Teoría Econométrica I - EAE- 250-A

Endogeneidad

Tatiana Rosá

Instituto de Economía - Pontificia Universidad Católica de Chile

Noviembre 2021

Introducción

- En esta sección vamos a levantar el supuesto de ortogonalidad entre los errores y los regresores ($\mathbb{E}(u|x) \neq 0$)
  - Vamos a hablar de regresores endógenos
  - Es quizás este uno de los aportes más grandes de la econometría a la estadística (y una diferencia!)
- Vamos a usar mucho el approach de Angrist y Pischke ("Mostly Harmless Econometrics")

Endogeneidad

- Un regresor es endógeno cuando $\mathbb{E}(u|x) \neq 0$
- Se viola el principio de identificación del OLS: ya no tendremos estimadores OLS insesgados ni consistentes
- Ya que

$$\hat{\beta}_{OLS} = \beta + \left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i'\mathbf{x}_i\right]^{-1} \left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i' u_i\right]$$

- Parámetro esta identificado: Cuando podemos expresar el parámetro como una función de momentos poblacionales de variables observables

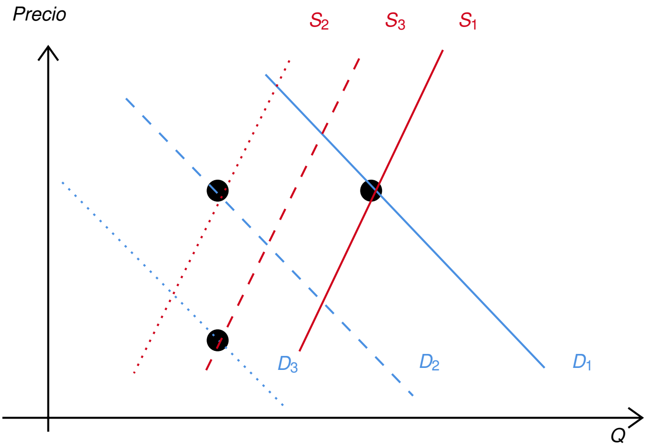
Causas de la endogeneidad

- En econometría de grado vemos que existen diferentes causas por las que $\mathbb{E}(u|x) \neq 0$
  - Error de medida: sesgo de atenuación
  - Simultaneidad: Oferta y demanda
  - Omisión de variables relevantes: generan sesgo de omisión de variables (OVB por sus siglas en inglés)
- Con regresores endógenos no podremos hablar de causalidad
- La preocupación de esto empieza en 1920 cuando Wright quiere estimar la pendiente de la curva de oferta y demanda del mercado agrícola

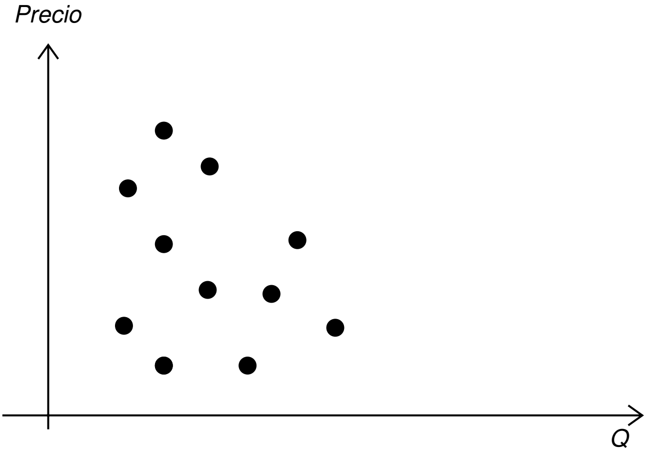
Ejemplo: Estimación de oferta y demanda

- Queremos estimar la demanda de un bien agrícola:

$$ln(Q_i) = \beta_0 + \underbrace{\beta_1}_{\text{Elastcidad precio}} ln(Pr_i) + u_i$$

