Z$ :

$$\frac{\text{Efecto de } Z \text{ en } Y}{\text{Efecto de } Z \text{ en } X}$$

## Ejemplo intuitivo

La edad de entrada y salida de la escuela, determinada por reglas externas, genera diferencias en años de educación independientes de la habilidad individual.



# Forma reducida y conexión con Wald

**Forma reducida (efecto de  $Z$  sobre  $Y$ ):**

$$Y_i = \gamma_0 + \gamma_1 Z_i + W_i' \gamma + \eta_i$$

**Primera etapa:**

$$X_i = \alpha_0 + \alpha_1 Z_i + W_i' \alpha + \nu_i$$

**Wald (para  $Z$  binaria):**

$$\hat{\beta}^{Wald} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[X|Z=1] - \mathbb{E}[X|Z=0]}$$

# Forma reducida y conexión con Wald

**Forma reducida (efecto de  $Z$  sobre  $Y$ ):**

$$Y_i = \gamma_0 + \gamma_1 Z_i + W_i' \gamma + \eta_i$$

**Primera etapa:**

$$X_i = \alpha_0 + \alpha_1 Z_i + W_i' \alpha + \nu_i$$

**Wald (para  $Z$  binaria):**

$$\hat{\beta}^{Wald} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[X|Z=1] - \mathbb{E}[X|Z=0]}$$

**Relación con 2SLS**

- Con un solo instrumento binario y sin controles, Wald y 2SLS coinciden.
- Con controles y/o múltiples instrumentos, 2SLS generaliza la misma intuición:

$$\hat{\beta}_{2SLS} \approx \frac{\hat{\gamma}_1}{\hat{\alpha}_1}$$

# Forma reducida y conexión con Wald

**Forma reducida (efecto de  $Z$  sobre  $Y$ ):**

$$Y_i = \gamma_0 + \gamma_1 Z_i + W_i' \gamma + \eta_i$$

**Primera etapa:**

$$X_i = \alpha_0 + \alpha_1 Z_i + W_i' \alpha + \nu_i$$

**Wald (para  $Z$  binaria):**

$$\hat{\beta}^{Wald} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[X|Z=1] - \mathbb{E}[X|Z=0]}$$

**Relación con 2SLS**

- Con un solo instrumento binario y sin controles, Wald y 2SLS coinciden.
- Con controles y/o múltiples instrumentos, 2SLS generaliza la misma intuición:

$$\hat{\beta}_{2SLS} \approx \frac{\hat{\gamma}_1}{\hat{\alpha}_1}$$

# Roadmap

## Ejemplo alternativo: Transporte escolar gratuito

**Pregunta:** ¿Cuál es el retorno de la educación secundaria en los ingresos laborales?

## Ejemplo alternativo: Transporte escolar gratuito

**Pregunta:** ¿Cuál es el retorno de la educación secundaria en los ingresos laborales?

**Problema:** La decisión de continuar en secundaria está correlacionada con factores no observados ( $U$ ), como motivación o apoyo familiar.

## Ejemplo alternativo: Transporte escolar gratuito

**Pregunta:** ¿Cuál es el retorno de la educación secundaria en los ingresos laborales?

**Problema:** La decisión de continuar en secundaria está correlacionada con factores no observados ( $U$ ), como motivación o apoyo familiar.

**Instrumento:** Disponibilidad de transporte escolar gratuito en el municipio ( $Z$ ).

- **Relevancia:** facilita la asistencia y aumenta los años de estudio ( $X$ ).
- **Exclusión:** no afecta los ingresos futuros directamente, salvo a través de la educación.

## Ejemplo alternativo: Transporte escolar gratuito

**Pregunta:** ¿Cuál es el retorno de la educación secundaria en los ingresos laborales?

**Problema:** La decisión de continuar en secundaria está correlacionada con factores no observados ( $U$ ), como motivación o apoyo familiar.

**Instrumento:** Disponibilidad de transporte escolar gratuito en el municipio ( $Z$ ).

- **Relevancia:** facilita la asistencia y aumenta los años de estudio ( $X$ ).
- **Exclusión:** no afecta los ingresos futuros directamente, salvo a través de la educación.

### Interpretación

El estimador IV mide el efecto de la educación sobre ingresos para los jóvenes cuya decisión de estudiar cambia gracias a la política de transporte gratuito (**cumplidores**).



