

Econometría 1 - Actividad 7

Simulaciones de Variables Instrumentales

Ejercicio 4: Instrumentos Débiles y Endogeneidad

División de Economía - CDE
Dr. Francisco Cabrera

1 Introducción: Por Qué Simulaciones

Las **simulaciones Monte Carlo** son una herramienta fundamental para entender las propiedades de los estimadores econométricos.

1.1 Ventajas de las Simulaciones

1. Conocemos la verdad:

- En datos reales: los parámetros verdaderos son desconocidos
- En simulaciones: nosotros especificamos los parámetros
- Podemos evaluar si los estimadores recuperan la verdad

2. Experimentación controlada:

- Cambiar un solo aspecto (e.g., fuerza del instrumento)
- Ver el efecto aislado
- Imposible en datos reales

3. Verificar propiedades teóricas:

- Consistencia: $\text{plim}_{n \rightarrow \infty} \hat{\beta} = \beta$
- Insesgamiento: $E[\hat{\beta}] = \beta$
- Distribución asintótica

2 Escenario Base: Instrumento Fuerte y Exógeno

2.1 Proceso Generador de Datos (DGP)

DGP del Escenario Base:

$$\begin{aligned}z_i &\sim N(0, 1) \quad (\text{instrumento}) \\u_i &\sim N(0, 1) \quad (\text{error}) \\x_i &= 0.5 z_i + 4 u_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, 1) \\y_i &= 3 x_i + 5 u_i\end{aligned}\tag{1}$$

Parámetro verdadero: $\beta_1 = 3$

2.2 Propiedades del DGP

1. Endogeneidad de x

$$\text{Cov}(x_i, u_i) = \text{Cov}(0.5z_i + 4u_i + \varepsilon_i, u_i) = 4 \text{Var}(u_i) = 4 \neq 0\tag{2}$$

Por lo tanto, x es **endógena**.

2. Relevancia del instrumento

$$\text{Cov}(z_i, x_i) = \text{Cov}(z_i, 0.5z_i + 4u_i + \varepsilon_i) = 0.5 \text{Var}(z_i) = 0.5 \neq 0\tag{3}$$

El instrumento es **relevante**.

3. Exogeneidad del instrumento

Como z_i y u_i son generados independientemente:

$$\text{Cov}(z_i, u_i) = 0\tag{4}$$

El instrumento es **exógeno**.

2.3 Sesgo del Estimador MCO

El estimador de MCO tiene el siguiente límite en probabilidad:

Sesgo de MCO por endogeneidad:

$$\text{plim } \hat{\beta}_1^{\text{OLS}} = \beta_1 + \frac{\text{Cov}(x_i, u_i)}{\text{Var}(x_i)}\tag{5}$$

En nuestro DGP:

- $\text{Cov}(x_i, u_i) = 4$ (positivo)
- Por lo tanto: $\text{plim } \hat{\beta}_1^{\text{OLS}} > \beta_1 = 3$

Conclusión: MCO **sobreestima** el efecto verdadero.

2.4 Consistencia del Estimador IV

Con un instrumento válido (z relevante y exógeno):

$$\text{plim}_{n \rightarrow \infty} \hat{\beta}_1^{\text{IV}} = \beta_1 = 3 \quad (6)$$

El estimador IV es **consistente**.

3 Ejercicio 4a: Evaluación del Instrumento

3.1 Condición 1: Relevancia

Regresión de primera etapa

$$x_i = \pi_0 + \pi_1 z_i + v_i \quad (7)$$

Hipótesis a probar:

$$\begin{aligned} H_0 : \pi_1 &= 0 \quad (\text{instrumento irrelevante}) \\ H_1 : \pi_1 &\neq 0 \quad (\text{instrumento relevante}) \end{aligned} \quad (8)$$

Regla práctica de Stock y Yogo (2005)

Instrumento débil vs fuerte:

Un instrumento se considera **DÉBIL** si:

$$F_{\text{first stage}} < 10 \quad (9)$$

donde F es el estadístico F de la prueba de $H_0 : \pi_1 = 0$.