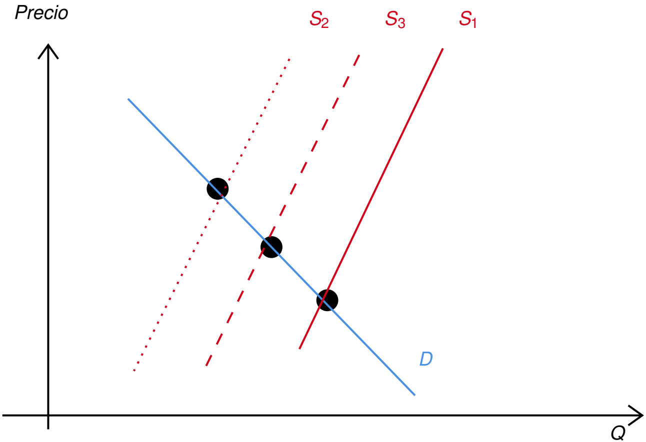


Interacción de la curva de oferta y de demanda



- Este es el scatterpolot de los precios y cantidades de equilibrio en los distintos períodos
- ¿Podríamos estimar una curva de demanda a partir de estos puntos?

¿Y si solo se moviera la curva de oferta?



- La idea es usar un **instrumento** que aísle los cambios en precios y cantidades que se deben exclusivamente a cambios en la oferta
- Un instrumento aquí sería algo que mueva la oferta y no la demanda

Variables Instrumentales

- Esta literatura tiene origen en los sistemas de ecuaciones simultaneas (SEMs por sus siglas en inglés)
- De ahí mucha de la terminología que usamos hoy cuando aplicamos variables instrumentales (IV)
- Incluso cuando las aplicaciones más populares de instrumentos sean para solucionar los problemas de errores de medida y OVB

Variables Instrumentales

- Pensemos en la ecuación de salarios y educación en la población (ecuación estructural)

$$f_i = \alpha + \rho s_i + \eta_i$$

- Podemos pensar esto en el marco de outcomes potenciales (A&P): observamos un salario para un nivel de educaciónpero ese individuo tendría otro salario (no observado) con otro nivel de educación. El efecto de la educación sería la diferencia entre esos salarios
- Ahora imaginamos que este error está compuesto de una variable inobservable *A*, llamada *habilidad* que nos da este selección en observables:

$$\eta_i = A_i' \gamma + \nu_i$$

donde γ es son los coeficientes poblaciones y A_i y ν_i son ortogonales por construcción

Variables Instrumentales

- η_i no será ortogonal con s_i si la habilidad y la educación están correlacionas
- Pero tenemos entonces $\mathbb{E}(s_i \nu_i) = 0$
- Podemos escribir la ecuación larga

$$f_i = \alpha + \rho s_i + A_i' \gamma + \nu_i$$

- Si pudieramos observar A_i podríamos estimar esta ecuación por OLS de forma consistente y su interpretación sería causal
- Pero como no lo observamos: vamos a usar un instrumento z_i que debe cumplir:
  1. $\mathbb{E}(z_i' \eta_i) = 0$: Restricción de exclusión
  2. $\mathbb{E}(z_i' s_i) \neq 0$: Relevancia

Estimador de Variables instrumentales

- Pre multiplicamos la ecuación estructural:

$$z_i' f_i = z_i' \alpha + \rho z_i' s_i + z_i' \eta_i$$

- escribimos todos los regresores juntos (x_i)

$$z_i' f_i = z_i' x_i \beta + z_i' \eta_i$$

- Tomamos esperanzas y usamos la restricción de exclusión:

$$\mathbb{E}(z_i' x_i)\beta = \mathbb{E}(z_i' y_i)$$

- donde $\mathbb{E}(z_i' x_i)$ es de orden $K \times K$ por tanto tenemos el mismo número de ecuaciones que de parámetros (en nuestro ejemplo 2)
- Invertimos la ecuación y encontramos una expresión para β

$$\beta = \mathbb{E}(z_i' x_i)^{-1}\mathbb{E}(z_i' y_i)$$

- Por principio de analogía:

$$\hat{\beta}_{IV} = \left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i' \mathbf{x}_i\right]^{-1} \left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i' y_i\right] \equiv (Z'X)^{-1} Z'Y$$

Forma reducida y primera etapa

- Otra manera de pensar este estimador es pensar la identificación del β partiendo de la restricción de exclusión:

$$Cov(\eta_i, z_i) = 0$$

$$Cov(y_i - \alpha - \rho s_i, z_i) = Cov(y_i, z_i) - \rho Cov(s_i, z_i) = 0$$

$$\rho = \frac{Cov(y_i, z_i)}{Cov(s_i, z_i)} = \frac{Cov(y_i, z_i)/V(z_i)}{Cov(s_i, z_i)/V(z_i)}$$

