Economía Aplicada

Variables Instrumentales

Departamento de Economía
Universidad Carlos III de Madrid

Basado en Stock y Watson (cap.12), Wooldridge (cap. 15) y Angrist y Pischke (cap. 4)

Modelos de Regresión Simple y Múltiple

- Comparemos los modelos simple y múltiple a través de un ejemplo:

Modelo de Regresión Múltiple (Modelo Largo)

- $w_i = \beta_0 + \beta_1 educ_i + \beta_2 IQ_i + u_i$
  Se verifican los supuestos del modelo de regresión lineal:
- $C(educ, u) = C(IQ, u) = 0$

Modelo de Regresión Simple (Modelo Corto)

- $w_i = \gamma_0 + \gamma_1 educ_i + v_i$

- ¿Existe alguna relación entre γ_1 y β_1?

¿Existe alguna relación entre γ_1 y β_1?

Usando el modelo largo (imponiendo $C(educ, u) = 0$):

$$
\begin{aligned}
C(educ, w) &= C(educ, \beta_0 + \beta_1 \, educ + \beta_2 IQ + u) \\
&= \beta_1 \, V(educ) + \beta_2 \, C(educ, IQ)
\end{aligned}
$$

Comparando ambos modelos:

$$
\gamma_1 = \frac{C(educ, w)}{V(educ)} = \beta_1 + \beta_2 \frac{C(educ, IQ)}{V(educ)}
$$

Sesgo de variable omitida

- Encontramos que imponiendo $C(educ, u) = 0$:

$$\gamma_1 = \beta_1 + \beta_2 \frac{C(educ, IQ)}{V(educ)}$$

- Notar que $\frac{C(educ, IQ)}{V(educ)}$ es la pendiente de una regresión de IQ en $educ$.
- Esta fórmula define el llamado Sesgo de Variable Omitida: $\gamma_1 - \beta_1$
- Es fácil ver que no hay Sesgo de Variable Omitida ($\gamma_1 = \beta_1$) si se verifica al menos una de las siguientes situaciones:
  - inteligencia no es relevante en la ecuación de salarios: $\beta_2 = 0$
  - educación no está correlacionada con inteligencia: $C(educ, IQ) = 0$

¿Es $\hat{\gamma}_1$ un estimador consistente del parámetro de interés?

Suponemos que el parámetro de interés es β_1 y utilizamos el modelo corto

$$\hat{\gamma}_1 = \frac{\hat{C}(educ, wages)}{\hat{V}(educ)} = \frac{\hat{C}(educ, \beta_0 + \beta_1 educ + \beta_2 IQ + u)}{\hat{V}(educ)}$$

$$= \beta_1 + \beta_2 \frac{\hat{C}(educ, IQ)}{\hat{V}(educ)}$$

$$\Rightarrow plim(\hat{\gamma}_1) = \beta_1 + \beta_2 \frac{C(educ, IQ)}{V(educ)}$$

- $plim(\hat{\gamma}_1) = \beta_1$ ($\hat{\gamma}_1$ es consistente) si
    - la inteligencia no es relevante: $\beta_2 = 0$
    - la educación no está correlacionada con la inteligencia: $C(educ, IQ) = 0$
- Se puede mostrar que $V(\hat{\gamma}_1) \leq V(\hat{\beta}_1)$

Controles no correlacionados

- Si *educ* e *IQ* **no están correlacionados** tenemos condiciones de primer orden sencillas para los estimadores MCO:

$$\hat{C}(educ, wages) = \hat{\beta}_1 \hat{V}(educ)$$
$$\hat{C}(IQ, wages) = \hat{\beta}_2 \hat{V}(IQ)$$

- Por lo tanto:

$$\hat{\beta}_1 = \frac{\hat{C}(educ, wages)}{\hat{V}(educ)}$$
$$\hat{\beta}_2 = \frac{\hat{C}(IQ, wages)}{\hat{V}(IQ)}$$