Consecuencias de instrumento débil:

- Sesgo del estimador IV hacia MCO
- Inferencia no válida (intervalos de confianza incorrectos)
- Distribución asintótica no aproxima bien la distribución finita

3.2 Condición 2: Exogeneidad

Restricción de exclusión:

$$\text{Cov}(z_i, u_i) = 0 \quad (10)$$

Esta condición **NO** es **testeable** directamente porque u_i no es observable.

Verificación por construcción del DGP

En nuestras simulaciones:

- $z_i \sim N(0, 1)$ generado independientemente
- $u_i \sim N(0, 1)$ generado independientemente
- Por construcción: $z_i \perp\!\!\!\perp u_i$

4 Ejercicio 4b: Instrumento Débil

4.1 Nuevo DGP con Instrumento Débil

DGP con relevancia reducida:

$$\begin{aligned} z_i &\sim N(0, 1) \\ u_i &\sim N(0, 1) \\ x_i &= 0.2 z_i + 4 u_i + \varepsilon_i \quad (\text{coeficiente de } z \text{ reducido}) \\ y_i &= 3 x_i + 5 u_i \end{aligned} \tag{11}$$

Cambio: $0.5 \rightarrow 0.2$ en la ecuación de x .

4.2 Sesgo con Instrumento Débil

Cuando el instrumento es débil, el estimador IV tiene **sesgo de muestra finita**.

Fórmula aproximada del sesgo (Bound, Jaeger, Baker, 1995):

$$\frac{E[\hat{\beta}_1^{\text{IV}} - \beta_1]}{\sigma} \approx \frac{1}{F} \cdot \frac{n - K}{n} \tag{12}$$

donde:

- F = F-estadístico de primera etapa
- n = tamaño de muestra
- K = número de instrumentos
- σ = desviación estándar de u

Implicación: El sesgo es inversamente proporcional a F .

- Si F es pequeño \rightarrow sesgo grande
- El estimador IV está sesgado hacia MCO

4.3 Dirección del Sesgo

Con instrumento débil:

$$E[\hat{\beta}_1^{IV}] \approx \lambda \beta_1^{OLS} + (1 - \lambda) \beta_1 \quad (13)$$

donde $0 < \lambda < 1$ depende inversamente de F .

Interpretación: $\hat{\beta}_1^{IV}$ es una mezcla entre el estimador consistente (β_1) y el estimador sesgado (MCO).

5 Ejercicio 4c: Error Estándar con Instrumento Débil

5.1 Fórmula del Error Estándar de IV

El error estándar asintótico del estimador IV es:

Error estándar de $\hat{\beta}_1^{IV}$:

$$SE(\hat{\beta}_1^{IV}) = \frac{\sigma_u}{\sigma_x \sqrt{n} \rho_{zx}} \quad (14)$$

donde:

- σ_u = desviación estándar del error u
- σ_x = desviación estándar de x
- ρ_{zx} = correlación entre z y x
- n = tamaño de muestra

5.2 Efecto de Instrumento Débil

Si la correlación ρ_{zx} disminuye:

$$\rho_{zx} \downarrow \Rightarrow SE(\hat{\beta}_1^{IV}) \uparrow \quad (15)$$

Implicaciones:

1. Mayor incertidumbre en la estimación
2. Intervalos de confianza más anchos
3. Menor poder estadístico en las pruebas

5.3 Demostración con la Relación F

El F-estadístico de primera etapa está relacionado con la correlación:

$$F \approx n \rho_{zx}^2 \quad (\text{aproximadamente, para un instrumento}) \quad (16)$$

Por lo tanto:

$$\text{SE}(\hat{\beta}_1^{\text{IV}}) \propto \frac{1}{\sqrt{F}} \quad (17)$$

Conclusión: El error estándar es inversamente proporcional a \sqrt{F} .

