Clase 10.A: Mapa de ruta

Introducción: Amenazas a la validez interna

- Consideremos el modelo de regresión lineal simple

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- Sabemos que $\mathbb{E}(u_i \mid X_i) \neq 0$ implica que los estimadores OLS de β_0 y β_1 son sesgados e inconsistentes.
  - Y el problema se extiende al modelo de regresión lineal múltiple.
- $\mathbb{E}(u_i \mid X_i) \neq 0$ puede tener diversas fuentes:
  - Variables omitidas
  - Errores de medición
  - Causalidad Simultánea
- En esta clase estudiaremos el **Estimador de Variables Instrumentales** (**VI**), que permite atacar este problema.

Variables instrumentales

- Variables Instrumentales (VI) es un método general para obtener estimadores consistentes de los coeficientes poblacionales del modelo de regresión lineal

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

  cuando X_i está correlacionada con u_i (independientemente de la fuente de esta correlación).
  - Cuando X_i está correlacinada (no correlacionada) con u_i decimos que X_i es endógena (exógena).
- Recordemos que $Cov\left(X_i, u_i\right) \neq 0 \Rightarrow \mathbb{E}(u_i \mid X_i) \neq 0$, provocando que los estimadores MCO sean sesgados e inconsistentes.
- Para usar VI necesitamos encontrar una nueva variable, Z, que llamaremos variable instrumental o instrumento (para X).
- ¿Qué clase de variable califica como un (buen) instrumento?

Variables instrumentales

- Un instrumento Z es válido si satisface dos condiciones:
  1. Relevancia
  $$Cov\left(Z_i, X_i\right) \neq 0$$
  2. Exogeneidad
  $$Cov\left(Z_i, u_i\right) = 0$$
- Si un instrumento es relevante, entonces la variación en el instrumento está relacionada con la variación en X_i. Si, además, el instrumento es exógeno, entonces esa parte de la variación de X_i captada por la variable instrumental es exógena. Por tanto, un instrumento que sea relevante y exógeno puede captar los movimientos de X_i que son exógenos. Esta variación exógena a su vez puede ser utilizada para estimar los coeficientes poblacionales β_0 y β_1.
- Como veremos, la relevancia y exogeneidad del instrumento son vitales para el funcionamiento del método de variables instrumentales.

Variables instrumentales

- Consideremos el modelo de regresión lineal simple:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- Luego:

$$\begin{aligned} Cov(Z_i, Y_i) &= Cov(Z_i, \beta_0 + \beta_1 X_i + u_i) \\ &= Cov(Z_i, \beta_0) + Cov(Z_i, \beta_1 X_i) + Cov(Z_i, u_i) \\ &= \beta_1 Cov(Z_i, X_i) + Cov(Z_i, u_i) \end{aligned}$$

- Si Z_i satisface la condición de exogeneidad $Cov(Z_i, u_i) = 0$, la expresión anterior se reduce a

$$Cov(Z_i, Y_i) = \beta_1 Cov(Z_i, X_i)$$

El estimador VI

- Supongamos, además, que Z_i satisface la condición de relevancia $Cov(Z_i, X_i) \neq 0$. Entonces, la expresión enterior implica:

$$\beta_1 = \frac{Cov(Z_i, Y_i)}{Cov(Z_i, X_i)} = \frac{\sigma_{ZY}}{\sigma_{ZX}}$$

- El estimador VI de β_1 se obtiene reemplazando las covarianzas poblacionales por las covarianzas muestrales:

$$\widehat{\beta}_1^{\text{vi}} = \frac{s_{ZY}}{s_{ZX}}$$

donde

$$s_{ZY} = \tfrac{1}{n-1} \sum_{i=1}^{n} (Z_i - \overline{Z})(Y_i - \overline{Y})$$

$$s_{ZX} = \tfrac{1}{n-1} \sum_{i=1}^{n} (Z_i - \overline{Z})(X_i - \overline{X})$$

  - Si X_i es exógena, entonces $Z_i = X_i$ (X_i es su propio instrumento) y la expresión anterior se reduce al estimador de MCO: $\widehat{\beta}_1 = \frac{s_{XY}}{s_X^2}$.

El estimador VI

- Reemplazando las fórmulas para s_{ZY} y s_{ZX} en $\widehat{\beta}_1^{\text{vi}} = \frac{s_{ZY}}{s_{ZX}}$ obtenemos:

$$\widehat{\beta}_1^{\text{iv}} = \frac{\sum_{i=1}^{n}(Z_i - \overline{Z})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})}$$

- El estimador VI de β_0 es simplemente

$$\widehat{\beta}_0^{\text{vi}} = \overline{Y} - \widehat{\beta}_1^{\text{vi}}\overline{X}$$

como en OLS, pero usando $\widehat{\beta}_1^{\text{vi}}$ en lugar de $\widehat{\beta}_1$.

- El valor de predicción es:

$$\widehat{Y}_i = \widehat{\beta}_0^{\text{vi}} + \widehat{\beta}_1^{\text{vi}} X_i$$

- Y el residuo es:

$$\begin{aligned} \widehat{u}_i &= Y_i - \widehat{Y}_i \\ &= Y_i - \widehat{\beta}_0^{\text{vi}} - \widehat{\beta}_1^{\text{vi}} X_i \end{aligned}$$

Supuestos de VI

- Recordemos el modelo de regresión poblacional:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

donde $\mathbb{E}\left(u_i \mid X_i\right) \neq 0$, y Z_i es un instrumento para X_i.