- el estimador MCO es el mismo que el que se obtiene de un modelo corto:

$$\hat{\beta}_1 = \frac{\hat{C}(educ, wages)}{\hat{V}(educ)} = \hat{\gamma}_1$$

Controles correlacionados

- Con **controles correlacionados**, en el modelo largo tenemos CPO más complicadas:

$$\hat{C}(educ, wages) = \hat{\beta}_1 \hat{V}(educ) + \hat{\beta}_2 \hat{C}(educ, IQ)$$
$$\hat{C}(IQ, wages) = \hat{\beta}_1 \hat{C}(IQ, educ) + \hat{\beta}_2 \hat{V}(IQ)$$

- dividiendo la primera condición por $\hat{V}(educ)$:

$$\frac{\hat{C}(educ,wages)}{\hat{V}(educ)} = \hat{\beta}_1 + \hat{\beta}_2 \frac{\hat{C}(educ,IQ)}{\hat{V}(educ)}$$

- el estimador MCO en el modelo corto $\hat{\gamma}_1 = \frac{\hat{C}(educ,wages)}{\hat{V}(educ)}$ se escribe:

$$\hat{\gamma}_1 = \hat{\beta}_1 + \hat{\beta}_2 \frac{\hat{C}(educ,IQ)}{\hat{V}(educ)}$$

Controles correlacionados (cont.)

- Volvemos al Sesgo de Variable Omitida:

$$\hat{\gamma}_1 - \hat{\beta}_1 = \hat{\beta}_2 \frac{\hat{C}(educ, IQ)}{\hat{V}(educ)}$$

- El estimador MCO en el modelo corto ($\hat{\gamma}_1$) captura dos efectos sobre los salarios:

1. los efectos de cambios independientes en la variable *educ*: $\hat{\beta}_1$
2. los efectos de cambios en *IQ* que son a la vez cambios en *educ*:

$$\hat{\beta}_2 \frac{\hat{C}(educ, IQ)}{\hat{V}(educ)}$$

- donde $\frac{\hat{C}(educ, IQ)}{\hat{V}(educ)}$ captura el cambio en *IQ* ante un cambio en *educ*

Independencia en Media Condicional

- Por lo tanto, si estimamos el modelo corto cuando el modelo largo es el correcto, no estamos identificando el efecto que queremos (en este ejemplo el impacto de la educación en el salario). ¿Por qué?
- Porque el supuesto de Independencia en Media Condicional no se cumple:

$$E(v|educ) \neq 0$$

- Cuando el supuesto de Independencia en Media Condicional no se cumple, se dice que existe un problema de endogeneidad.
- Si, por cualquier razón, X_j está correlacionada con u, decimos que X_j es una variable endógena.

Primera Aplicación

- El método de variables instrumentales suele usarse para resolver problemas de endogeneidad relacionados con variables omitidas como el que vimos antes.
- Sin embargo, las primeras aplicaciones se vinculan a intentos de estimar elasticidades de la demanda y la oferta de productos agrícolas.
- En concreto Philip Wright (1928) usó la idea de lo que serían variables instrumentales para estimar la elasticidad de la demanda de linaza utilizando una ecuación de demanda simple: $ln(Q_i) = \beta_0 + \beta_1 ln(P_i) + u_i$, donde Q es cantidad y P precio.
- Problema: los precios y cantidades de equilibrio se determinan simultáneamente en el cruce de las curvas de oferta y demanda.

Primera Aplicación (cont.)

- Por lo tanto una estimación MCO de cantidades en precios no logra identificar ni la curva de demanda ni la de oferta.
- La solución propuesta por Wright fue encontrar dos tipos de factores: "(A) que afecten las condiciones de la demanda sin afectar los costos o que (B) afecten los costos sin afectar las condiciones de la demanda".
- Factores del tipo (A) ayudan a identificar la curva de oferta, factores del tipo (B) ayudan a identificar la curva de demanda.
- Wright propuso varios de esos factores: por ejemplo el precio de bienes sustitutos como factor que afecta la demanda pero no la oferta, y factores climáticos como factores que afectan la oferta sin afectar la demanda.

VI: Introducción 1/2

Supongamos que queremos estimar el siguiente modelo:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \text{ , donde } C(X_i, u_i) \neq 0$$

- Si $C(X_i, u_i) \neq 0$, entonces X es una variable endógena y el estimador MCO es inconsistente.
- La estimación con Variables Instrumentales (VI) utiliza una variable adicional (Z) para aislar la parte de X que no está correlacionada con u.
- Para que sea útil Z pedimos que cumpla dos condiciones.