Si el F-estadístico disminuye de F_1 a F_2 , el error estándar aumenta en un factor:

$$\frac{\text{SE}_2}{\text{SE}_1} \approx \sqrt{\frac{F_1}{F_2}} \quad (18)$$

6 Ejercicio 4d: Instrumento Endógeno

6.1 Nuevo DGP con Instrumento Endógeno

DGP con violación de exogeneidad:

$$\begin{aligned} v_i &\sim N(0, 1) \quad (\text{variable no observada}) \\ z_{2i} &= v_i + \varepsilon_{zi}, \quad \varepsilon_{zi} \sim N(0, 1) \\ u_{2i} &= 0.1 v_i + \varepsilon_{ui}, \quad \varepsilon_{ui} \sim N(0, 1) \\ x_{2i} &= 0.2 z_{2i} + 4 u_{2i} + \varepsilon_{xi} \\ y_{2i} &= 3 x_{2i} + 5 u_{2i} \end{aligned} \quad (19)$$

Problema: Tanto z_{2i} como u_{2i} dependen de v_i (no observada).

6.2 Violación de Exogeneidad

$$\text{Cov}(z_{2i}, u_{2i}) = \text{Cov}(v_i + \varepsilon_{zi}, 0.1v_i + \varepsilon_{ui}) = 0.1 \text{Var}(v_i) = 0.1 \neq 0 \quad (20)$$

Conclusión: El instrumento z_2 NO es exógeno.

6.3 Inconsistencia del Estimador IV

Cuando $\text{Cov}(z, u) \neq 0$, el estimador IV converge a:

Límite en probabilidad con instrumento endógeno:

$$\text{plim}_{n \rightarrow \infty} \hat{\beta}_1^{\text{IV}} = \frac{\text{Cov}(z_i, y_i)}{\text{Cov}(z_i, x_i)} \quad (21)$$

Sustituyendo $y_i = \beta_1 x_i + \gamma u_i$ (donde $\gamma = 5$ en nuestro caso):

$$\begin{aligned} \text{plim} \hat{\beta}_1^{\text{IV}} &= \frac{\text{Cov}(z_i, \beta_1 x_i + \gamma u_i)}{\text{Cov}(z_i, x_i)} \\ &= \frac{\beta_1 \text{Cov}(z_i, x_i) + \gamma \text{Cov}(z_i, u_i)}{\text{Cov}(z_i, x_i)} \\ &= \beta_1 + \gamma \frac{\text{Cov}(z_i, u_i)}{\text{Cov}(z_i, x_i)} \end{aligned} \quad (22)$$

Sesgo asintótico:

$$\text{Sesgo} = \gamma \frac{\text{Cov}(z_i, u_i)}{\text{Cov}(z_i, x_i)} = 5 \times \frac{\text{Cov}(z_i, u_i)}{\text{Cov}(z_i, x_i)} \neq 0 \quad (23)$$

\Rightarrow El estimador IV es **INCONSISTENTE**.

6.4 Comparación: OLS vs IV con Instrumento Endógeno

Ambos estimadores están sesgados:

- **MCO:** sesgado por $\text{Cov}(x, u) \neq 0$
- **IV:** sesgado por $\text{Cov}(z, u) \neq 0$

¿Cuál es mejor?

Depende de la magnitud relativa de los sesgos:

$$\begin{aligned} \text{Sesgo}_{\text{OLS}} &= \frac{\text{Cov}(x, u)}{\text{Var}(x)} \\ \text{Sesgo}_{\text{IV}} &= \gamma \frac{\text{Cov}(z, u)}{\text{Cov}(z, x)} \end{aligned} \quad (24)$$

Es posible que: $|\text{Sesgo}_{\text{IV}}| > |\text{Sesgo}_{\text{OLS}}|$

\Rightarrow Un instrumento endógeno puede ser **PEOR** que MCO.