  - VI Supuesto 1 (S1): **Validez del Instrumento**

$$Cov\left(Z_i, X_i\right) \neq 0 \qquad \text{(relevancia)}$$

$$Cov\left(Z_i, u_i\right) = 0 \qquad \text{(exogeneidad)}$$

  - VI Supuesto 2 (S2): **Muestreo Aleatorio Simple**

$\left(X_i, Z_i, Y_i\right)$ son extracciones *iid* de su distribución conjunta

  - VI Supuesto 3 (S3): **Valores Extremos Improbables**

X_i, Z_i, Y_i tienen momentos de cuarto orden no nulos y finitos

Propiedades del Estimador VI

- Bajo VI S1-S3, $\widehat{\beta}_1^{\text{vi}}$ es **consistente** para β_1.
  - Como $s_{ZY} \xrightarrow{p} \sigma_{ZY}$ y $s_{ZX} \xrightarrow{p} \sigma_{ZX}$, se cumple:

$$\widehat{\beta}_1^{\text{vi}} = \frac{s_{ZY}}{s_{ZX}} \xrightarrow{p} \frac{\sigma_{ZY}}{\sigma_{ZX}} = \beta_1$$

  - Ver el Apéndice I para un argumento más formal.
- Bajo IV S1-S3, el estimador VI es **sesgado** (ver Apéndice I).
- Bajo IV S1-S3, el estimador VI es **asintóticamente normal**:

$$\widehat{\beta}_1^{\text{vi}} \overset{a}{\sim} N\left(\beta_1, \sigma^2_{\widehat{\beta}_1^{\text{vi}}}\right)$$

con

$$\sigma^2_{\widehat{\beta}_1^{\text{vi}}} = \frac{1}{n}\frac{Var[(Z_i - \mu_Z)u_i]}{[Cov(Z_i, X_i)]^2}$$

donde $Cov\left(Z_i, X_i\right) \neq 0$ por la relevancia del instrumento (VI S1).

  - Recordemos que la varianza (asintótica) del estimador MCO es $\sigma^2_{\widehat{\beta}_1} = \frac{1}{n}\frac{Var[(X_i - \mu_X)u_i]}{[Var(X_i)]^2}$.

Propiedades del Estimador VI

- Podemos reescribir la varianza de $\widehat{\beta}_1^{\text{vi}}$ de la siguiente manera:

$$\sigma^2_{\widehat{\beta}_1^{\text{vi}}} = \frac{1}{n}\frac{\sigma_q^2}{\left[\sigma_{ZX}\right]^2}$$

donde $q_i \equiv (Z_i - \mu_Z)\,u_i$, $\sigma_q^2 \equiv Var(q_i)$, y $\sigma_{ZX} \equiv Cov\,(Z_i, X_i)$.

- La varianza anterior puede estimarse de manera consistente de la siguiente manera:

$$\widehat{\sigma}^2_{\widehat{\beta}_1^{\text{vi}}} = \frac{1}{n}\frac{\widehat{\sigma}_q^2}{\left[\widehat{\sigma}_{ZX}\right]^2}$$

donde

$$\widehat{\sigma}_q^2 = \tfrac{1}{n-2}\textstyle\sum_{i=1}^n (Z_i - \overline{Z})^2 \widehat{u}_i^2$$

$$\widehat{\sigma}_{ZX} = s_{ZX} = \tfrac{1}{n-1}\textstyle\sum_{i=1}^n (X_i - \overline{X})(Z_i - \overline{Z})$$

- Luego:

$$SE(\widehat{\beta}_1^{\text{vi}}) = \sqrt{\widehat{\sigma}^2_{\widehat{\beta}_1^{\text{vi}}}}$$

Mínimos Cuadrados en dos Etapas

- La derivación del estimador IV puede llevarse a cabo de una forma alternativa (e interesante).
- El nuevo procedimiento requiere la estimación por MCO de dos regresiones. Por esta razón este método se denomina **Mínimos Cuadrados en Dos Etapas** (**MC2E**).
- ¿Cómo funciona MC2E?
- Supongamos que

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

  con $Cov\left(X_i, u_i\right) \neq 0$ (de manera que MCO S1 no se cumple).
- El estimador MC2E se obtiene en dos etapas.
  1. La primera etapa descompone a X_i en dos partes:
     - Una parte problemática, que correlaciona con u_i.
     - Una parte no problemática, que no correlaciona con u_i.
  2. La segunda etapa usa la parte no problemática de X_i para etimar β_1.

Mínimos Cuadrados en dos Etapas

- **Primera Etapa**
  - Para aislar la parte de X_i que no está correlacionada con u_i, estimamos la siguiente regresión por MCO:

$$X_i = \pi_0 + \pi_1 Z_i + v_i$$

  - Obtenemos:

$$\widehat{\pi}_1 = \frac{\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})}{\sum_{i=1}^{n}(Z_i - \overline{Z})^2} = \frac{s_{ZX}}{s_Z^2}$$

$$\widehat{\pi}_0 = \overline{X} - \widehat{\pi}_1 \overline{Z}$$

  - Computamos los valores de predicción

$$\widehat{X}_i = \widehat{\pi}_0 + \widehat{\pi}_1 Z_i$$

y los residuos

$$\widehat{v}_i = X_i - \widehat{X}_i$$

Mínimos Cuadrados en dos Etapas

- **Primera Etapa** (cont.)
  - A partir de la expresión anterior obtenemos
$$X_i = \widehat{X}_i + \widehat{v}_i$$
  - Hemos descompuesto a X_i en dos partes, $\widehat{X}_i$ y $\widehat{v}_i$.
  - Sabemos por las propiedades algebraicas de MCO que $\widehat{X}_i$ y $\widehat{v}_i$ son ortogonales:
$$Cov(\widehat{X}_i, \widehat{v}_i) = 0$$
  - Además:
$Cov(\widehat{X}_i, u_i) = Cov(\widehat{\pi}_0 + \widehat{\pi}_1 Z_i, u_i) = \widehat{\pi}_1 Cov(Z_i, u_i) = \widehat{\pi}_1 \times 0$
(por la exogeneidad del instrumento) $\Rightarrow$
$$Cov(\widehat{X}_i, u_i) = 0$$
    - $\widehat{X}_i$ es la parte no problemática de X_i: es la parte de X_i que no correlaciona con el error u_i.
    - El otro componente, $\widehat{v}_i$, es problemático ya que $Cov(\widehat{v}_i, u_i) = Cov(X_i - \widehat{X}_i, u_i) = Cov(X_i, u_i) - Cov(\widehat{X}_i, u_i) = Cov(X_i, u_i) \neq 0$.

