

Variables Instrumentales

Econometría Aplicada y Ciencia de Datos - CIDE Fall 2026

Carlos Brito

- 1 Repaso MCO
- 2 Motivación - Variables Instrumentales
- 3 Definición de Instrumento
- 4 Estimador de Variables Instrumentales
- 5 Lenguaje de Variables Instrumentales
- 6 2SLS
- 7 Propiedades 2SLS
- 8 Aplicación: Card (1995)
- 9 Aplicación: Oferta y Demanda

Repaso MCO

Repaso: Mínimos Cuadrados Ordinarios (MCO)

Modelo lineal poblacional: $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + u$

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

Inssegadez:

- Linealidad en los parámetros
- Muestra aleatoria
- No colinealidad perfecta
- Media condicional cero:
 $E(u | x_1, \dots, x_K) = 0$

$$\hookrightarrow E(\hat{\beta}_{OLS}) = \beta$$

Consistencia:

- Muestra aleatoria
- No colinealidad perfecta
- Ortogonalidad:
 $E(u) = 0, \text{Cov}(x_j, u) = 0 \quad \forall j$

$$\hookrightarrow \hat{\beta}_{OLS} \xrightarrow{p} \beta$$

- Inssegadez + homoced. ($\text{Var}(u | x_1, \dots, x_K) = \sigma^2$) $\Rightarrow$ Gauss–Markov:

Repaso: Insesgadez de MCO

$$y = X\beta + u, \quad \hat{\beta} = (X'X)^{-1}X'y$$

Supuestos: linealidad, muestra aleatoria, no colinealidad perfecta ($X'X$ invertible), y $E(u | X) = 0$

Repaso: Consistencia de MCO

$$\hat{\beta} = \beta + (X'X)^{-1}X'u = \beta + \left(\frac{1}{N}X'X\right)^{-1}\left(\frac{1}{N}X'u\right)$$

Supuestos: muestra aleatoria i.i.d., $E(u) = 0$, $\text{Cov}(x_j, u) = 0 \forall j$, y $E(x'x)$ finita y no singular.

$$y = X\beta + u, \quad u \mid X \sim (0, \sigma^2 I_n)$$

$$\hat{\beta} = (X'X)^{-1}X'y$$

With normality:

$$\hat{\beta} \mid X \sim N(\beta, \sigma^2(X'X)^{-1})$$

Model and Estimator

$$y = X\beta + u, \quad u | X \sim (0, \sigma^2 I_n)$$

$$\hat{\beta} = (X'X)^{-1}X'y$$

With normality:

$$\hat{\beta} | X \sim N(\beta, \sigma^2(X'X)^{-1})$$

Without normality:
asymptotically (CLT):

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1}), \quad Q = \text{plim } \frac{1}{n} X'X$$

Feasible variance estimate:

$$\widehat{\text{Var}}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}, \quad \hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n - k}$$

t -test: Single Coefficient

Test: $H_0 : \beta_j = \beta^*$ vs. $H_1 : \beta_j \neq \beta^*$

$$t_j = \frac{\hat{\beta}_j - \beta^*}{\text{se}(\hat{\beta}_j)} \sim t_{n-k} \quad \text{under } H_0$$

- **Reject H_0** at level α if $|t_j| > t_{n-k, \alpha/2}$.
- $(1 - \alpha)$ confidence interval: $\hat{\beta}_j \pm t_{n-k, \alpha/2} \cdot \text{se}(\hat{\beta}_j)$
- As $n - k \rightarrow \infty$: $t_{n-k} \rightarrow N(0, 1)$.

F-test: Joint Restrictions

1. Test q linear restrictions $H_0 : R\beta = r$:

$$F = \frac{(\text{RSS}_R - \text{RSS}_{UR})/q}{\text{RSS}_{UR}/(n-k)} \sim F_{q, n-k} \quad \text{under } H_0$$

- Reject H_0 if $F > F_{q, n-k, \alpha}$.