- El termino de arriba lo podemos pensar como el coeficiente poblacional de regresar y_i sobre z_i (forma reducida)
- El denominador lo podemos pensar como el coeficiente poblacional de regresar x_i sobre z_i (primera etapa)
- Nuevamente el estimador de VI sería el análogo en la muestra de esta expresión

$$\hat{\beta}_{IV} = \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i'\mathbf{x}_i\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i'y_i\right] \equiv (Z'X)^{-1}Z'Y$$

Supuestos del Instrumento e Instrumentos débiles

- Cuando buscamos instrumentos para una variable endógena, ambos supuestos son igualmente importantes para identificar β
- Sin embargo,la restricción de exclusión no puede ser testeada y debe ser mantenido. Tip: Pensar en variables determinadas aleatoriamente
- El supuesto de relevancia puede y debe ser testeado. Más adelante veremos que es relativamente sencillo hacerlo y no requiere más tecnología que un test-t o F
- Cuando la correlación de las variables instrumentales con las endógenas es pequeña se dice que estamos en presencia de **instrumentos débiles** o ***weak instruments***.

Consistencia y Normalidad Asintótica del estimador VI

- Podemos escribir el estimador de variables instrumentales como sigue:

$$\hat{\beta}_{IV} = \beta + \left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i'\mathbf{x}_i\right]^{-1} \left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i' u_i\right]$$

- Aplicamos LLN para probar consistencia:

Distribución asintótica

- Podemos generar la expresión típica ajustada por $\sqrt{n}$:

$$\sqrt{n}(\hat{\beta}_{IV} - \beta) = \left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i'\mathbf{x}_i\right]^{-1} \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbf{z}_i' u_i\right]$$

- El primer término del lado derecho de la ecuación convergerá a $\mathbb{E}(\mathbf{z}'\mathbf{x}) = M_{zx}$ por LLN
- El segundo término converge en distribución a una Normal por el Teorema central del límite (CLT)

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbf{z}_i' u_i \xrightarrow{d} \mathcal{N}(0, V_0)$$

donde $V_0 = \mathbb{E}(u^2\mathbf{z}'\mathbf{z})$. Por lo tanto, por LLN y CLT tenemos que

$$\sqrt{n}(\hat{\beta}_{IV} - \beta) \xrightarrow{d} \mathcal{N}(0, M_{zx}^{-1} V_0 (M_{zx}^{-1})')$$

- Para obtener un estimador de la varianza debemos usar el *sample analog* de M_{zx} que es trivial y un estimador para V_0 (White, Newey West)

Ejemplos de variables endógenas e instrumentos

- **Retornos de la educaciónon:**
  - dummy trimestre de nacimiento (Angrist and Krueger, 1991)
  - proximidad geográfica con alguna Universidad (Card, 1995)
- **Retornos participación en programas de capacitación:**

$$Salario_i = \beta_0 + \beta_1 Participa_i + u_i$$

  - sorteos aleatorios si los hay
- **Elasticidad precio de la demanda**

$$Cantidad_i = \beta_0 + \beta_1 Precio_i + u_i$$

  - Caracteristicas de bienes similares (Berry, Levinsohn and Pakes, 1995)
- **Efecto de fumar durante el embarazo en la salud de los nacidos:**

$$PesoNacer_i = \beta_0 + \beta_1 Cigarrillos_i + u_i$$

  - Cambios en los impuestos sobre el tabaco (Liens and Evans,2005)

Múltiples instrumentos y Mínimos cuadrados en dos etapas

- Sea Z la matriz de instrumentos de orden $n \times L$ y X es la matriz de variables independientes o explicativas de orden $n \times K$
- Cuando hay más instrumentos que variables endógenas (o más de un instrumento para una variable endógena) tenemos el caso “sobre identificado”: más ecuaciones que incógnitas
- Opción poco eficiente: descartar instrumentos
- Opción más eficiente: combinarlos para que las matrices queden conformables
- Armamos una combinación lineal de los instrumentos utilizando la matriz Π tal que $Z\Pi$ sea de dimensión $n \times K$