VI: Introducción 2/2

- Las dos condiciones que pedimos a un buen instrumento son:
  - Z no correlacionada con el error: $C(Z_i, u_i) = 0$. No afecta directamente a la variable de interés. Las variables no correlacionadas con el error se denominan variables exógenas. Por ello esta condición se denomina **exogeneidad del instrumento**.
  - Z correlacionada (en términos de correlación parcial si hay otros regresores) con X (la variable endógena) : $C(Z, X) \neq 0$. Esta condición se denomina **relevancia del instrumento**.
- Si Z es relevante, su variación está relacionada con la variación de X. Si además Z es exógena, la parte de la variación de X capturada por Z es exógena. La única razón para que exista una relación entre Y y Z es por la relevancia de Z.
- Si se cumplen estas condiciones, Z nos permite obtener estimadores consistentes incluso en caso de endogeneidad.

Ejemplo Retorno a la Educación 1/3

Ejemplo: ecuación de salarios

$$salario_i = \beta_0 + \beta_1 educ_i + u_i$$

- ¿Es razonable asumir que $C(educ_i, u_i) = 0$?
- Podemos argumentar que por ejemplo la habilidad de la persona es una variable omitida en el modelo. Si el nivel educativo de una persona está correlacionado con su habilidad, el supuesto de independencia en media condicional no se verifica.
- Un buen instrumento necesita estar correlacionado con la variable endógena *educ* pero no con el error, que en nuestro ejemplo simple podría ser la habilidad. ¿Alguna idea?

Ejemplo 2/3

- Algunos ejemplos de VI para educación utilizados en la literatura: educación de los padres, número de hermanos, distancia a la universidad, fecha de nacimiento.
- Por ejemplo, Card(1995) utilizó datos de salario y educación de una muestra de hombres de 1976 para estimar el retorno a la educación. Estimó una ecuación similar a la anterior con controles adicionales: experiencia, raza, región.
- Utilizó una variable artificial que toma el valor uno si la persona creció cerca de una universidad como instrumento para el nivel educativo alcanzado.

Ejemplo 3/3

- Relevancia: aquéllos que crecieron más cerca de una universidad tienen una probabilidad mayor de asistir. ¿algún argumento en contra?
- Exogeneidad: la distancia al centro educativo no debería estar correlacionada con la habilidad de los individuos, ni afectar directamente a los salarios. ¿algún argumento en contra?
- Card encuentra que la estimación por VI del coeficiente de la educación es casi dos veces mayor que la estimación por MCO (13,2% vs. 7,5%).
- Pero también encuentra que el error estándar de la estimación por VI es 18 veces más grande. El intervalo de confianza al 95% va de 0,024 a 0,239. El precio a pagar por conseguir estimadores consistentes.

Validez del instrumento: Exogeneidad

¿Cómo saber si tenemos un instrumento válido? Exogeneidad del instrumento:

- En el caso de una variable endógena y un instrumento no es posible contrastar si se verifica la condición de exogeneidad: $C(Z,u)=0$.
- En el ejemplo de Card debemos argumentar que la distancia al centro de estudios no afecta al salario a través de otro mecanismo. ¿Qué sucedería si la distancia estuviera correlacionada con el ingreso familiar y el ingreso familiar tuviese un efecto directo en los salarios?
- En el ejemplo de Wright debemos argumentar que los factores climáticos no tienen una influencia directa sobre la demanda del bien.

Validez del instrumento: Relevancia

¿Cómo saber si tenemos un instrumento válido? Relevancia del instrumento:

- La condición de relevancia: $C(X_i|Z_i) \neq 0$ es verificable ya que ambas variables son observadas por el investigador:
    - Estimamos una regresión de X en Z:

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

      y contrastamos la siguiente hipótesis: $H_0: \ \pi_1 = 0$.
    - Si rechazamos H_0, encontramos evidencia de que X y Z están correlacionadas y por tanto el instrumento es relevante.
    - Si no rechazamos H_0, se dice que el instrumento es débil, lo cual representa un problema que veremos más adelante en más detalle.