2. **Overall significance** ($H_0 : \beta_1 = \cdots = \beta_{k-1} = 0$):

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F_{k-1, n-k}$$

- Reject H_0 if $F > F_{k-1, n-k, \alpha}$.

Motivación - Variables Instrumentales

Motivación: ¿Por qué necesitamos Variables Instrumentales?

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + u$$

- Cuando $\text{Cov}(x_j, u) = 0$ falla: MCO es **sesgado e inconsistente**.
- Fuentes de endogeneidad:
 - Variables omitidas (p. ej., habilidad no observada)
 - Error de medición en los regresores
 - Simultaneidad / causalidad inversa

Sin endogeneidad

x y

u

x y u son causas independientes de y : MCO consistente.

Regresor endógeno

x y

u

u afecta tanto a x como a y : MCO captura un efecto *directo + indirecto* $\Rightarrow$ sesgo.

Motivación: ¿Por qué necesitamos Variables Instrumentales?

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + u$$

- Cuando $\text{Cov}(x_j, u) = 0$ falla: MCO es **sesgado e inconsistente**.
- Fuentes de endogeneidad:
 - Variables omitidas (p. ej., habilidad no observada)
 - Error de medición en los regresores
 - Simultaneidad / causalidad inversa

Sin endogeneidad

x y

u

x y u son causas independientes de y : MCO consistente.

Regresor endógeno

x y

u

u afecta tanto a x como a y : MCO captura un efecto *directo + indirecto* $\Rightarrow$ sesgo.

- **Variables Instrumentales:** explotar una fuente externa de variación *exógena* en el regresor endógeno.

Definición de Instrumento

Consideremos el modelo

$$y = \beta_0 + \beta_1 x + u, \quad \text{cov}(x, u) \neq 0$$

Una variable z_1 es un **instrumento válido**:

1. Exogeneidad

z no está correlacionada con el error:

$$\text{cov}(z, u) = 0$$

2. Relevancia

z está correlacionada con el regresor endógeno:

$$\text{cov}(z, x) \neq 0$$

Estimador de Variables Instrumentales

Tomando la covarianza con z en ambos lados de $y = \beta_0 + \beta_1 x + u$:

$$\text{cov}(y, z) = \beta_1 \text{cov}(x, z) + \text{cov}(u, z)$$

Como $\text{cov}(u, z) = 0$ por exogeneidad, despejamos:

$$\hat{\beta}_1 = \frac{\hat{\text{cov}}(y, z)}{\hat{\text{cov}}(x, z)}, \quad \text{siempre que } \text{cov}(x, z) \neq 0$$

- Por una Ley de Grandes Números:

$\hat{\beta}_1$ es **consistente**

Tomando la covarianza con z en ambos lados de $y = \beta_0 + \beta_1 x + u$:

$$\text{cov}(y, z) = \beta_1 \text{cov}(x, z) + \text{cov}(u, z)$$

Como $\text{cov}(u, z) = 0$ por exogeneidad, despejamos:

$$\hat{\beta}_1 = \frac{\hat{\text{cov}}(y, z)}{\hat{\text{cov}}(x, z)}, \quad \text{siempre que } \text{cov}(x, z) \neq 0$$

- Por una Ley de Grandes Números:

$\hat{\beta}_1$ es **consistente**

- Sin embargo, $\hat{\beta}_1$ es **siempre sesgado** en muestras finitas.
- El sesgo puede ser sustancial si la muestra es pequeña $\Rightarrow$ tener precaución con el tamaño muestral.

Estimador de VI

Modelo: $y = \beta x + u$, instrumento z .