Mínimos cuadrados en dos etapas

- Premultiplicando el modelo tenemos que:

$$\Pi' z_i y_i = \Pi' z_i' x_i \beta + \Pi' z_i' u_i$$

- Explotando la condición de identificación $\mathbb{E}(z_i' u_i) = 0$ podemos identificar β tomando valor esperado,

$$\beta = [\mathbb{E}(\Pi' \mathbf{z}' \mathbf{x})]^{-1} \mathbb{E}[\Pi' \mathbf{z}' y]$$

- Siguiendo el *analogy principle* el estimador estaría dado por

$$\hat{\beta} = \left[\frac{1}{n} \sum_{i=1}^{n} \Pi' \mathbf{z}_i' \mathbf{x}_i \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^{n} \Pi' \mathbf{z}_i' y_i \right] \equiv (\Pi' Z' X)^{-1} \Pi' Z' y$$

- No hemos dicho nada de la matriz Π. Esta puede ser desconocida para lo cual necesitaremos un estimador de Π

Mínimos cuadrados en dos etapas

- Supongamos que tenemos un estimador de Π dado por $\hat{\Pi}$, luego el estimador **generalizado de variables instrumentales (GIV)** (de la familia del GMM) está dado por

$$\hat{\beta}_{GIV} = \left[\frac{1}{n}\sum_{i=1}^{n}\hat{\Pi}'\mathbf{z}_i'\mathbf{x}_i\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\hat{\Pi}'\mathbf{z}_i'y_i\right] \equiv (\hat{\Pi}'Z'X)^{-1}\hat{\Pi}'Z'y$$

- La elección clásica de matriz es $\hat{\Pi} = (Z'Z)^{-1}Z'X$ que corresponde al estimador MCO de la regression de X sobre Z, así el estimador de mínimos cuadrados en dos etapas (2SLS) es,

$$\begin{aligned} \hat{\beta}_{2SLS} &= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y && (1) \\ &= (\hat{X}'\hat{X})^{-1}\hat{X}'y && (2) \end{aligned}$$

donde $\hat{X} = Z(Z'Z)^{-1}Z'X$. Note que el nombre de mínimos cuadrados en dos etapas (2SLS) viene de la interpretación que muestra que el estimador se puede obtener en dos etapas

Estimación en dos etapas

- Tenemos:

$$\hat{\beta}_{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y \tag{3}$$

- se puede escribir como un sistema de ecuaciones donde tenemos la primera etapa y luego la ecuación estructural:

$$\begin{aligned} X &= Z\Pi + e & (4) \\ y &= X\beta + u & (5) \end{aligned}$$

- Se estima la primera etapa y se ocupan los valores estimados $\hat{X} = Z\hat{\Pi} = Z(Z'Z)^{-1}Z'X$ en la segunda etapa.
- **Observación:** Note que si reemplazamos la primera etapa en la ecuación estructural obtenemos

$$\begin{aligned} X &= Z\Pi\beta + (u + e\beta) & (6) \\ y &= Z\gamma + \tilde{u} & (7) \end{aligned}$$

y esta expresión se conoce como la **forma reducida**. Cuando $K = L = 1$ se puede recuperar β dividiendo el parámetro de la forma reducida por el de la primera etapa $\beta = \gamma/\Pi$

Consistencia y Normalidad Asintótica

- Si $\frac{1}{\sqrt{N}} Z' u \xrightarrow{d} (0, V_0)$ donde $V_0 = \mathbb{E}(u^2 \mathbf{z}'\mathbf{z})$ y $\hat{\Pi} \xrightarrow{p} \Pi$ es fácil demostrar que

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} \mathcal{N}(0, [\Pi' M_{zx}]^{-1} \Pi' V_0 \Pi [\Pi' M_{zx}^{-1}]')$$

- Expresión que depende de Π, V_0 y M_{zx}
- No conocemos Π, podemos estimarla como la proyección ortogonal de X sobre Z
- Por ley de grandes números sabemos que

$$\hat{\Pi} \xrightarrow{p} \Pi \equiv [\mathbb{E}(\mathbf{z}'\mathbf{z})]^{-1} \mathbb{E}(\mathbf{z}'\mathbf{x}) = M_{zz}^{-1} M_{zx}$$