Mínimos Cuadrados en dos Etapas

- **Segunda Etapa**
  - Ahora estimamos β_1 (y β_0) corriendo la siguiente regresión por MCO:
  $$Y_i = \beta_0 + \beta_1 \widehat{X}_i + e_i$$
  donde la variable explicativa es $\widehat{X}_i$ en lugar de X_i.
  - El resultado es el estimador de mínimos cuadrados en dos etapas (MC2E):
  $$\widehat{\beta}_1^{\text{MC2E}} = \frac{\sum_{i=1}^{n}(\widehat{X}_i - \overline{\widehat{X}})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(\widehat{X}_i - \overline{\widehat{X}})^2} = \frac{s_{\widehat{X}Y}}{s^2_{\widehat{X}}}$$
  - ¿Por qué funciona este procedimiento?
  - Si sustituimos $X_i = \widehat{X}_i + \widehat{v}_i$ en $Y_i = \beta_0 + \beta_1 X_i + u_i$ obtenemos:
  $$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + u_i \\ &= \beta_0 + \beta_1(\widehat{X}_i + \widehat{v}_i) + u_i \\ &= \beta_0 + \beta_1 \widehat{X}_i + (\beta_1 \widehat{v}_i + u_i) \end{aligned}$$

Mínimos Cuadrados en dos Etapas

- **Segunda Etapa** (cont.)
  - Luego:
  $$Y_i = \beta_0 + \beta_1 \widehat{X}_i + e_i$$
  donde
  $$e_i \equiv \beta_1 \widehat{v}_i + u_i$$
  - Recordemos que $Cov(\widehat{X}_i, \widehat{v}_i) = 0$ y $Cov(\widehat{X}_i, u_i) = 0$. Luego: $Cov(\widehat{X}_i, e_i) = Cov(\widehat{X}_i, \beta_1 \widehat{v}_i + u_i) =$ $\beta_1 Cov(\widehat{X}_i, \widehat{v}_i) + Cov(\widehat{X}_i, u_i) = \beta_1 \times 0 + 0 \Rightarrow$
  $$Cov(\widehat{X}_i, e_i) = 0$$
  - Vemos que el nuevo término de error (e_i) no está correlacionado con el nuevo regresor ($\widehat{X}_i$).
    - Recuérdese que $Cov(\widehat{X}_i, e_i) = 0$ es una condición más débil que $\mathbb{E}(e_i \mid \widehat{X}_i) = 0$. La covarianza nula es suficiente para obtener consistencia, pero no alcanza para obtener insesgamiento.
  - A continuación demostraremos que $\widehat{\beta}_1^{\text{MC2E}} = \widehat{\beta}_1^{\text{vi}}$.

Mínimos Cuadrados en dos Etapas

- Tenemos:

$$
\begin{aligned}
\widehat{\beta}_1^{\text{MC2E}} &= \frac{s_{\widehat{X}Y}}{s^2_{\widehat{X}}} \\
&= \frac{\widehat{\pi}_1 s_{ZY}}{\left(\widehat{\pi}_1\right)^2 s^2_Z} && \text{ya que } \widehat{X}_i = \widehat{\pi}_0 + \widehat{\pi}_1 Z_i \\
&= \frac{s_{ZY}}{\widehat{\pi}_1 s^2_Z} = \frac{s_{ZY}}{\frac{s_{ZX}}{s^2_Z} s^2_Z} && \text{ya que } \widehat{\pi}_1 = \frac{s_{ZX}}{s^2_Z} \\
&= \frac{s_{ZY}}{s_{ZX}} \\
&= \widehat{\beta}_1^{\text{vi}}
\end{aligned}
$$

- Como $\widehat{\beta}_1^{\text{MC2E}} = \widehat{\beta}_1^{\text{vi}}$, concluimos que $\widehat{\beta}_1^{\text{MC2E}}$ es consistente para β_1 y asintóticamente normal, con

$$\widehat{\beta}_1^{\text{MC2E}} \overset{a}{\sim} N\left(\beta_1, \sigma^2_{\widehat{\beta}_1^{\text{MC2E}}}\right) \quad \text{y} \quad \sigma^2_{\widehat{\beta}_1^{\text{MC2E}}} = \frac{1}{n}\frac{Var[(Z_i-\mu_Z)u_i]}{[Cov(Z_i,X_i)]^2}$$

- El error estándar es

$$SE(\widehat{\beta}_1^{\text{MC2E}}) = \sqrt{\widehat{\sigma}^2_{\widehat{\beta}_1^{\text{MC2E}}}} \quad \text{con} \quad \widehat{\sigma}^2_{\widehat{\beta}_1^{\text{MC2E}}} = \frac{1}{n}\frac{\widehat{\sigma}_q^2}{\left[\widehat{\sigma}_{ZX}\right]^2}$$

donde

$$\widehat{\sigma}_q^2 = \frac{1}{n-2}\sum_{i=1}^n (Z_i - \overline{Z})^2 \widehat{u}_i^2$$

$$\widehat{\sigma}_{ZX} = s_{ZX} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})(Z_i - \overline{Z})$$

  - Los residuos son $\widehat{u}_i = Y_i - \widehat{Y}_i$, donde los valores de predicción se calculan usando el verdadero regresor X (no $\widehat{X}$): $\widehat{Y}_i = \widehat{\beta}_0^{\text{MC2E}} + \widehat{\beta}_1^{\text{MC2E}} X_i$.