Estimador de mínimos cuadrados en dos etapas: MC2E

Si el instrumento Z cumple las dos condiciones, es posible obtener estimadores consistentes para β_2 mediante el llamado estimador de mínimos cuadrados en dos etapas: MC2E.

Las dos etapas:

- La primera etapa regresa X en Z: $X_i = \pi_0 + \pi_1 Z_i + v_i$
  La idea es utlizar el componente de X que puede predecirse con Z: $\pi_0 + \pi_1 Z_i$ que, al ser Z exógena, estará incorrelacionado con u. Como los parámetros π_0 y π_1 son desconocidos, en esta primera etapa se obtienen estimaciones MCO de dichos parámetros y se computa $\hat{X}$.
- La segunda etapa es la regresión de Y en $\hat{X}$ utilizando MCO. Los estimadores de la regresión de la segunda etapa son los estimadores MC2E.

Estimador MC2E (cont.)

- En la práctica las dos etapas se realizan automáticamente en cualquier paquete econométrico. Si se realizan las dos etapas por separado, los errores estándar en la segunda etapa deben ajustarse ya que se están usando variables estimadas en una etapa anterior.
- Fórmula: en el caso de un regresor y un instrumento es muy sencilla:

$$\hat{\beta}_1^{MC2E} = \frac{s_{ZY}}{s_{ZX}},$$

  donde s representa la covarianza muestral entre dos variables.
- Es posible demostrar que el estimador MC2E es consistente y que en muestras grandes distribuye normalmente.

Estimador MC2E: consistencia

Veamos la consistencia utilizando la ecuación original $Y_i = \beta_0 + \beta_1 X_i + u_i$ y aplicando propiedades de las covarianzas:

$$C(Z,Y) = \beta_1 C(Z,X) + C(Z,u)$$

Bajo el supuesto de exogeneidad del instrumento: $C(Z,u) = 0$ y entonces $\beta_1 = C(Z,Y)/C(Z,X)$. Dado que la covarianza muestral es un estimador consistente de la covarianza poblacional se puede mostrar que:

$$\hat{\beta}_1^{MC2E} = \frac{s_{ZY}}{s_{ZX}} \xrightarrow{p} \frac{C(Z,Y)}{C(Z,X)} = \beta_1$$

MC2E vs MCO

- MC2E:

$$\hat{\beta}_1^{MC2E} = \frac{s_{ZY}}{s_{ZX}}$$

Su sesgo se puede obtener sin imponer el supuesto de exogeneidad:

$$\hat{\beta}_1^{MC2E} \xrightarrow{p} \frac{C(Z,Y)}{C(Z,X)} = \beta_1 + \frac{C(Z,u)}{C(Z,X)}.$$

El sesgo depende de las dos condiciones: exogeneidad y relevancia.

- MCO:

$$\hat{\beta}_1^{MCO} = \frac{s_{XY}}{s_X^2}$$

Su sesgo se puede obtener de manera similar:

$$\hat{\beta}_1^{MCO} \xrightarrow{p} \frac{C(X,Y)}{V(X)} = \beta_1 + \frac{C(X,u)}{V(X)}.$$

El sesgo depende de la exogeneidad de la variable X.

Modelo general

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$$
$$+ \beta_{k+1} W_{1i} + \ldots + \beta_{k+r} W_{ri} + u_i$$

- Es posible tener varios regresores endógenos (las X, potencialmente correlacionadas con u) y varios regresores exógenos (las W, no correlacionadas con u).
- Para aplicar MC2E se requieren al menos tantos instrumentos (que llamaremos $Z_1, Z_2, \ldots, Z_m$) como variables endógenas.
- Terminología: se dice que los coeficientes están exactamente identificados si el número de instrumentos es igual al número de regresores endógenos y que están sobreidentificados si tenemos más instrumentos que regresores endógenos.