Derivación intuitiva (Heckman)

$$\beta_{VI} = \frac{dy/dz}{dx/dz}$$

Se estiman dy/dz y dx/dz mediante pendientes simples de MCO:

$$\hat{\beta}_{VI} = \frac{(z'z)^{-1}z'y}{(z'z)^{-1}z'x} = (z'x)^{-1}z'y$$

Forma de covarianza

$$\hat{\beta}_{VI} = \frac{\text{Cov}[z, y]}{\text{Cov}[z, x]} = \frac{r_{zy}}{r_{zx}} \sqrt{\frac{y'y}{x'x}}$$

Ejemplo: Rendimientos a la Educación

- Card (1995) estudia los **retornos a la educación** - muestra 3,000 hombres 24-34 años.
- Interesa la relación entre educación e ingreso, pero la habilidad no se observa:
estimar $\ln(w_i) = \beta_0 + \beta_1 educ_i + u_i$ produce un $\hat{\beta}_1$ sesgado.

Ejemplo: Rendimientos a la Educación

- Card (1995) estudia los **retornos a la educación** - muestra 3,000 hombres 24-34 años.
- Interesa la relación entre educación e ingreso, pero la habilidad no se observa:
estimar $\ln(w_i) = \beta_0 + \beta_1 educ_i + u_i$ produce un $\hat{\beta}_1$ sesgado.
- **Instrumento propuesto:** una variable binaria que indica si en el municipio de la persona i había una universidad cercana.
- **Intuición:** la cercanía de la universidad reduce el costo de asistir, pero no afecta directamente el ingreso (solo a través de la escolaridad).

Encontrar un buen instrumento casi nunca es sencillo: exige entender los mecanismos, las instituciones y el contexto detrás de la variable endógena.

Lenguaje de Variables Instrumentales

Ecuación estructural:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + u, \quad \text{Cov}(x_K, u) \neq 0$$

- x_K : variable endógena
- z_1 : instrumento válido

Ecuación estructural:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + u, \quad \text{Cov}(x_K, u) \neq 0$$

- x_K : variable endógena
- z_1 : instrumento válido

Forma reducida:

$$x_K = \delta_0 + \delta_1 x_1 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + r_K \text{ (forma reducida de } x_K \text{)}$$

Sustituyendo en la ecuación estructural se obtiene la forma reducida de y :

$$y = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_{K-1} x_{K-1} + \lambda_1 z_1 + v, \quad v = u + \beta_K r_K \text{ (forma reducida de } y \text{)}$$

$$\text{con } \alpha_j = \beta_j + \beta_K \delta_j \text{ y } \lambda_1 = \beta_K \theta_1$$

Ecuación estructural:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + u, \quad \text{Cov}(x_K, u) \neq 0$$

- x_K : variable endógena
- z_1 : instrumento válido

Forma reducida:

$$x_K = \delta_0 + \delta_1 x_1 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + r_K \text{ (forma reducida de } x_K \text{)}$$

Sustituyendo en la ecuación estructural se obtiene la forma reducida de y :

$$y = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_{K-1} x_{K-1} + \lambda_1 z_1 + v, \quad v = u + \beta_K r_K \text{ (forma reducida de } y \text{)}$$

$$\text{con } \alpha_j = \beta_j + \beta_K \delta_j \text{ y } \lambda_1 = \beta_K \theta_1$$

- MCO estima de forma consistente los parámetros de la forma reducida α_j, λ_1 .
- Las estimaciones de forma reducida a veces interesan por sí mismas, pero el parámetro estructural β_K suele ser el de interés.

La **primera etapa** especifica la relación entre la variable endógena y el instrumento:

$$x_K = \delta_0 + \delta_1 x_1 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + r_K$$

- La condición de relevancia se reescribe como $\theta_1 \neq 0$.
- La primera etapa **puede y debe** probarse empíricamente.

La **primera etapa** especifica la relación entre la variable endógena y el instrumento:

$$x_K = \delta_0 + \delta_1 x_1 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + r_K$$

- La condición de relevancia se reescribe como $\theta_1 \neq 0$.
- La primera etapa **puede y debe** probarse empíricamente.

Restricción de exclusión

No es posible probar directamente que $\text{cov}(z_i, u_i) = 0$. Esta restricción debe sustentarse en la teoría económica, el conocimiento institucional, o la exogeneidad de un experimento natural.

Más de un Instrumento

- Si existen J variables z_{ij} candidatas a instrumento: cada una debe cumplir $\text{cov}(u_i, z_{ij}) = 0$ y correlacionarse con x_{1i} .