Mínimos Cuadrados en dos Etapas

- Importante

  - Los errores estándar que se obtienen al estimar la segunda etapa por OLS son incorrectos pues no tienen en cuenta que el regresor $\widehat{X}_i$ es una estimación (que proviene de la primera etapa).

  - En la práctica se usa una función especializada que computa el estimador MC2E con los errores estándar correctos.
    - En R lo podemos hacer con **ivreg()**, que viene incluida en el paquete **AER**.

  - Como siempre, usamos errores estándar robustos a la presencia de heterocedasticidad.
    - En R los podemos obtener usando **coeftest**, con las opciones **vcov = vcovHC** y **type = "HC1"**.

Causalidad Simultánea

- Hemos visto que VI nos permite corregir los problemas generados por $\mathbb{E}(u_i \mid X_i) \neq 0$, independientemente de la fuente de la endogeneidad.
- Una de las fuentes de endogeneidad es la **causalidad simultánea**.
- Como su nombre lo indica, la causalidad simultánea ocurre cuando X afecta a Y pero a la vez Y afecta a X.
- Supongamos que queremos estimar el efecto causal del tamaño del curso (STR) sobre el las notas de los alumnos (TS). El modelo propuesto es

$$TS_i = \beta_0 + \beta_1 STR_i + u_i$$

  donde se espera $\beta_1 < 0$.
- Ahora imaginemos que los directores de escuela reaccionan a las bajas notas contratando más docentes, reduciendo de esta manera el tamaño promedio de los cursos:

$$STR_i = \alpha_0 + \alpha_1 TS_i + v_i$$

Causalidad Simultánea

- Esto es un problema, ya que ahora habrá correlación entre STR_i y u_i.
- Veamos por qué.
- Supongamos que uno de los factores omitidos no directamente correlacionado con STR provoca una caída en las notas:

$$\downarrow TS_i = \beta_0 + \beta_1 STR_i + \downarrow u_i$$

- El directorio del colegio reacciona a este shock reduciendo el tamaño del curso:

$$\downarrow STR_i = \alpha_0 + \alpha_1 \downarrow TS_i + v_i \quad \text{(recordemos que } \alpha_1 > 0\text{)}$$

- Pero entonces

$$\uparrow\downarrow TS_i = \beta_0 + \beta_1 \downarrow STR_i + \downarrow u_i \quad \text{(recordemos que } \beta_1 < 0\text{)}$$

- Luego, habrá una correlación positiva entre STR_i y u_i : $Corr(STR_i, u_i) > 0$. Como consecuencia, $\mathbb{E}(u_i \mid STR_i) \neq 0$, y MCO S1 no se cumple.

Causalidad Simultánea

- Como $\mathbb{E}(u_i \mid STR_i) \neq 0$, el estimador MCO de β_1 será inconsistente (y sesgado).
- De hecho, en nuestra clase sobre SVO vimos que

$$\widehat{\beta}_1 \xrightarrow{p} \beta_1 + \frac{\sigma_u}{\sigma_X}\rho_{Xu}$$

- En nuestro caso particular:

$$\widehat{\beta}_1 \xrightarrow{p} \beta_1 + \frac{\sigma_u}{\sigma_{STR}}\rho_{STRu}$$

- Con $\beta_1 < 0$ y $\rho_{STRu} > 0$ la fórmula anterior implica que MCO tenderá a subestimar el impacto del tamaño del curso sobre las notas ($\widehat{\beta}_1$ será menos negativo de lo que debería, e incluso podría ser positivo).
  - Es posible obtener esta misma conclusión observando las flechas que aparecen en la última ecuación de la transparencia anterior.
- Si se consigue un buen instrumento para STR es posible resolver este problema estimando β_1 con $\widehat{\beta}_1^{\text{vi}}$ $(= \widehat{\beta}_1^{\text{MC2E}})$.

Causalidad Simultánea

- Un ejemplo clásico de causalidad simultánea ocurre cuando se quiere estimar una función de demanda (o de oferta).
- Supongamos que la demanda de cierto producto (manteca, por ejemplo) es

$$Q_i = \beta_0 + \beta_1 P_i + u_i \qquad \beta_1 < 0$$

  donde Q_i es la cantidad demandada, P_i es el precio y u_i captura determinantes de la demanda no observados, como el ingreso y las preferencias.
- ¿Por qué no usar MCO para estimar esta función de demanda?

Causalidad Simultánea

- Porque hay otra función, la curva de oferta, que relaciona la cantidad y el precio:

$$Q_i = \alpha_0 + \alpha_1 P_i + v_i \qquad \alpha_1 > 0$$

  donde v_i captura determinantes de la oferta no observados (como los precios de los insumos).
- Como el precio y la cantidad de equilibrio se determinan conjuntamente, tendremos un problema de causalidad simultánea. Mostraremos por qué, de dos maneras distintas.
- Comenzamos invirtiendo la función de oferta:

$$P_i = \gamma_0 + \gamma_1 Q_i + \eta_i \qquad \gamma_1 > 0$$

  donde $\gamma_0 \equiv -\frac{\alpha_0}{\alpha_1}$, $\gamma_1 \equiv \frac{1}{\alpha_1} > 0$ y $\eta_i \equiv -\frac{v_i}{\alpha_1}$.