# Ejemplo: Transporte escolar gratuito (modelo)

## Primera etapa (Relevancia):

$$X_i = \alpha_0 + \alpha_1 Z_i + W_i' \alpha + \nu_i$$

- $X_i$ : años de educación secundaria completados.
- $Z_i$ : acceso a transporte escolar gratuito.
- $\hat{\alpha}_1 > 0$ : el transporte aumenta la probabilidad de terminar la secundaria.

# Ejemplo: Transporte escolar gratuito (modelo)

## Primera etapa (Relevancia):

$$X_i = \alpha_0 + \alpha_1 Z_i + W_i' \alpha + \nu_i$$

- $X_i$ : años de educación secundaria completados.
- $Z_i$ : acceso a transporte escolar gratuito.
- $\hat{\alpha}_1 > 0$ : el transporte aumenta la probabilidad de terminar la secundaria.

## Forma reducida:

$$Y_i = \gamma_0 + \gamma_1 Z_i + W_i' \gamma + \eta_i$$

- $Y_i$ : ingresos laborales en la adultez.
- $\hat{\gamma}_1$ : efecto total de  $Z$  sobre ingresos.

# Ejemplo: Transporte escolar gratuito (modelo)

## Primera etapa (Relevancia):

$$X_i = \alpha_0 + \alpha_1 Z_i + W_i' \alpha + \nu_i$$

- $X_i$ : años de educación secundaria completados.
- $Z_i$ : acceso a transporte escolar gratuito.
- $\hat{\alpha}_1 > 0$ : el transporte aumenta la probabilidad de terminar la secundaria.

## Forma reducida:

$$Y_i = \gamma_0 + \gamma_1 Z_i + W_i' \gamma + \eta_i$$

- $Y_i$ : ingresos laborales en la adultez.
- $\hat{\gamma}_1$ : efecto total de  $Z$  sobre ingresos.

## Segunda etapa (2SLS):

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + W_i' \beta + u_i$$

- $\hat{X}_i$ : parte de la educación explicada solo por  $Z$ .
- $\hat{\beta}_1$ : efecto causal local de la educación sobre ingresos para los jóvenes cuya decisión cambia gracias al transporte (*cumplidores*).

## Ejemplo numérico (Wald): Transporte escolar como $Z$

**Supongamos (cifras ilustrativas):**

$$\hat{\alpha}_1 = 0.50 \quad (\text{años}) \quad \hat{\gamma}_1 = 1000 \quad (\text{USD/año})$$

# Ejemplo numérico (Wald): Transporte escolar como $Z$

Supongamos (cifras ilustrativas):

$$\hat{\alpha}_1 = 0.50 \quad (\text{años}) \quad \hat{\gamma}_1 = 1000 \quad (\text{USD/año})$$

## Interpretación de etapas

- **Primera etapa** ( $Z \rightarrow X$ ): acceso a transporte escolar gratuito aumenta en 0.50 años la educación secundaria completada.
- **Forma reducida** ( $Z \rightarrow Y$ ): acceso a transporte incrementa el ingreso anual en 1000 USD.

## Ejemplo numérico (Wald): Transporte escolar como $Z$

Supongamos (cifras ilustrativas):

$$\hat{\alpha}_1 = 0.50 \quad (\text{años}) \quad \hat{\gamma}_1 = 1000 \quad (\text{USD/año})$$

### Interpretación de etapas

- **Primera etapa** ( $Z \rightarrow X$ ): acceso a transporte escolar gratuito aumenta en 0.50 años la educación secundaria completada.
- **Forma reducida** ( $Z \rightarrow Y$ ): acceso a transporte incrementa el ingreso anual en 1000 USD.

### Estimador de Wald

$$\hat{\beta}^{Wald} = \frac{\hat{\gamma}_1}{\hat{\alpha}_1} = \frac{1000}{0.50} = 2000 \text{ USD por año adicional de educación.}$$

## Ejemplo numérico (Wald): Transporte escolar como $Z$

Supongamos (cifras ilustrativas):

$$\hat{\alpha}_1 = 0.50 \quad (\text{años}) \quad \hat{\gamma}_1 = 1000 \quad (\text{USD/año})$$

### Interpretación de etapas

- **Primera etapa** ( $Z \rightarrow X$ ): acceso a transporte escolar gratuito aumenta en 0.50 años la educación secundaria completada.
- **Forma reducida** ( $Z \rightarrow Y$ ): acceso a transporte incrementa el ingreso anual en 1000 USD.