7 Ejercicio 4e: Mejora en Relevancia

7.1 Nuevo DGP con Relevancia Mejorada

DGP con instrumento endógeno pero más relevante:

$$\begin{aligned}v_i &\sim N(0, 1) \\z_{2i} &= v_i + \varepsilon_{zi} \\u_{2i} &= 0.1 v_i + \varepsilon_{ui} \\x_{3i} &= 3 z_{2i} + 4 u_{2i} + \varepsilon_{xi} \quad (\text{coeficiente aumentado a 3}) \\y_{2i} &= 3 x_{3i} + 5 u_{2i}\end{aligned}\tag{25}$$

Cambio: $0.2 \rightarrow 3$ en la ecuación de x .

7.2 Regla Práctica: Estadístico t

Regla práctica alternativa para relevancia:

Un instrumento se considera **RELEVANTE** si:

$$|t_{\text{first stage}}| > 3.16\tag{26}$$

donde t es el estadístico t del coeficiente del instrumento en la primera etapa.

Relación con la regla F:

Para un solo instrumento ($K = 1$):

$$F = t^2\tag{27}$$

Por lo tanto:

$$|t| > 3.16 \quad \Leftrightarrow \quad F > (3.16)^2 \approx 10\tag{28}$$

Ambas reglas son equivalentes.

7.3 ¿Por Qué el Estimador es Más Consistente?

Aunque el instrumento sigue siendo endógeno ($\text{Cov}(z_2, u_2) \neq 0$), el sesgo asintótico se reduce.

Descomposición del límite en probabilidad

$$\text{plim } \hat{\beta}_1^{\text{IV}} = \beta_1 + \gamma \frac{\text{Cov}(z_i, u_i)}{\text{Cov}(z_i, x_i)}\tag{29}$$

Componentes del sesgo:

- Numerador: $\text{Cov}(z_i, u_i)$ (endogeneidad del instrumento)
- Denominador: $\text{Cov}(z_i, x_i)$ (relevancia del instrumento)

Efecto de aumentar la relevancia

Cuando aumentamos el coeficiente de z en la ecuación de x :

- $\text{Cov}(z_i, u_i)$ permanece **igual** (no cambia la dependencia de v)
- $\text{Cov}(z_i, x_i)$ **aumenta** (mayor efecto directo)

Por lo tanto:

$$\text{Cov}(z_i, x_i) \uparrow \Rightarrow \frac{\text{Cov}(z_i, u_i)}{\text{Cov}(z_i, x_i)} \downarrow \Rightarrow |\text{Sesgo}| \downarrow \quad (30)$$

Intuición: Señal vs Ruido

- **Señal:** $\text{Cov}(z, x)$ = variación en x inducida por z
- **Ruido:** $\text{Cov}(z, u)$ = correlación espuria entre z y u

Ratio señal/ruido:

$$S/R = \frac{\text{Cov}(z, x)}{\text{Cov}(z, u)} \quad (31)$$

Al mejorar la relevancia:

- La señal aumenta
- El ruido permanece constante
- El ratio S/R mejora
- El sesgo asintótico disminuye

7.4 Fórmula del Sesgo Asintótico

Sesgo asintótico con instrumento endógeno:

$$\text{Sesgo}_\infty = \beta_1^{\text{IV}} - \beta_1 = \gamma \frac{\text{Cov}(z, u)}{\text{Cov}(z, x)} \quad (32)$$

Factores que reducen el sesgo:

1. **Disminuir** $\text{Cov}(z, u)$ (mejorar exogeneidad) \rightarrow IDEAL
2. **Aumentar** $\text{Cov}(z, x)$ (mejorar relevancia) \rightarrow AYUDA PERO NO RESUELVE

Conclusión:

- La exogeneidad ($\text{Cov}(z, u) = 0$) es **esencial**
- La relevancia ($\text{Cov}(z, x)$ grande) es **importante** pero **secundaria**

7.5 Ejercicio 4f: Discusión del F-estadístico

El F-estadístico de primera etapa es:

$$F = \frac{\text{ESS}/K}{\text{RSS}/(n - K - 1)} \quad (33)$$

Para un instrumento:

$$F \approx \frac{n \rho_{zx}^2}{1 - \rho_{zx}^2} \quad (34)$$

Interpretación de $F > 10$:

1. Relevancia confirmada:

- El instrumento tiene suficiente correlación con x
- La primera etapa es fuerte

2. Sesgo de muestra finita pequeño:

- El sesgo hacia MCO es mínimo
- La aproximación asintótica es buena

3. Inferencia válida:

- Los errores estándar son confiables
- Las pruebas t e intervalos de confianza son válidos

IMPORTANTE: $F > 10$ NO garantiza exogeneidad

Un F-estadístico alto indica:

- El instrumento es **relevante**
- No hay problemas de instrumento **débil**
- NO dice nada sobre **exogeneidad**

Es posible tener $F > 10$ y aún así $\text{Cov}(z, u) \neq 0$.

⇒ La exogeneidad debe argumentarse por otras vías:

- Conocimiento del contexto institucional
- Aleatorización o cuasi-aleatorización
- Pruebas con variables observables
- Argumentos económicos

8 Resumen: Jerarquía de Propiedades del Instrumento

Orden de importancia de las propiedades de un instrumento:

1. **EXOGENEIDAD** ($\text{Cov}(z, u) = 0$):

- **Absolutamente esencial**
- Sin exogeneidad, IV es inconsistente
- No hay forma de corregirlo en la estimación
- Debe argumentarse a priori

2. **RELEVANCIA** ($\text{Cov}(z, x) \neq 0$, $F > 10$):

- **Muy importante** para precisión
- Afecta sesgo de muestra finita y varianza
- Pero un instrumento débil es mejor que uno endógeno
- Puede mejorarse con controles o múltiples instrumentos

Casos extremos:

- Exógeno + Fuerte ($F > 10$): **IDEAL**
- Exógeno + Débil ($F < 10$): **USABLE** \bigcirc (con precaución)
- Endógeno + Fuerte: **MALO** (inconsistente)
- Endógeno + Débil: **MUY MALO** (peor que MCO)

8.1 Comparación de Escenarios

Escenario	$\text{Cov}(z, u)$	F	Consistencia	Uso
Base (ideal)	0	Alto	Sí	Recomendado
Débil	0	Bajo	Sí	\bigcirc Con precaución
Endógeno débil	$\neq 0$	Bajo	No	Evitar
Endógeno fuerte	$\neq 0$	Alto	No	Evitar

Table 1: Comparación de propiedades y usabilidad de instrumentos

9 Lecciones para la Práctica

9.1 Recomendaciones para Investigación Aplicada

1. Priorizar exogeneidad sobre relevancia:

- Buscar fuentes de variación "as-if random"
- Argumentar extensamente por qué $\text{Cov}(z, u) = 0$
- Usar aleatorización, discontinuidades, o variación histórica

2. Verificar relevancia:

- Reportar siempre F -estadístico de primera etapa
- Verificar $F > 10$ (o $|t| > 3.16$)
- Si $F < 10$, reconocer las limitaciones

3. Pruebas indirectas de exogeneidad:

- Probar correlación de z con variables observables (e.g., IQ, educación de padres)
- Usar controles para hacer z exógeno condicionalmente
- Realizar pruebas de sobreidentificación si hay múltiples instrumentos

4. Ser transparente sobre limitaciones:

- Discutir amenazas a la validez del instrumento
- Reconocer si el instrumento es débil
- Interpretar cuidadosamente (LATE vs ATE)

5. Comparar con MCO:

- Mostrar siempre estimaciones OLS e IV
- Discutir por qué difieren
- Si $IV \approx OLS$, discutir implicaciones

9.2 Trade-offs en Variables Instrumentales

Conclusión: Solo usar IV cuando:

- La endogeneidad es una preocupación seria
- Tenemos un instrumento creíblemente válido
- Estamos dispuestos a identificar LATE en vez de ATE

Aspecto	MCO	IV
Consistencia (si x endógena)	No	Sí (si z válido)
Eficiencia	Alta	Baja
Complejidad	Simple	Requiere instrumento
Interpretación	ATE (con supuestos)	LATE
Robustez	A especificación	A validez de z

Table 2: Trade-offs entre MCO y IV