Causalidad Simultánea

- Ahora supongamos que se se produce un shock no observable ($\uparrow u_i$) que aumenta la cantidad demandada para cada precio (*e.g.*, un aumento del ingreso o un cambio en las preferencias). Luego:

$$\uparrow Q_i = \beta_0 + \beta_1 P_i + \uparrow u_i$$

- Pero la función (inversa) de oferta implica:

$$\uparrow P_i = \gamma_0 + \gamma_1 \uparrow Q_i + \eta_i \quad \text{(recordemos que } \gamma_1 > 0)$$

- Volviendo a la función de demanda:

$$\downarrow\uparrow Q_i = \beta_0 + \beta_1 \uparrow P_i + \uparrow u_i \quad \text{(recordemos que } \beta_1 < 0)$$

- Luego, P_i y u_i están positivamente correlacionados: $Corr(P_i, u_i) > 0$. Entonces $\mathbb{E}(u_i \mid P_i) \neq 0$, y MCO S1 no se cumple.

Causalidad Simultánea

- Recordemos que
$$\widehat{\beta}_1 \xrightarrow{p} \beta_1 + \frac{\sigma_u}{\sigma_P}\rho_{Pu}$$
- Con $\beta_1 < 0$ y $\rho_{Pu} > 0$ la fórmula anterior implica que MCO tenderá a subestimar el efecto del precio sobre la cantidad demandada ($\widehat{\beta}_1$ será menos negativo de lo que debería, y hasta podría ser positivo).
    - La misma conclusión se obtiene si se analizan con cuidado las flechas que aparecen en la última ecuación de la transparencia anterior.
- Con un buen instrumento para P podríamos estimar β_1 de manera consistente usando VI.
- ¿Pero cómo encontramos un buen instrumento?
- Antes de contestar esta pregunta, analicemos el problema de causalidad simultánea desde otra perspectiva.

Causalidad Simultánea

- Consideremos el sistema de ecuaciones formado por las funciones de oferta y demanda:

$$\begin{aligned} \text{Demanda} &: \quad Q_i = \beta_0 + \beta_1 P_i + u_i \qquad \beta_1 < 0 \\ \text{Oferta} &: \quad Q_i = \alpha_0 + \alpha_1 P_i + v_i \qquad \alpha_1 > 0 \end{aligned}$$

- Resolviendo el sistema obtenemos:

$$P_i = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{u_i - v_i}{\alpha_1 - \beta_1}$$

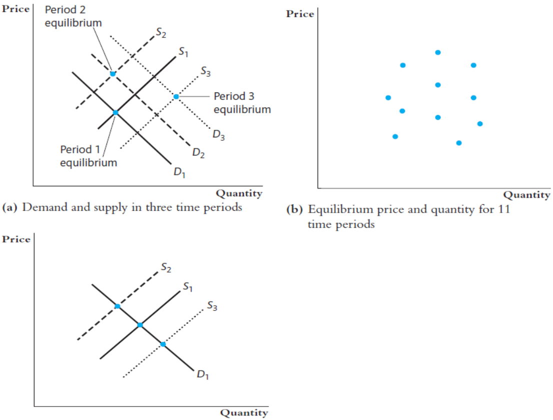
$$Q_i = \frac{\beta_0 \alpha_1 - \beta_1 \alpha_0}{\alpha_1 - \beta_1} + \frac{\alpha_1 u_i - \beta_1 v_i}{\alpha_1 - \beta_1}$$

donde $\alpha_1 - \beta_1 > 0$.

- De la solución anterior obtenemos $Corr(P_i, u_i) > 0$.
    - También se cumple $Corr(P_i, v_i) < 0$, $Corr(Q_i, u_i) > 0$ y $Corr(Q_i, v_i) > 0$.
- Luego, no podemos usar MCO para estimar la función de demanda. Y si queremos usar VI, necesitamos un instrumento para P.

Causalidad Simultánea

- El siguiente gráfico ayuda a entender el problema que presenta la determinación simultánea del precio y la cantidad de equilibrio:



(a) Demand and supply in three time periods

(b) Equilibrium price and quantity for 11 time periods

(c) Equilibrium price and quantity when only the supply curve shifts

Causalidad Simultánea

- El gráfico anterior sugiere una manera de encontrar un instrumento para P que nos permita identificar la demanda: necesitamos una variable Z que mueva la curva de oferta pero no la de demanda.
  - El corrimiento de la curva de oferta inducido por Z generará un cambio en el precio de equilibrio (a lo largo de la curva de demanda). Luego, $Cov\left(Z_i, P_i\right) \neq 0$ (Z es relevante).
  - Si Z mueve la oferta pero no la demanda, y Z no cambia en respuesta a los factores que corren la demanda, tendremos $Cov\left(Z_i, u_i\right) = 0$ (Z es exógeno), precisamente porque los factores que mueven la curva de demanda son parte de u_i.
- En nuestro ejemplo para el mercado de manteca podríamos elegir $Z =$ *cantidad de lluvia en las regiones de producción lechera*.
  - $Cov\left(Z_i, P_i\right) \neq 0$ es plausible: poca lluvia $\Rightarrow$ menos pasturas $\Rightarrow$ menos producción de leche $\Rightarrow$ menor oferta de manteca $\Rightarrow$ mayores precios.
  - $Cov\left(Z_i, u_i\right) = 0$ es plausible: las condiciones meteorológicas en las regiones de producción lechera no deberían afectar la demanda de manteca (y, la lluvia no responde a la demanda).

Causalidad Simultánea

- Formalicemos esta idea.
- El nuevo sistema de ecuaciones es

$$\begin{aligned} \text{Demanda} &: \quad Q_i = \beta_0 + \beta_1 P_i + u_i & \beta_1 < 0 \\ \text{Oferta} &: \quad Q_i = \alpha_0 + \alpha_1 P_i + \alpha_2 Z_i + v_i & \alpha_1 > 0 \end{aligned}$$

  donde Z_i aparece solamente en la función de oferta, y $\alpha_2 > 0$ si más lluvia aumenta la oferta.
- Las funciones de oferta y demanda se denominan **ecuaciones estructurales** ya que se pueden obtener usando la teoría económica y los coeficientes tienen una interpretación causal.
- Q_i y P_i son variables **endógenas** (son determinadas por el sistema de ecuaciones) mientras que Z_i es una variable **exógena** (se determina fuera del sistema).