### Estimador de Wald

$$\hat{\beta}^{\text{Wald}} = \frac{\hat{\gamma}_1}{\hat{\alpha}_1} = \frac{1000}{0.50} = 2000 \text{ USD por año adicional de educación.}$$

### Lectura

Usamos sólo la variación de  $X$  inducida por  $Z$ . El retorno local a un año adicional de educación (para quienes cambian su conducta por el transporte) es de  $\sim \$2000$  anuales.

## Lectura LATE del ejemplo

**Supongamos además:** sólo un 30% de los estudiantes (*cumplidores*) cambia su decisión de asistir gracias al transporte. Descomposición intuitiva:

$$\underbrace{\hat{\gamma}_1}_{\text{E. Promedio inducido por } Z \text{ en } Y} = \underbrace{\hat{\alpha}_1}_{\text{V. Promedio inducido por } Z \text{ en } X} \times \underbrace{\hat{\beta}^{\text{Wald}}}_{\text{efecto local (LATE)}}$$
$$1000 = 0.50 \times 2000$$



## Lectura LATE del ejemplo

**Supongamos además:** sólo un 30% de los estudiantes (*cumplidores*) cambia su decisión de asistir gracias al transporte. Descomposición intuitiva:

$$\underbrace{\hat{\gamma}_1}_{\text{E. Promedio inducido por } Z \text{ en } Y} = \underbrace{\hat{\alpha}_1}_{\text{V. Promedio inducido por } Z \text{ en } X} \times \underbrace{\hat{\beta}^{\text{Wald}}}_{\text{efecto local (LATE)}}$$

$$1000 = 0.50 \times 2000$$

**Intuición con cumplidores:** Si sólo el 30% responde al instrumento, entonces el cambio medio en escolaridad es mayor.

$$\Delta X_C \approx \frac{\hat{\alpha}_1}{p_c} = \frac{0.50}{0.30} \approx 1.67 \text{ años.}$$

En general, con  $p_c$  (proporción de cumplidores):

## Lectura LATE del ejemplo

**Supongamos además:** sólo un 30% de los estudiantes (*cumplidores*) cambia su decisión de asistir gracias al transporte. Descomposición intuitiva:

$$\underbrace{\hat{\gamma}_1}_{\text{E. Promedio inducido por } Z \text{ en } Y} = \underbrace{\hat{\alpha}_1}_{\text{V. Promedio inducido por } Z \text{ en } X} \times \underbrace{\hat{\beta}^{Wald}}_{\text{efecto local (LATE)}}$$

$$1000 = 0.50 \times 2000$$

**Intuición con cumplidores:** Si sólo el 30% responde al instrumento, entonces el cambio medio en escolaridad es mayor.

$$\Delta X_C \approx \frac{\hat{\alpha}_1}{p_c} = \frac{0.50}{0.30} \approx 1.67 \text{ años.}$$

En general, con  $p_c$  (proporción de cumplidores):

$$\hat{\alpha}_1 = p_c \Delta X_C, \quad \hat{\gamma}_1 = p_c \Delta Y_C, \quad \hat{\beta}^{Wald} = \frac{\hat{\gamma}_1}{\hat{\alpha}_1} = \frac{\Delta Y_C}{\Delta X_C} = \text{LATE.}$$

## Lectura LATE del ejemplo

**Supongamos además:** sólo un 30% de los estudiantes (*cumplidores*) cambia su decisión de asistir gracias al transporte. Descomposición intuitiva:

$$\underbrace{\hat{\gamma}_1}_{\text{E. Promedio inducido por } Z \text{ en } Y} = \underbrace{\hat{\alpha}_1}_{\text{V. Promedio inducido por } Z \text{ en } X} \times \underbrace{\hat{\beta}^{Wald}}_{\text{efecto local (LATE)}}$$

$$1000 = 0.50 \times 2000$$

**Intuición con cumplidores:** Si sólo el 30% responde al instrumento, entonces el cambio medio en escolaridad es mayor.

$$\Delta X_C \approx \frac{\hat{\alpha}_1}{p_c} = \frac{0.50}{0.30} \approx 1.67 \text{ años.}$$