- Estos momentos podemos estimarlos consistentemente y reemplazando tenemos:

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} \mathcal{N}(0, [M'_{xz} M_{zz}^{-1} M_{zx}]^{-1} M'_{xz} M_{zz}^{-1} V_0 M_{zz}^{-1} M_{zx} [M'_{xz} M_{zz}^{-1} M_{zx}]')$$

Varianza Asintótica con u y z independientes

- Si u es independiente de **z**, la fórmula se simplifica debido a que $V_0 = \sigma^2 M_{zz}$, entonces:

- Y llegamos a :

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2 [M'_{xz} M_{zz}^{-1} M_{zx}]^{-1})$$

- ¿Cómo estimamos σ^2? Tenemos $\hat{u} = y - \mathbf{x}\hat{\beta}_{2SLS}$

$$\hat{\sigma}^2 = \frac{1}{n-k} \sum_{i=1}^{n} \hat{u}_i^2$$

- **Ojo! no son los residuos de la segunda etapa!!**

Varianza Asintótica con u y z no independientes

- Si u y $\mathbf{z}$ no son independientes (a nivel de distribución, pero $\mathbb{E}(\mathbf{z}'u) = 0$) se puede estimar consistentemente la matriz de varianzas y covarianzas usando Eicker-Huber-White o Newey-West dependiendo si los errores son autocorrelacionados
- El estimador de Eicker-Huber-White de la varianza asintótica de $\hat{\beta}_{2SLS}$ está dado por

$$Avar(\beta_{2SLS}) = (\hat{X}'\hat{X})^{-1}\left(\sum_{i=1}^{n} \hat{u}_i^2 \hat{\mathbf{x}}_i' \hat{\mathbf{x}}_i\right)(\hat{X}'\hat{X})^{-1} \quad (8)$$

- En Stata 11-14, usando el comando `ivregress 2sls` con la opción `vc`$\mathbb{E}$(*robust*) nos entregará los errores estándar de la matriz recién descrita. También puede usar el comando `ivreg2` con la opción `robust`

Método Generalizado de los Momentos: repaso

- Un estimador alternativo a 2SLS en presencia de endogeneidad y variables instrumentales es el estimador de GMM.
- Definamos las **condiciónes de momento** como

$$m(\mathbf{z}, \mathbf{x}, \beta) = \mathbf{z}'(y - \mathbf{x}\beta)$$

  donde $\mathbf{z}$ es una realización del vector de $L \times 1$ de instrumentos y $\mathbf{x}$ es una realización del vector de variables endógenas de $K \times 1$
- Suponiendo que $L > K$ tenemos un sistema sobre identificado. Dado el supuesto de identificación $\mathbb{E}(\mathbf{z}u) = 0$ tenemos que,

$$\mathbb{E}(m(\mathbf{z}, \mathbf{x}, \beta)) = 0$$

- El valor esperado de cada condición de momento es cero.
- Cada condición de momento poblacional tiene su contraparte muestral dada por,

$$\overline{m}(\mathbf{z}, \mathbf{x}, \beta) = \frac{1}{n} \sum \mathbf{z}'(y - \mathbf{x}\beta) = Z'u$$

GMM

- Tenemos el mismo problema que antes: si $L \geq K$ tenemos más condiciones que incógnitas
- En el caso que $L = K$ se tiene un sistema exactamente identificado y luego la solución está dada por $\overline{m}(\mathbf{z}, \mathbf{x}, \beta) = 0$ con lo cual se obtiene la misma solución que $\hat{\beta}_{IV}$
- Para el caso sobre-identificado el estimador de GMM es aquel que minimiza la siguiente forma cuadrática,

$$\min_{\beta} n \, \overline{m}(\mathbf{z}, \mathbf{x}, \beta)' \, \mathrm{W}^{-1} \, \overline{m}(\mathbf{z}, \mathbf{x}, \beta)$$

donde W^{-1} es una matriz de $L \times L$ con lo cual el sistema es de $K \times K$.