Causalidad Simultánea

- Resolviendo el sistema anterior obtenemos el precio y la cantidad de equilibrio:

$$P_i = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} - \frac{\alpha_2}{\alpha_1 - \beta_1} Z_i + \frac{u_i - v_i}{\alpha_1 - \beta_1}$$

$$Q_i = \frac{\beta_0 \alpha_1 - \beta_1 \alpha_0}{\alpha_1 - \beta_1} - \frac{\beta_1 \alpha_2}{\alpha_1 - \beta_1} Z_i + \frac{\alpha_1 u_i - \beta_1 v_i}{\alpha_1 - \beta_1}$$

- Las ecuaciones anteriores conforman la **forma reducida** del sistema. Expresan a las variables endógenas (P y Q) como funciones de la/s variable/s exógena/s (Z) y de los shocks no observables (u y v).
- En notación más compacta:

$$\begin{aligned} P_i &= \pi_0 + \pi_1 Z_i + \varepsilon_i \\ Q_i &= \theta_0 + \theta_1 Z_i + \xi_i \end{aligned}$$

Causalidad Simultánea

- Vimos que la forma reducida para P es

$$P_i = \pi_0 + \pi_1 Z_i + \varepsilon_i$$

  donde $\pi_1 \equiv -\frac{\alpha_2}{\alpha_1 - \beta_1}$.
- La expresión anterior muestra que, si $\pi_1 \neq 0$, P_i está correlacionada con $Z_i : Cov(Z_i, P_i) \neq 0$.
  - Nótese que $\pi_1 \neq 0$ requiere $\alpha_2 \neq 0$. Esto significa que Z_i tiene algún efecto sobre la curva de oferta (como hemos asumido).
- Luego, Z_i es un instrumento relevante para P_i.
- Como Z_i no es parte de la ecuación de demanda y tampoco responde a los factores que mueven la demanda, se cumple $Cov(Z_i, u_i) = 0$. Luego, Z_i es un instrumento exógeno para P_i.
- Concluimos entonces que Z_i es un instrumento válido para P_i.

Causalidad Simultánea

- Nótese, además, que

$$P_i = \pi_0 + \pi_1 Z_i + \varepsilon_i$$

  es exactamente la regresión que tendríamos que estimar en la primera etapa de MC2E.

- Una vez que estimamos la ecuación anterior por MCO obtenemos:

$$\widehat{P}_i = \widehat{\pi}_0 + \widehat{\pi}_1 Z_i$$

- Luego, en la segunda etapa, usamos MCO para estimar

$$Q_i = \beta_0 + \beta_1 \widehat{P}_i + e_i$$

  y obtenemos $\widehat{\beta}_1^{\text{MC2E}}$.

  - Recordemos nuestro comentario previo sobre los errores estándar.

Ejemplo: Demanda de Cigarrillos

- El debate político sobre los riesgos de fumar suele referirse a la necesidad de la intervención estatal (externalidades).
- Una opción de política es la introducción de un impuesto que desincentive a los fumadores (a través de un aumento del precio).
- ¿Pero cuál debería ser el nivel del impuesto?
- La respuesta depende de la elasticidad de la demanda de cigarrillos.
  - ¿En qué porcentaje caerá la cantidad demandada por cada 1 % de aumento del precio?

Ejemplo: Demanda de Cigarrillos

- Queremos estimar la siguiente función de demanda de cigarrillos:

$$\ln Q_i = \beta_0 + \beta_1 \ln P_i + u_i$$

donde β_1 es la elasticidad de la demanda.

- Tenemos datos sobre el consumo anual de cigarrilos y el precio promedio por paquete (incluyendo impuestos) para los 48 estados continentales de E.E.U.U. en 1995.
  - Q_i es el número de paquetes de cigarrillos por habitante vendidos en cada estado ($media = 96$).
  - P_i es el precio promedio por paquete incluyendo impuestos.
- La variable instrumental propuesta es

$$Z_i = SalesTax_i$$

donde $SalesTax_i$ es el impuesto general a las ventas (medido en centavos de dólar por paquete), en el estado i.

Ejemplo: Demanda de Cigarrillos

- ¿Es $SalesTax_i$ un instrumento relevante: $Cov\left(SalesTax_i, P_i\right) \neq 0$?
  - Razonable: el impuesto a las ventas afecta el precio de venta.
    - El impuesto traslada la curva de oferta.
- ¿Es $SalesTax_i$ un instrumento exógeno: $Cov\left(SalesTax_i, u_i\right) = 0$?
  - ¿Está $SalesTax$ correlacionado con Q solamente a través de P?
    - ¿El impuesto a las ventas afecta la cantidad demandada solamente a través del precio?
  - Tal vez, pero los impuestos pueden estar afectados por las mismas condiciones locales que afectan a la demanda (actitudes frente al fumar, lobby de la industria del tabaco, etc.).
    - Podría ocurrir que algunos de los factores no observables (u) que trasladan la curva de demanda induzcan a las autoridades a cambiar los impuestos, generando $Cov\left(SalesTax_i, u_i\right) \neq 0$ (la oferta se corre cuando se corre la demanda).
  - Sin embargo, esto es probablemente más problemático para el impuesto específico sobre los cigarrillos que para el impuesto general a las ventas.