En general, con  $p_c$  (proporción de cumplidores):

$$\hat{\alpha}_1 = p_c \Delta X_C, \quad \hat{\gamma}_1 = p_c \Delta Y_C, \quad \hat{\beta}^{Wald} = \frac{\hat{\gamma}_1}{\hat{\alpha}_1} = \frac{\Delta Y_C}{\Delta X_C} = \text{LATE.}$$

### Moraleja

Cuando la relevancia es baja ( $\hat{\alpha}_1$  pequeño), el efecto promedio inducido por el instrumento ( $\hat{\gamma}_1$ ) se diluye y es más difícil de detectar (menor poder).

# Roadmap

## ¿Como encontrar un buen instrumento?



# Limitaciones de los Instrumentos

**Aun cuando  $Z$  parece válido, hay limitaciones importantes:**

# Limitaciones de los Instrumentos

Aun cuando  $Z$  parece válido, hay limitaciones importantes:

- **Validez externa limitada:** El efecto estimado (LATE) aplica solo a los *cumplidores*, no a toda la población.

# Limitaciones de los Instrumentos

Aun cuando  $Z$  parece válido, hay limitaciones importantes:

- **Validez externa limitada:** El efecto estimado (LATE) aplica solo a los *cumplidores*, no a toda la población.
- **Instrumentos débiles:** Si  $Z$  explica muy poca variación en  $X$ , las estimaciones IV/2SLS son imprecisas y sesgadas hacia OLS.



# Limitaciones de los Instrumentos

Aun cuando  $Z$  parece válido, hay limitaciones importantes:

- **Validez externa limitada:** El efecto estimado (LATE) aplica solo a los *cumplidores*, no a toda la población.
- **Instrumentos débiles:** Si  $Z$  explica muy poca variación en  $X$ , las estimaciones IV/2SLS son imprecisas y sesgadas hacia OLS.
- **Difícil de defender la exclusión:** Puede existir un canal no observado  $Z \rightarrow Y$  distinto a través de  $X$ . Ejemplo: clima afecta educación pero también productividad agrícola.

# Limitaciones de los Instrumentos

Aun cuando  $Z$  parece válido, hay limitaciones importantes:

- **Validez externa limitada:** El efecto estimado (LATE) aplica solo a los *cumplidores*, no a toda la población.
- **Instrumentos débiles:** Si  $Z$  explica muy poca variación en  $X$ , las estimaciones IV/2SLS son imprecisas y sesgadas hacia OLS.
- **Difícil de defender la exclusión:** Puede existir un canal no observado  $Z \rightarrow Y$  distinto a través de  $X$ . Ejemplo: clima afecta educación pero también productividad agrícola.
- **Sensibles a la especificación:** Resultados pueden depender de los controles  $W$ , definiciones de muestra o clusters de errores.

# Limitaciones de los Instrumentos

Aun cuando  $Z$  parece válido, hay limitaciones importantes:

- **Validez externa limitada:** El efecto estimado (LATE) aplica solo a los *cumplidores*, no a toda la población.
- **Instrumentos débiles:** Si  $Z$  explica muy poca variación en  $X$ , las estimaciones IV/2SLS son imprecisas y sesgadas hacia OLS.
- **Difícil de defender la exclusión:** Puede existir un canal no observado  $Z \rightarrow Y$  distinto a través de  $X$ . Ejemplo: clima afecta educación pero también productividad agrícola.
- **Sensibles a la especificación:** Resultados pueden depender de los controles  $W$ , definiciones de muestra o clusters de errores.
- **Interpretación restringida:** IV no recupera el ATE ni el TOT en general, sino el efecto local para los inducidos por  $Z$ .

# Limitaciones de los Instrumentos

Aun cuando  $Z$  parece válido, hay limitaciones importantes:

- **Validez externa limitada:** El efecto estimado (LATE) aplica solo a los *cumplidores*, no a toda la población.
- **Instrumentos débiles:** Si  $Z$  explica muy poca variación en  $X$ , las estimaciones IV/2SLS son imprecisas y sesgadas hacia OLS.
- **Difícil de defender la exclusión:** Puede existir un canal no observado  $Z \rightarrow Y$  distinto a través de  $X$ . Ejemplo: clima afecta educación pero también productividad agrícola.
- **Sensibles a la especificación:** Resultados pueden depender de los controles  $W$ , definiciones de muestra o clusters de errores.
- **Interpretación restringida:** IV no recupera el ATE ni el TOT en general, sino el efecto local para los inducidos por  $Z$ .