MC2E en el modelo general: varios instrumentos

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 W_{1i} + \ldots + \beta_{1+r} W_{ri} + u_i$$

- Con más de un instrumento para X_1 tendríamos más de un estimador de VI, pero ninguno sería eficiente: el mejor instrumento es una combinación lineal de todos los instrumentos.
- Primera etapa (relaciona X_1 con los m instrumentos y los r regresores exógenos):
  $X_{1i} = \pi_0 + \pi_1 Z_{1i} + \pi_2 Z_{2i} + \ldots + \pi_m Z_{mi} + \pi_{m+1} W_{1i} + \ldots + \pi_{m+r} W_{ri} + v_i$
- Segunda etapa: Y_i se regresa sobre $\hat{X}_i$ y los regresores exógenos en la ecuación original $(W_{1i}, \ldots, W_{ri})$ utilizando MCO.
- Relevancia: al menos una Z útil para predecir X, dado W.
- Exogeneidad: cada instrumento Z debe estar incorrelacionado con u.

MC2E en el modelo general: múltiples regresores endógenos

- El procedimiento de MC2E es similar al ya explicado, pero en este caso cada regresor endógeno requiere su propia regresión de primera etapa. La ecuación para cada X contiene los mismos regresores: todos los instrumentos y todos los regresores exógenos de la ecuación original.
- En la segunda etapa se estima la ecuación original por MCO sustituyendo las X_j por sus valores estimados respectivos $(\hat{X}_j)$ obtenidos en la primera etapa.
- Como dijimos, en la práctica las dos etapas de MC2E se realizan automáticamente en cualquier paquete econométrico, como veremos en el caso de `gretl`. Y así se obtienen los errores estándar correctos.

Contrastando endogeneidad: Test de Hausman

- Si no existe endogeneidad en el modelo original, tanto MCO como VI son consistentes, pero MCO es más eficiente.
- Si existe endogeneidad, solamente VI es consistente.
- Por lo tanto es importante realizar un contraste de endogeneidad para evitar usar VI cuando no es necesario.
- Se puede utilizar el contraste de Hausman de endogeneidad (H_0 : Exogeneidad).

Contraste de endogeneidad 1/2

Dado el siguiente modelo muy simplificado:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 W_{1i} + u_i$$

- Si contamos con una variable exógena adicional (Z_1), para contrastar si X_1 es endógena se puede aplicar un procedimiento de dos etapas:
- **Primera etapa:** regresar X_1 en todas las exógenas (en nuestro ejemplo Z_1 y W_1) y computar los residuos $\hat{v}$:

$$X_1 = \pi_0 + \pi_1 Z_1 + \pi_2 W_1 + v$$

- En caso de exogeneidad de X_1, dado que Z_1 y W_1 (suponemos que) no están correlacionadas con u, los residuos $\hat{v}$ tampoco deberían estarlo.

Contraste de endogeneidad 2/2

- **Segunda etapa:** estimar por MCO el modelo original agregando $\hat{v}$ a la ecuación:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 W_{1i} + \alpha \hat{v}_i + \varepsilon_i$$

- Se contrasta la hipótesis nula de que X_1 es exógena. Bajo exogeneidad, el coeficiente de $\hat{v}$ no debería ser significativo: $H_0) \alpha = 0$.
- Si rechazamos H_0, encontramos evidencia en contra de la exogeneidad de X_1, y por tanto en contra de estimar el modelo por MCO.
- Notar que es necesario contar con un instrumento exógeno para llevar a cabo el contraste.

Validez de los instrumentos: contraste de relevancia

- En el caso de un regresor endógeno y varios instrumentos: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 W_{1i} + u_i$, con m variables adicionales exógenas: $Z_1, ..., Z_m$.
- La relevancia se analiza en la ecuación de la primera etapa: $X_{1i} = \pi_0 + \pi_1 Z_{1i} + ... + \pi_m Z_{mi} + \pi_{m+1} W_{1i} + v_i$
- Se contrasta la hipótesis nula de que los coeficientes de las Z son conjuntamente iguales a cero: $H0)\pi_1 = ... = \pi_m = 0$. El estadístico F correspondiente da una medida de la información incluida en los instrumentos.
- Los instrumentos son "débiles" si explican una pequeña proporción de la variación de X_1 (regla práctica sencilla: F menor que 10).

Instrumentos débiles

- Si los instrumentos son débiles, los estimadores MC2E estarán sesgados y la inferencia estadística usando MC2E no será correcta (los estadísticos t y los intervalos de confianza serán poco fiables). El problema persiste incluso en muestras grandes.
- ¿Qué hacer? ¡Conseguir mejores instrumentos! Nada fácil...
- Si se tienen varios instrumentos, se suele proceder descartando el más débil y con los restantes volver a realizar el análisis hasta encontrar un subconjunto de instrumentos que se puedan considerar relevantes.