- Comenzamos estimando la función de demanda por MCO, ignorando la endogeneidad de P:

```r
# Comenzamos estimando la función de demanda por MCO
# sin tener en cuenta la endogeneidad de P
cig_s0 <- lm(log(packs) ~ log(rprice), data = c1995)
coeftest(cig_s0, vcov = vcovHC, type = "HC1")
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 10.33892    0.93482 11.0598 1.502e-14 ***
## log(rprice) -1.21306    0.19459 -6.2339 1.290e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Luego:

$$\widehat{\ln Q_i} = \underset{(0{,}93)}{10{,}34} - \underset{(0{,}19)}{1{,}21}\ \ln P_i$$

- La elasticidad estimada es alta.

- Ahora estimamos la elasticidad usando MC2E.
  - Usamos MCO para estimar cada etapa por separado:
- Primera etapa ($\ln P_i = \pi_0 + \pi_1 SalesTax_i + \varepsilon_i$):

```
# Realizamos la primera etapa de la regresión
cig_s1 <- lm(log(rprice) ~ salestax, data = c1995)
coeftest(cig_s1, vcov = vcovHC, type = "HC1")
```

```
## t test of coefficients:
##
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) 4.6165463  0.0289177 159.6444 < 2.2e-16 ***
## salestax    0.0307289  0.0048354   6.3549 8.489e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Luego: $\widehat{\ln P_i} = \underset{(0,03)}{4{,}62} + \underset{(0,005)}{0{,}031}\, SalesTax_i$

```
# Almacenamos los valores predichos
lcigp_pred <- cig_s1$fitted.values
```

- Segunda etapa ($\ln Q_i = \beta_0 + \beta_1 \widehat{\ln P}_i + e_i$):

```r
# Corremos la regresión de la segunda etapa
cig_s2 <- lm(log(c1995$packs) ~ lcigp_pred)
coeftest(cig_s2, vcov = vcovHC)
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  9.71988    1.70304  5.7074 7.932e-07 ***
## lcigp_pred  -1.08359    0.35563 -3.0469  0.003822 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Luego:

$$\widehat{\ln Q_i} = \underset{(1{,}70)}{9{,}72} - \underset{(0{,}36)}{1{,}08} \ln P_i$$

- La elasticidad estimada por MC2E es menor que la de MCO.
    - Recordemos que los errores estándar anteriores son incorrectos.
    - Recordemos también que los valores de predicción se computan usando $\ln P_i$ a pesar de que el regresor es $\widehat{\ln P}_i$.

- Para obtener los errores estándar correctos usamos la función ivreg():

```r
# Realizamos la estimación de MC2E usando 'ivreg()'
cig_ivreg <- ivreg(log(packs) ~ log(rprice) | salestax,data=c1995)
coeftest(cig_ivreg, vcov = vcovHC, type = "HC1")
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  9.71988    1.52832  6.3598 8.346e-08 ***
## log(rprice) -1.08359    0.31892 -3.3977  0.001411 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Luego:

$$\widehat{\ln Q_i} = \underset{(1{,}53)}{9{,}72} - \underset{(0{,}32)}{1{,}08} \ln P_i$$

- En este caso particular no hay una gran diferencia entre los errores estándar correctos y los obtenidos usando MCO, pero en otros casos las diferencias pueden ser importantes.

Ejercitación I (20 minutos)

En este ejercicio, utilizaremos R para estimar el efecto de la fertilidad sobre la oferta laboral de las mujeres. El conjunto de datos **fertility.csv** contiene información sobre mujeres casadas de 21 a 35 años con dos o más hijos. Las variables que usaremos son:

- weeksworked: las semanas trabajadas por la mujer en 1979
- morekids: variable ficticia igual a 1 si la mujer tuvo más de 2 hijos
- twoboys: variable ficticia igual a 1 si los dos primeros hijos son varones
- twogirls: variable ficticia igual a 1 si los dos primeros hijos son niñas
- edad: edad de la mujer en el censo de 1980

Con estos datos, nos interesa responder la siguiente pregunta: ¿cuánto cae la oferta laboral de una mujer cuando tiene un hijo adicional? Por tanto, la regresión que queremos estimar es:

$$\text{weeksworked}_i = \beta_0 + \beta_1 \text{morekids}_i + \beta_2 \text{age}_i + u_i$$

Usando R, realice lo siguiente:

(cont...)

Ejercitación I (20 minutos)

(cont...)

1. Estime la primera etapa del modelo de la siguiente forma:

$$\text{morekids}_i = \pi_0 + \pi_1 \text{twoboys}_i + \pi_2 \text{twogilrs}_i + \pi_3 age_i + \varepsilon_i$$

   - Obtenga los errores estándar robustos
   - Interprete los resultados
   - Obtenga los valores predichos

2. Estime la segunda etapa del modelo:

$$\text{weeksworked}_i = \beta_0 + \beta_1 \widehat{\text{morekids}}_i + \beta_2 \text{age}_i + u_i$$

   - Obtenga los errores estándar robustos
   - Interprete los resultados

3. Realice la regresión VI pero utilizando el comando **ivreg()** de R. Verifique que los coeficientes que obtuvo son los mismos que obtuvo anteriormente, pero que los errores estándar son diferentes; ¿Por qué?

4. En función de la regresión con **ivreg()**:
   - ¿Cuántos instrumentos tenemos? ¿Cuáles son?
   - ¿Cuántos regresores endógenos tenemos? ¿Cuáles son?