GMM eficiente

- Se define el **estimador de GMM eficiente (EGMM)** aquel que utiliza como weighting matrix

$$W = Var(m(\mathbf{z}, \mathbf{x}, \beta)) = \mathbb{E}(u^2 \mathbf{z}' \mathbf{z}) = V_0$$

- En el caso que $W = V_0$ tenemos que $\hat{\beta}_{GMM}$ minimiza la siguiente expresión

$$\min_{\beta} \frac{1}{n} (y - X\beta)' Z V_0^{-1} Z' (y - X\beta)$$

- Luego,

$$\hat{\beta}_{EGMM} = [X' Z V_0^{-1} Z' X]^{-1} X' Z V_0^{-1} Z' y$$

- Sólo nos falta un estimador consistente de V_0
- Bajo el supuesto de heterocedasticidad, podemos usar el estimador de Eicker-White antes mencionado, con lo cual la varianza asintótica estará dada por

$$Avar(\hat{\beta}_{EGMM}) = (M_{zx}' V_0^{-1} M_{zx})^{-1}$$

Implementación del EGMM

- Se puede implementar el estimador EGMM en tres etapas:
  1. Estime el modelo por 2SLS y obtenga los residuos de la manera antes descrita $\hat{u} = y - \mathbf{x}\hat{\beta}_{2SLS}$.
  2. Construya la matriz $\hat{V}_0 = \frac{1}{n}\sum \hat{u}_i^2 \mathbf{z}_i' \mathbf{z}_i$.
  3. Estime mediante EGMM usando $\hat{V}_0$ como weighting matrix

  En Stata esto se puede implementar con el comando `ivreg2` con la opción `gmm`.

2SLS y EGMM

- En el caso general, cuando los errores son heterocedásticos y/o autocorrelacionados y $V_0 \neq \sigma^2 M_{zz}$ el estimador 2SLS no tendrá la menor varianza asintótica
- Para obtener un estimador eficiente necesitamos escoger una matriz Π que minimize la varianza asintótica.
- Queremos minimizar con respecto a Π la siguiente expresión,

$$Avar(\hat{\beta}_{GIV}) = [\Pi' M_{zx}]^{-1} \Pi' V_0 \Pi([\Pi M_{zx}]^{-1})'$$

- Se puede demostrar (bastante engorroso) que

$$\Pi^* = V_0^{-1} M_{zx} = \text{argmin}_{\Pi} Avar(\hat{\beta}_{GIV}(\Pi))$$

- No disponemos disponer de Π^*, falta V_0 y M_{zx}

2SLS y EGMM

- La ley débil de los grandes números nos garantiza que si $\{x_i z_i\}$ son *i.i.d* con primer y segundo momento acotados,

$$\hat{\Pi}^* = V_0^{-1}\left(\frac{1}{n}\sum \mathbf{z}_i'\mathbf{x}_i\right) \xrightarrow{p} V_0^{-1}M_{zx} = \ Pi^*$$

el estimador eficiente 2SLS (EGIV) corresponde al estimador EGMM y es igual a

$$\hat{\beta}_{EGIV} = \hat{\beta}_{EGMM} = [X'ZV_0^{-1}Z'X]^{-1}X'ZV_0^{-1}Z'y$$

con distribución asintótica

$$\sqrt{n}(\hat{\beta}_{EGMM} - \beta) \xrightarrow{d} \mathcal{N}(0, (M_{zx}'V_0^{-1}M_{zx})^{-1})$$

Test de Sobre identificación

- Cuando tenemos más instrumentos que variables endógenas decimos que el modelo esta sobre identificado
- Podemos usar esas reestricciones ”extra” para testear la validez de los instrumentos (hasta un cierto punto)
- En el caso de GMM, esto se hace testeando que las condiciones de momento muestrales, en conjunto, no sean diferentes de cero (en términos estadísticos)
- Limitación: vamos a estar testeando conjuntamente la ”exogeneidad” y la forma funcional. Podría ser que los instrumentos sean válidos pero el modelo este más especificado en sentido de las restricciones de exclusión
- En los test de sobre identificación la nula siempre es que todos los instrumentos son válidos frente a la alternativa que alguno no lo es.