## Moraleja

Un buen instrumento es raro: debe ser **relevante, creíble en la exclusión y suficientemente fuerte**. La interpretación debe hacerse con cautela.

# Roadmap

## Acerca de este paper

- Pregunta central: ¿La asistencia escolar obligatoria afecta la escolaridad y los ingresos?
- Contexto: En EE.UU., la edad de inicio escolar y las leyes de asistencia obligatoria generan variación en la escolaridad según el trimestre de nacimiento.
- Enfoque: Usar el trimestre de nacimiento como **variable instrumental** para educación.
- Conclusión: La escolaridad inducida por leyes obligatorias aumenta los ingresos.
- Contribución: Evidencia empírica novedosa sobre retornos a la educación con un “experimento natural”.

# Estrategía de Identificación

- **Idea clave:**

- ▶ La edad mínima para ingresar a la escuela es 6 años cumplidos antes de enero.
- ▶ Los nacidos a inicio de año (Q1) entran a la escuela más viejos que los nacidos a final de año (Q4).
- ▶ Con leyes de asistencia obligatoria (hasta 16–17 años), los nacidos en Q1 pueden abandonar antes y acumulan menos años de estudio.

# Estrategía de Identificación

- **Idea clave:**

- ▶ La edad mínima para ingresar a la escuela es 6 años cumplidos antes de enero.
- ▶ Los nacidos a inicio de año (Q1) entran a la escuela más viejos que los nacidos a final de año (Q4).
- ▶ Con leyes de asistencia obligatoria (hasta 16–17 años), los nacidos en Q1 pueden abandonar antes y acumulan menos años de estudio.

- **Fuente de variación:**

- ▶ La interacción entre *reglas de entrada escolar* y *edad mínima de abandono*.
- ▶ Genera diferencias “forzadas” en años de escolaridad según trimestre de nacimiento.



# Estrategía de Identificación

- **Idea clave:**

- ▶ La edad mínima para ingresar a la escuela es 6 años cumplidos antes de enero.
- ▶ Los nacidos a inicio de año (Q1) entran a la escuela más viejos que los nacidos a final de año (Q4).
- ▶ Con leyes de asistencia obligatoria (hasta 16–17 años), los nacidos en Q1 pueden abandonar antes y acumulan menos años de estudio.

- **Fuente de variación:**

- ▶ La interacción entre *reglas de entrada escolar* y *edad mínima de abandono*.
- ▶ Genera diferencias “forzadas” en años de escolaridad según trimestre de nacimiento.

- **Supuesto central:**

- ▶ El trimestre de nacimiento es **asignado aleatoriamente** respecto a habilidad, motivación o condiciones familiares.
- ▶ No hay un efecto directo de nacer en cierto trimestre sobre salarios, excepto vía educación.

## Estimación IV

- **Trimestre de nacimiento**  $\Rightarrow$  fuente de variación exógena en educación.
- Se convierte en un **instrumento válido** bajo el supuesto de exclusión:
  - ▶ Afecta la educación acumulada.
  - ▶ No afecta directamente los salarios, excepto a través de la educación.

## Estimación IV

- **Trimestre de nacimiento**  $\Rightarrow$  fuente de variación exógena en educación.
- Se convierte en un **instrumento válido** bajo el supuesto de exclusión:
  - ▶ Afecta la educación acumulada.
  - ▶ No afecta directamente los salarios, excepto a través de la educación.
- Estimación con *Variables Instrumentales* (2SLS):

$$\underbrace{\text{Educación}_i}_{\text{1ra etapa}} \leftarrow f(\text{Trimestre de Nacimiento, cohorte, } X_i)$$

$$\underbrace{\ln(\text{Salario}_i)}_{\text{2da etapa}} \leftarrow \beta \widehat{\text{Educación}_i} + X_i + \varepsilon_i$$