Ejemplo: dos instrumentos, la primera etapa es

$$x_K = \delta_0 + \delta_1 x_1 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + \theta_2 z_2 + r_K$$

con $\text{cov}(z_1, r_K) = \text{cov}(z_2, r_K) = \text{cov}(x_j, r_K) = 0, \forall j \neq k$.

- Para lograr identificación, basta con $\theta_1 \neq 0$ o $\theta_2 \neq 0$.

- Si existen J variables z_{ij} candidatas a instrumento: cada una debe cumplir $\text{cov}(u_i, z_{ij}) = 0$ y correlacionarse con x_{1i} .

Ejemplo: dos instrumentos, la primera etapa es

$$x_K = \delta_0 + \delta_1 x_1 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + \theta_2 z_2 + r_K$$

con $\text{cov}(z_1, r_K) = \text{cov}(z_2, r_K) = \text{cov}(x_j, r_K) = 0, \forall j \neq k$.

- Para lograr identificación, basta con $\theta_1 \neq 0$ o $\theta_2 \neq 0$.

Exactamente identificado: número de instrumentos = número de variables endógenas

Sobreidentificado: número de instrumentos $>$ número de variables endógenas

2SLS

Mínimos Cuadrados en Dos Etapas

El procedimiento para estimar β_1 en dos pasos:

- 1 Regresión de x_{1i} sobre los instrumentos y las variables exógenas $\Rightarrow$ valores ajustados $\hat{x}_{1i}$.

x_K sobre $1, x_1, \dots, x_{K-1}, z_1, \dots, z_M$

Mínimos Cuadrados en Dos Etapas

El procedimiento para estimar β_1 en dos pasos:

- 1 Regresión de x_{1i} sobre los instrumentos y las variables exógenas $\Rightarrow$ valores ajustados $\hat{x}_{1i}$.

$$x_K \text{ sobre } 1, x_1, \dots, x_{K-1}, z_1, \dots, z_M$$

- 2 Regresión de y_i sobre las variables exógenas y $\hat{x}_{1i}$.

$$y \text{ sobre } 1, x_1, \dots, x_{K-1}, \hat{x}_K$$

Es como *purgar* a x_{1i} de su correlación con u_i .

Mínimos Cuadrados en Dos Etapas

El procedimiento para estimar β_1 en dos pasos:

- 1 Regresión de x_{1i} sobre los instrumentos y las variables exógenas $\Rightarrow$ valores ajustados $\hat{x}_{1i}$.

$$x_K \text{ sobre } 1, x_1, \dots, x_{K-1}, z_1, \dots, z_M$$

- 2 Regresión de y_i sobre las variables exógenas y $\hat{x}_{1i}$.

$$y \text{ sobre } 1, x_1, \dots, x_{K-1}, \hat{x}_K$$

Es como *purgar* a x_{1i} de su correlación con u_i .

Advertencia

- Nunca omita los regresores exógenos $x_1, \dots, x_{K-1}$ en la regresión de primera etapa (produce estimadores inconsistentes)
- Nunca se hace este procedimiento “a mano” — los errores estándar de la segunda etapa serían incorrectos.

Propiedades 2SLS

Supuestos Generales para 2SLS (Wooldridge)

Supuesto 2SLS.1 (Ortogonalidad $\approx$ Exogeneidad)

Para el vector $1 \times L$ z : $E(z'u) = 0$.