Apéndice I

- Vamos a demostrar que el estimador VI de β_1 es consistente pero sesgado.
- Sabemos que

$$\widehat{\beta}_1^{\text{vi}} = \frac{\sum_{i=1}^{n}(Z_i - \overline{Z})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})}$$

- Usando $Y_i = \beta_0 + \beta_1 X_i + u_i$, $\overline{Y} = \beta_0 + \beta_1 \overline{X} + \overline{u}$ obtenemos:

$$\begin{aligned}
\widehat{\beta}_1^{\text{vi}} &= \frac{\sum_{i=1}^{n}(Z_i - \overline{Z})(\beta_0 + \beta_1 X_i + u_i - \beta_0 - \beta_1 \overline{X} - \overline{u})}{\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})} \\
&= \frac{\sum_{i=1}^{n}(Z_i - \overline{Z})[\beta_1 (X_i - \overline{X}) + (u_i - \overline{u})]}{\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})} \\
&= \beta_1 \frac{\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})}{\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})} + \frac{\sum_{i=1}^{n}(Z_i - \overline{Z})(u_i - \overline{u})}{\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})}
\end{aligned}$$

Apéndice I

- Luego:

$$\widehat{\beta}_1^{\text{vi}} = \beta_1 + \frac{\sum_{i=1}^{n}(Z_i - \overline{Z})(u_i - \overline{u})}{\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})}$$

- Para demostrar la consistencia de $\widehat{\beta}_1^{\text{vi}}$ reescribimos la expresión anterior de la siguiente manera:

$$\widehat{\beta}_1^{\text{vi}} = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z})(u_i - \overline{u})}{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})}$$

- Pero sabemos que

$$\frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z})(u_i - \overline{u}) \xrightarrow{p} Cov(Z_i, u_i)$$

$$\frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X}) \xrightarrow{p} Cov(Z_i, X_i)$$

- Luego:

$$\widehat{\beta}_1^{\text{vi}} = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z})(u_i - \overline{u})}{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})} \xrightarrow{p} \beta_1 + \frac{Cov(Z_i, u_i)}{Cov(Z_i, X_i)}$$

Apéndice I

- Pero:

$$\begin{array}{rcll} Cov(Z_i, u_i) & = & 0 & \text{(exogeneidad)} \\ Cov(Z_i, X_i) & \neq & 0 & \text{(relevancia)} \end{array}$$

- Luego:

$$\frac{Cov(Z_i, u_i)}{Cov(Z_i, X_i)} = 0$$

- Luego: $\widehat{\beta}_1^{\text{vi}} \xrightarrow{p} \beta_1 + \frac{Cov(Z_i,u_i)}{Cov(Z_i,X_i)} \Rightarrow$

$$\widehat{\beta}_1^{\text{vi}} \xrightarrow{p} \beta_1$$

- Concluimos que $\widehat{\beta}_1^{\text{vi}}$ es un estimador consistente de β_1.

Apéndice I

- Para demostrar que $\widehat{\beta}_1^{\text{vi}}$ es en general sesgado, notemos que

$$\begin{aligned}
\widehat{\beta}_1^{\text{vi}} &= \beta_1 + \frac{\sum_{i=1}^{n}(Z_i - \overline{Z})(u_i - \overline{u})}{\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})} \\
&= \beta_1 + \frac{\sum_{i=1}^{n}(Z_i - \overline{Z})u_i}{\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})} - \frac{\overline{u}\sum_{i=1}^{n}(Z_i - \overline{Z})}{\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})} \\
&= \beta_1 + \frac{\sum_{i=1}^{n}(Z_i - \overline{Z})u_i}{\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})}
\end{aligned}$$

donde hemos usado $\sum_{i=1}^{n}(Z_i - \overline{Z}) = 0$.

- Luego:

$$\mathbb{E}\left(\widehat{\beta}_1^{\text{vi}}\right) = \mathbb{E}\left(\beta_1 + \frac{\sum_{i=1}^{n}(Z_i - \overline{Z})u_i}{\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})}\right)$$

Apéndice I

- Luego:

$$
\begin{aligned}
\mathbb{E}\left(\widehat{\beta}_1^{\text{vi}}\right) &= \beta_1 + \mathbb{E}\left(\frac{\sum_{i=1}^{n}(Z_i-\overline{Z})u_i}{\sum_{i=1}^{n}(Z_i-\overline{Z})(X_i-\overline{X})}\right) \\
&= \beta_1 + \mathbb{E}\left(\mathbb{E}\left[\left.\frac{\sum_{i=1}^{n}(Z_i-\overline{Z})u_i}{\sum_{i=1}^{n}(Z_i-\overline{Z})(X_i-\overline{X})}\right| X_1,\ldots X_n, Z_1,\ldots,Z_n\right]\right) \\
&= \beta_1 + \mathbb{E}\left(\frac{\sum_{i=1}^{n}(Z_i-\overline{Z})\mathbb{E}(u_i|X_1,\ldots X_n,Z_1,\ldots,Z_n)}{\sum_{i=1}^{n}(Z_i-\overline{Z})(X_i-\overline{X})}\right) \\
&= \beta_1 + \mathbb{E}\left(\frac{\sum_{i=1}^{n}(Z_i-\overline{Z})\mathbb{E}(u_i|X_i,Z_i)}{\sum_{i=1}^{n}(Z_i-\overline{Z})(X_i-\overline{X})}\right)
\end{aligned}
$$

donde el último paso utiliza la hipótesis de independencia (que se sigue de VI S2).

Apéndice I

- Como $\mathbb{E}(u_i \mid X_i) \neq 0$ (por hipótesis), se sigue que $\mathbb{E}(u_i \mid X_i, Z_i) \neq 0$. Entonces, en general:

$$\mathbb{E}\left(\frac{\sum_{i=1}^{n}(Z_i-\overline{Z})\mathbb{E}(u_i|X_i,Z_i)}{\sum_{i=1}^{n}(Z_i-\overline{Z})(X_i-\overline{X})}\right) \neq 0$$

- Luego

$$\begin{aligned}\mathbb{E}\left(\widehat{\beta}_1^{\text{vi}}\right) &= \beta_1 + \mathbb{E}\left(\frac{\sum_{i=1}^{n}(Z_i-\overline{Z})\mathbb{E}(u_i|X_i,Z_i)}{\sum_{i=1}^{n}(Z_i-\overline{Z})(X_i-\overline{X})}\right) \\ &\neq \beta_1\end{aligned}$$

- Concluimos que $\widehat{\beta}_1^{\text{vi}}$ es un estimador sesgado de β_1.