Test de Hansen(1982)

- El estadístico no es más que la función objetivo evaluada en $\hat{\beta}_{EGMM}$ y se distribuye como χ^2_{L-k}, así

$$J(\hat{\beta}_{EGMM}) = n\,\overline{m}(\mathbf{z}, \mathbf{x}, \hat{\beta}_{EGMM})'\,\hat{V}_0^{-1}\overline{m}(\mathbf{z}, \mathbf{x}, \hat{\beta}_{EGMM}) \xrightarrow{d} \chi^2_{L-K}$$

- Valor grandes de este estadístico nos harán rechazar H_0: problemas o bien de los instrumentos o de la especificación del modelo

Test de Sargan

- Para el caso del $\hat{\beta}_{2SLS}$ tenemos el test de Sagan(1952)

$$Sargan = \frac{\hat{u}' P_Z \hat{u}}{\hat{u}' \hat{u}/n} \xrightarrow{d} \chi^2_{L-K}$$

donde $P_z = Z(Z'Z)^{-1}Z$ es la matriz de proyección sobre Z

- Una manera sencilla de obtener el estadístico de Sargan es correr el modelo por 2SLS, obtener $\hat{u}$. Correr la regressión auxiliar de $\hat{u}$ sobre todas las variables exógenas (**x** y **z**'s) y obtener el R^2. Se puede demostrar que $Sargan = n \times R^2$

Fallas de identificación

- Considere el caso en que $K = L = 1$. Luego, el modelo puede ser escrito de la siguiente manera
$$y_i = x_i\beta + e_i$$
$$x_i = z_i\pi + u_i$$
donde ey u son i.i.d normalmente distribuidos y $\pi = \mathbb{E}(z_i x_i)/\mathbb{E}(z_i^2)$.
- Podemos ver que β está identificado si y sólo si $\pi \neq 0$, lo cual ocurre cuando $E(z_i x_i) \neq 0$
- Suponga que esta condición falla, luego $\mathbb{E}(z_i x_i) = 0$.
- Note que por CLT,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i e_i \xrightarrow{d} N_1 \sim \mathcal{N}(0, \mathbb{E}(z_i^2 e_i^2))$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i x_i = \frac{1}{n} \sum_{i=1}^{n} z_i u_i \xrightarrow{d} N_2 \sim \mathcal{N}(0, \mathbb{E}(z_i^2 u_i^2))$$

Fallas de identificación

- Entonces
$$\hat{\beta} - \beta = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i e_i}{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i x_i} = \frac{N_1}{N_2} \sim Cauchy$$
dado que el cuociente de dos variables aleatorias normalmente distribuidas sigue una distribución de Cauchy
- Esto es grave puesto que $(\hat{\beta} - \beta)$ no converge a cero y el estimador es inconsistente
- Esto es particularmente desagradable puesto que la distribución de Cauchy no tiene primer momento finito.
- Este resultado se puede extender para otros casos y fue examinado por **?** y Choi (1992).

Instrumentos débiles

- Cuando tenemos instrumentos débiles puede ser peor el remedio que la enfermedad
- Frente a mínimas desviaciones en la restricción de exclusión el estimador por VI puede ser inconsistente
- La varianza de los estimadores por VI se incrementará considerablemente (IC muy grandes)
- Para el trabajo empírico se utiliza muchas veces una ”regla”: se considera que los instrumentos son débiles si el F test es menor que 10
- Esta regla esta basada en el trabajo de Steigner y Stock (1997) sobre el coeficiente de concentración pero no está aplicada correctamente ya que si bien el estadístico F esta relacionado con el parámetro de concentración el mismo es un estimador sesgado de este parámetro

El parámetro de concentración

- Considere el siguiente *setup* para un regresor endógeno

$$y = X\beta + e \tag{9}$$
$$X = Z\Pi + u \tag{10}$$

donde, y es un vector de $n \times 1$, X es una matriz de $n \times 1$, Z es una matriz de $n \times l$ y e y u son vectores de $n \times 1$ con varianzas σ_e^2 y σ_u^2 respectivamente y con coeficiente de correlación igual a ρ.

- Una medida de fortaleza de los instrumentos está dada por el **parámetro de concentración** (*signal to noise ratio*),

$$\mu^2 = \Pi' Z' Z\Pi / \sigma_u^2 \tag{11}$$

- Este parámetro está muy relacionado con el estadístico F de la primera etapa para testear la hipótesis de relevancia $\Pi = 0$. Se puede demostrar que valores grandes de μ^2/k mueven la distribución del estadístico F y que $F - 1$ es un estimador de μ^2/k.