Supuesto 2SLS.2 (Condiciones de rango $\approx$ Relevancia)

(a) $\text{rank } E(z'z) = L$ (b) $\text{rank } E(z'x) = K$

Supuesto 2SLS.3 (Homocedasticidad)

$$E(u^2 z'z) = \sigma^2 E(z'z), \quad \sigma^2 = E(u^2)$$

Condición de orden: $L \geq K$

- 1 Bajo 2SLS.1–2SLS.2: el estimador 2SLS obtenido de una muestra aleatoria es **consistente** para β :

$$\hat{\beta} \xrightarrow{p} \beta$$

- ❶ Bajo 2SLS.1–2SLS.2: el estimador 2SLS obtenido de una muestra aleatoria es **consistente** para β :

$$\hat{\beta} \xrightarrow{p} \beta$$

- ❷ Bajo 2SLS.1–2SLS.3:

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \text{Normal}\left(0, \sigma^2 ([E(x'z)] [E(z'z)]^{-1} [E(z'x)])^{-1}\right)$$

- Matriz de varianza estimada: $\hat{\sigma}^2 (\hat{X}'\hat{X})^{-1}$, con $\hat{\sigma}^2 = (N - K)^{-1} \sum_i \hat{u}_i^2$ usando los residuos *estructurales* de 2SLS $\hat{u}_i = y_i - x_i \hat{\beta}$ (**no** los residuos de MCO de la segunda etapa).

- ❶ Bajo 2SLS.1–2SLS.2: el estimador 2SLS obtenido de una muestra aleatoria es **consistente** para β :

$$\hat{\beta} \xrightarrow{p} \beta$$

- ❷ Bajo 2SLS.1–2SLS.3:

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \text{Normal}\left(0, \sigma^2 ([E(x'z)] [E(z'z)]^{-1} [E(z'x)])^{-1}\right)$$

- Matriz de varianza estimada: $\hat{\sigma}^2(\hat{X}'\hat{X})^{-1}$, con $\hat{\sigma}^2 = (N - K)^{-1} \sum_i \hat{u}_i^2$ usando los residuos *estructurales* de 2SLS $\hat{u}_i = y_i - x_i\hat{\beta}$ (**no** los residuos de MCO de la segunda etapa).
- ❸ Bajo 2SLS.1–2SLS.3: 2SLS es **eficiente** dentro de la clase de todos los estimadores de VI que usan instrumentos lineales en z .
- Si x es exógena, 2SLS = MCO
 - Si falla el Supuesto 2SLS.3, existe un estimador GMM más eficiente (Hansen 1982; White 1982)

Aplicación: Card (1995)

Ejemplo Card (1995): Rendimientos a la Educación

En el problema de Card (1995), la primera etapa es

$$esc_i = \pi_0 + \pi_1 X_i + \phi unicerca_i + \nu_i, \quad unicerca_i = \begin{cases} 1 & \text{había universidad en el municipio} \\ 0 & \text{en otro caso} \end{cases}$$

Y la forma reducida es

$$\ln(w_i) = \gamma_0 + \gamma_1 X_i + \delta unicerca_i + \varepsilon_i$$

El salario está correlacionado con la presencia de la universidad, pero esa relación ocurre **a través de la escolaridad**.

Ejemplo Card (1995): Estimación en R

Datos de Card, variable instrumental `nearc4` (universidad cercana):

R

```
library(readr)
card <- read_csv(
  "your folder address/card1995.csv"
)

# OLS
ols_formula <- lwage ~ educ + exper + black + south + married + smsa
ols_fit <- lm(ols_formula, data = card)

print(summary(ols_fit))

# Variables instrumentales
iv_formula <- lwage ~ educ + exper + black + south + married + smsa |
  nearc4 + exper + black + south + married + smsa
iv_fit <- ivreg(iv_formula, data = card)

print(summary(iv_fit))
```

Ejemplo Card (1995): MCO vs. VI

El $\hat{\beta}_1$ de MCO es inconsistente; comparamos con la estimación de VI:

	MCO	VI
educ	0.071 (0.003)	0.124 (0.050)
exper	0.034 (0.002)	0.056 (0.020)
N	3,003	3,003
F	219.15	—

El coeficiente de VI (0.124) es notablemente mayor que el de MCO (0.071), con un error estándar bastante más grande — patrón típico cuando el instrumento no es muy fuerte.