Validez de los instrumentos: exogeneidad

- No es posible contrastar la exogeneidad de los instrumentos si los coeficientes están exactamente identificados.
- Si disponemos de más instrumentos que variables endógenas, es posible realizar un contraste que se llama de sobreidentificación (Sargan).
- Es un contraste sobre la exogeneidad de los instrumentos "adicionales" bajo el supuesto de que existen suficientes instrumentos válidos para identificar los coeficientes de interés.

Contraste de sobreidentificación

1. Primera etapa: se estima el modelo original via MC2E y se obtienen los residuos ($\hat{u}^{MC2E}$).
2. Segunda etapa: se regresan esos residuos, usando MCO, sobre todas las variables exógenas del modelo:
   $\hat{u}^{MC2E} = \delta_0 + \delta_1 Z_{1i} + ... + \delta_m Z_{mi} + \delta_{m+1} W_{1i} + ... + \delta_{m+r} W_{ri} + v_i$

- El estadístico para el contraste es nR^2, que bajo la hipótesis nula de que los instrumentos son exógenos verifica:

$$LM = nR^2 \rightarrow \chi^2_q$$

donde q es el número de instrumentos adicionales (el grado de sobreidentificación).

Aplicación

- Se quiere estimar la elasticidad de la demanda de cigarrillos usando datos de consumo anual en EEUU con la siguiente ecuación donde Q_i es el número de paquetes vendidos per cápita y P_i el precio real promedio por paquete incluidos todos los impuestos en el estado i.

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

- ¿Sería correcto utilizar MCO para estimar la elasticidad?
- Si queremos aplicar MC2E necesitamos al menos un instrumento. Usaremos el impuesto sobre los cigarrillos medido en dólares por paquete en cada estado: $ImpV_i$.

Aplicación (cont.)

- Requisitos para que *ImpV* sea válido:
  1. Relevante: correlacionada con P
     Usando datos de 1995 se obtienen los siguientes resultados para la primera etapa (archivo cig_ch10.gdt):

     $$\widehat{ln(P_i)} = \underset{(0.0289)}{4.6165} + \underset{(0.0048)}{0.0307}\, ImpV \quad T = 48 \quad R^2 = 0.4710$$

     El coeficiente de *ImpV* es positivo y significativamente distinto de 0: a mayor nivel impositivo, mayores precios después de impuestos. La variación en el impuesto a las ventas de cigarrillos explica el 47% de la varianza de los precios de los cigarrillos entre estados.
  2. Exógena: no correlacionada con u: no es posible realizar un contraste formal. Para que el impuesto sea una variable exógena, debe afectar a la demanda de cigarrillos solamente a traés del precio, lo que parece razonable.

Aplicación (cont.)

- Estimación MC2E utilizando *ImpV* como instrumento para el precio con errores estándar robustos:

$$\widehat{ln(Q_i)} = \underset{(1.5283)}{9.7199} - \underset{(0.3189)}{1.0836}\, ln(P_i)$$

  La elasticidad estimada implica que un aumento en el precio de un 1% reduce en promedio el consumo en un 1.08%

- Un posible problema es que existan variables omitidas correlacionadas con el impuesto sobre las ventas: lo que implicaría que *ImpV* no sería exógena (¿y entonces?).
- Por ejemplo, los estados con mayores ingresos podrían tener menores impuestos a las ventas y a su vez tener mayores niveles de consumo.

Aplicación (cont.)

- Para resolver este problema incorporamos la variable ingreso (que suponemos es exógena):

$$ln(Q_i) = \beta_0 + \beta_1 ln(P_i) + \beta_2 ln(Ing_i) + u_i$$

- Estimación MC2E utilizando *ImpV* como instrumento para el precio con errores estándar robustos:

$$\widehat{ln(Q_i)} = \underset{(1.2594)}{9.4307} - \underset{(0.3723)}{1.1434}\, ln(P_i) + \underset{(0.3117)}{0.2145}\, ln(Ing_i)$$

- Para esta estimación se utilizó un único instrumento: la elasticidad de la demanda está exactamente identificada.
- Se puede usar como instrumento adicional los impuestos específicos a los cigarrillos: con dos instrumentos la elasticidad de la demanda está sobreidentificada.