El parámetro de concentración y el estimador 2SLS

- El estimador 2SLS minimiza $(y - X\beta)' P_Z (y - X\beta)$ y se define como $\hat{\beta}_{2SLS} = (X' P_Z X)^{-1}(X' P_Z y)$ donde P_Z es la matriz de proyección a las columnas del espacio de las Z
- Rottenberg (1984) muestra que a medida que μ^2 crece, el estimador 2SLS converge en probabilidad y su distribución es estándar

$$\mu(\hat{\beta}_{2SLS} - \beta) = (\sigma_e/\sigma_u)\frac{z_e + S_{eu}/\mu}{1 + 2z_u/\mu + S_u u/\mu^2}$$

donde

$$z_e = (\Pi' Z' e)/(\sigma_e \sqrt{\Pi' Z' Z \Pi}); \qquad z_u = (\Pi' Z' u)/(\sigma_u \sqrt{\Pi' Z' Z \Pi})$$
$$S_{eu} = (u' P_Z e)/(\sigma_u \sigma_v); \qquad S_{uu} = (u' P_Z u)/\sigma_u^2$$

- Se puede demostrar que bajo los supuestos de instrumentos fijos y errores normales, z_e y z_u son variables aleatorias normales con coeficientes de correlación ρ, y S_{eu} y S_{uu} son formas cuadráticas de variables aleatorias normales con respecto a la matriz de proyección P_z.

Parámetro de concentración igual a cero

- Note que μ^2 juega el rol del tamaño muestral, i.e. si μ^2 es suficientemente grande entonces tenemos la aproximación normal usual. Por otro lado, si μ^2 es pequeño, la distribución asintótica no es estándar.
- Es fácil demostrar que cuando $\mu^2 = 0$, entonces $\mathrm{plim}(\hat{\beta}_{2SLS}) = \beta + (\sigma_u^2/\sigma_v^2)\rho$. Este caso extremo deja en evidencia cuán sensible puede ser el estimador 2SLS a la fuerza de los instrumentos
  El siguiente ejercicio de **Monte Carlo** considerado por Staigner-Stock:1997, Nelson-Starz:1990, y Stock-Wright-Yogo:2002 nos puede ayudar a ver cuán sensible es la distribución al parámetro de concentración.

Montecarlo para la consistencia del estimador 2SLS

- El modelo es

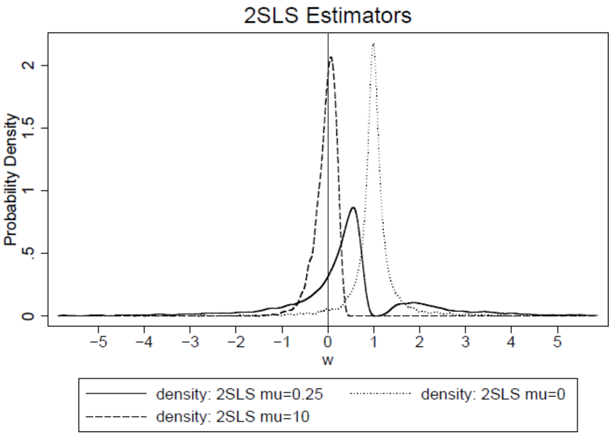
$$y = \beta x + e \tag{12}$$
$$X = \pi z + v \tag{13}$$

- El verdadero valor de $\beta = 0$, lo errores son i.i.d normalmente distribuidos con medias iguales a 0, varianzas iguales a 1 y covarianza $\rho = 0,99$.
- Para cada muestra generada fijamos $n = 20$. Se fija un valor para μ^2 y se generan número aleatorios para **z**, con ellos se obtiene un valor para π

$$\pi^2 = \mu^2 / \sum_{i=1}^{n} z_i^2$$

- se construye **x** y luego **y** generando (e,v) de una normal bivariada
- Se estima β mediante 2SLS. Este procedimiento se hace 10,000 veces.

Montecarlo


2SLS Estimators
Probability Density
0
.5
1
1.5
2
−5
−4
−3
−2
−1
0
1
2
3
4
5
w
density: 2SLS mu=0.25
density: 2SLS mu=0
density: 2SLS mu=10