Ejemplo Card (1995): Primera Etapa

	MCO	VI	Primera etapa
educ	0.071 (0.003)	0.124 (0.050)	
exper	0.034 (0.002)	0.056 (0.020)	-0.404 (0.009)
nearc4			0.327 (0.082)
N	3,003	3,003	3,003
F	219.15	—	456.14

El coeficiente de `nearc4` en la primera etapa (0.327) es positivo y significativo: la cercanía a una universidad aumenta la escolaridad esperada.

Ejemplo Card (1995): Fuerza del Instrumento

Debe verificarse el estadístico F de la primera etapa asociado *solo* a los instrumentos excluidos:

R

```
pe_vi <- lm(educ ~ nearc4 + exper + black + south + married + smsa,  
            data = data.ingresos)  
  
linearHypothesis(pe_vi, c("nearc4 = 0"))
```

Modelo	Res. Df	RSS	F	Pr(>F)
Restringido	2,997	11,304		
Completo	2,996	11,245	15.77	7.3×10^{-5}

$F \approx 15.8$ para un único instrumento: por encima de la regla práctica de $F > 10$, aunque no de forma holgada.

Aplicación: Oferta y Demanda

Ejemplo Oferta y Demanda

Supongamos que interesa estimar la elasticidad de la demanda de mantequilla:

$$\ln(q_i) = \alpha + \beta \ln(p_i) + \varepsilon_i$$

con datos de precio y consumo para una muestra grande de localidades i .

El problema de simultaneidad

Si estimamos esta ecuación por MCO, $\hat{\beta}$ es **inconsistente**: precio y cantidad se determinan simultáneamente en el equilibrio de oferta y demanda. Un choque a cualquiera de las dos curvas mueve ambas variables a la vez.

Ejemplo Oferta y Demanda: la Solución de VI

Para estimar la elasticidad de la **demanda** necesitamos una variable z que:

- afecte solo a la **oferta** (p. ej., precipitación o temperatura, que inciden en la producción);
- no afecte directamente a la demanda.

Intuición gráfica: al desplazar la curva de oferta mientras la demanda permanece fija, los distintos puntos de equilibrio observados *trazan* la curva de demanda, permitiendo identificar y estimar su elasticidad.

Ejemplo Oferta y Demanda: Identificación Gráfica en R

Para ilustrar la intuición, simulamos una demanda fija y tres curvas de oferta desplazadas por un instrumento (p. ej., condiciones climáticas):

R

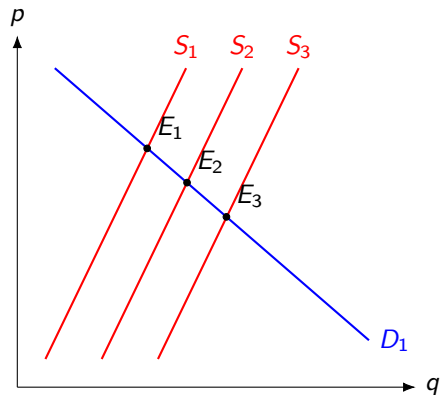
```
p <- seq(1, 9, by = 0.1)

demanda <- 10 - p # curva de demanda fija
oferta_1 <- p - 1 # oferta con clima promedio
oferta_2 <- p - 3 # oferta con clima favorable (mas produccion)
oferta_3 <- p - 5 # oferta con clima muy favorable

plot(demanda, p, type = "l", col = "blue", lwd = 2,
      xlab = "Cantidad", ylab = "Precio", xlim = c(0, 9))
lines(oferta_1, p, col = "red")
lines(oferta_2, p, col = "red", lty = 2)
lines(oferta_3, p, col = "red", lty = 3)
```

Cada desplazamiento de la oferta genera un nuevo punto de equilibrio observado (precio, cantidad); estos puntos son los que usamos para estimar la curva de demanda.

Ejemplo Oferta y Demanda: Identificación Gráfica



Cada choque exógeno a la oferta ($S_1 \rightarrow S_2 \rightarrow S_3$) genera un nuevo equilibrio observado (E_1, E_2, E_3) **sobre la curva de demanda fija** D_1 — así es como VI identifica su pendiente (elasticidad).