## Estimación IV

- **Trimestre de nacimiento**  $\Rightarrow$  fuente de variación exógena en educación.
- Se convierte en un **instrumento válido** bajo el supuesto de exclusión:
  - ▶ Afecta la educación acumulada.
  - ▶ No afecta directamente los salarios, excepto a través de la educación.
- Estimación con *Variables Instrumentales* (2SLS):

$$\underbrace{\text{Educación}_i}_{\text{1ra etapa}} \leftarrow f(\text{Trimestre de Nacimiento, cohorte, } X_i)$$

$$\underbrace{\ln(\text{Salario}_i)}_{\text{2da etapa}} \leftarrow \beta \widehat{\text{Educación}_i} + X_i + \varepsilon_i$$

- **Interpretación:**  $\beta$  mide el retorno a la educación inducida por leyes de asistencia obligatoria.

- **Fuentes:** Censos de Población de EE.UU. (1960, 1970, 1980).
- **Muestra principal:**
  - ▶ Hombres nacidos entre 1920–1949 en EE.UU.
  - ▶ Se usan diferentes cohortes para observar efectos en distintos períodos.
- **Variables clave:**
  - ▶ Años de educación completados (años escolares).
  - ▶ Tasas de matrícula y abandono escolar (en la adolescencia).
  - ▶ Ingresos semanales (log del salario).
  - ▶ Controles: edad precisa (años + trimestre), estado civil, residencia en zona metropolitana, raza.
- **Estrategia de análisis:**
  - ▶ Comparación de cohortes por **trimestre de nacimiento**.
  - ▶ Construcción de instrumentos: interacciones  $QOB \times cohorte$  y  $QOB \times estado$ .

# Resultados

## ● Evidencia descriptiva:

- ▶ Los nacidos en el primer trimestre (Q1) acumulan, en promedio, menos años de escolaridad.
- ▶ Menor probabilidad de graduarse de secundaria en comparación con los nacidos en el último trimestre (Q4).
- ▶ Patrón consistente en varias cohortes (1930s, 1940s), aunque se atenúa con el tiempo.

# Resultados

## ● Evidencia descriptiva:

- ▶ Los nacidos en el primer trimestre (Q1) acumulan, en promedio, menos años de escolaridad.
- ▶ Menor probabilidad de graduarse de secundaria en comparación con los nacidos en el último trimestre (Q4).
- ▶ Patrón consistente en varias cohortes (1930s, 1940s), aunque se atenúa con el tiempo.

## ● Mecanismo:

- ▶ La brecha se origina en la interacción entre **edad de entrada escolar** y **edad mínima legal de abandono**.
- ▶ Evidencia adicional: diferencias de matrícula al comparar estados con leyes de abandono a los 16 vs 17 años.

# Resultados

## ● Evidencia descriptiva:

- ▶ Los nacidos en el primer trimestre (Q1) acumulan, en promedio, menos años de escolaridad.
- ▶ Menor probabilidad de graduarse de secundaria en comparación con los nacidos en el último trimestre (Q4).
- ▶ Patrón consistente en varias cohortes (1930s, 1940s), aunque se atenúa con el tiempo.

## ● Mecanismo:

- ▶ La brecha se origina en la interacción entre **edad de entrada escolar** y **edad mínima legal de abandono**.
- ▶ Evidencia adicional: diferencias de matrícula al comparar estados con leyes de abandono a los 16 vs 17 años.

## ● Estimaciones econométricas:

- ▶ Retorno a un año adicional de educación (2SLS): **7–10%**.
- ▶ Estimaciones cercanas a OLS  $\Rightarrow$  sesgo por habilidad omitida es pequeño o incluso negativo.
- ▶ Robustez: no se encuentran efectos de trimestre de nacimiento en educación superior ni en graduados universitarios (apoya la validez del instrumento).



# Roadmap

# Replication Package

**Repositorio en R (GitHub):** Descarga directa del paquete de replicación (código y datos en R): [replication\\_package.zip](#)

**Instrucciones:** Antes de salir al break:

- Descarguen y descompriman el paquete.
- Ejecuten el script inicial para cargar librerías:

**Script en R:**

- `require(pacman)`
- `p_load(tidyverse, rio, fixest, broom)`

Cuando regresemos de la pausa, las librerías ya estarán cargadas y listas para usar.