Aplicación (cont.)

- Estimación MC2E utilizando *ImpV* e *ImpCig* como instrumentos para el precio con errores estándar robustos:

$$\widehat{ln(Q_i)} = \underset{(0.9592)}{9.8950} - \underset{(0.2496)}{1.2774}\, ln(P_i) + \underset{(0.2539)}{0.2804}\, ln(Ing_i)$$

- Comparar el error estándar con uno y dos instrumentos: ¿qué nos dice esta comparación?.
- ¿Son confiables estas estimaciones? Depende de la validez de los instrumentos.

Aplicación (cont.)

- Relevancia de los instrumentos: regresión de la primera etapa

$$ln(P_i) = \pi_0 + \pi_1 ImpV_i + \pi_2 ImpC_i + \pi_3 ln(Ing_i) + u_i$$

El estadístico F correspondiente a $H0)\pi_1 = \pi_2 = 0$ es 209.676, lo que lleva a rechazar la hipótesis nula de que los instrumentos son débiles.

- Exogeneidad: dado que hay dos instrumentos y una variable endógena es posible realizar un contraste de sobreidentificación. El contraste de Sargan da un estadístico de 0.33 que dada la distribución χ^2_1, tiene un p-valor asociado de 0.5641. Por tanto no se rechaza la hipótesis nula de que ambos instrumentos son exógenos.

Examples of Studies That Use Instrumental Variables to Analyze Data From Natural and Randomized Experiments

Outcome Variable	*Endogenous Variable*	*Source of Instrumental Variable(s)*	*Reference*
		1. Natural Experiments	
Labor supply	Disability insurance replacement rates	Region and time variation in benefit rules	Gruber (2000)
Labor supply	Fertility	Sibling-Sex composition	Angrist and Evans (1998)
Education, Labor supply	Out-of-wedlock fertility	Occurrence of twin births	Bronars and Grogger (1994)
Wages	Unemployment insurance tax rate	State laws	Anderson and Meyer (2000)
Earnings	Years of schooling	Region and time variation in school construction	Duflo (2001)
Earnings	Years of schooling	Proximity to college	Card (1995)
Earnings	Years of schooling	Quarter of birth	Angrist and Krueger (1991)
Earnings	Veteran status	Cohort dummies	Imbens and van der Klaauw (1995)
Earnings	Veteran status	Draft lottery number	Angrist (1990)
Achievement test scores	Class size	Discontinuities in class size due to maximum class-size rule	Angrist and Lavy (1999)
College enrollment	Financial aid	Discontinuities in financial aid formula	van der Klaauw (1996)
Health	Heart attack surgery	Proximity to cardiac care centers	McClellan, McNeil and Newhouse (1994)
Crime	Police	Electoral cycles	Levitt (1997)
Employment and Earnings	Length of prison sentence	Randomly assigned federal judges	Kling (1999)
Birth weight	Maternal smoking	State cigarette taxes	Evans and Ringel (1999)
		2. Randomized Experiments	
Earnings	Participation in job training program	Random assignment of admission to training program	Bloom et al. (1997)
Earnings	Participation in Job Corps program	Random assignment of admission to training program	Burghardt et al. (2001)
Achievement test scores	Enrollment in private school	Randomly selected offer of school voucher	Howell et al. (2000)
Achievement test scores	Class size	Random assignment to a small or normal-size class	Krueger (1999)
Achievement test scores	Hours of study	Random mailing of test preparation materials	Powers and Swinton (1984)
Birth weight	Maternal smoking	Random assignment of free smoker's counseling	Permutt and Hebel (1989)

Usando el archivo mroz.gdt, que tiene información para mujeres que participan en el mercado de trabajo, se estima una ecuación de salarios:

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exp + \beta_3 exp^2 + \varepsilon$$

1. Analizar si *educ* es una variable exógena (utilizar educación del esposo y de los padres como variables exógenas).
2. Evaluar si la educación del esposo y de los padres podrían ser instrumentos adecuados para *educ*.
3. Estimar el efecto de la educación en el salario utilizando el método más apropiado: MCO o MC2E.