

## Economía Aplicada Variables Instrumentales

1. Queremos analizar el efecto de fumar sobre el peso de los niños al nacer. Usamos para eso el archivo `bwght.gdt`. Un modelo simple sería:

$$\log(bwght) = \beta_0 + \beta_1 packs + u$$

donde *bwght* es el peso del niño al nacer y *packs* el número diario de paquetes de cigarrillos fumado por la madre durante el embarazo.

- (a) Estime este modelo e interprete  $\hat{\beta}_1$ . ¿Cree que  $E(u|packs) = 0$  es un supuesto razonable? Mencione dos factores que puedan estar incluidos en  $u$
- (b) Un modelo más interesante debería incluir más variables:

$$\log(bwght) = \beta_0 + \beta_1 packs + \beta_2 male + \beta_3 parity + \beta_4 lfaminc + u$$

donde *male* es una variable binaria que toma el valor 1 cuando el bebé es varón, *parity* representa el orden de nacimiento y *lfaminc* representa el logaritmo del ingreso familiar (medido en miles de dólares).

Estime este nuevo modelo e interprete  $\hat{\beta}_1$ . ¿Cambia esta estimación respecto a la anterior? ¿Cree que  $E(u|packs, male, parity, lfaminc) = 0$  es un supuesto razonable?

- (c) Si sospechamos que *packs* puede estar correlacionada con  $u$  (por ejemplo podemos pensar que el consumo de cigarrillos está correlacionado con otros factores que afectan a la salud de la madre o con un buen cuidado prenatal), usar una estimación por variables instrumentales (VI) es probablemente una buena idea. Una posible VI para *packs* es el precio medio de los cigarrillos en el lugar de residencia, *cigprice*. Si los cigarrillos son un bien de consumo habitual, la teoría económica básica sugiere que *packs* y *cigprice* están correlacionadas de forma negativa, por lo que se puede utilizar *cigprice* como VI para *packs*.

Estime el modelo anterior usando *cigprice* como VI para *packs*. ¿Cuál es su conclusión sobre el efecto de fumar en el peso de los bebés?

- (d) ¿Es *cigprice* un buen instrumento? ¿Qué condiciones debe cumplir *cigprice* para considerarlo un instrumento válido? Evalúe esas condiciones, si es posible, y decida si *cigprice* es un buen instrumento.
- (e) ¿Cuál estimación prefiere, MCO o VI? Explicar.

2. Usaremos el archivo `wage2.gdt` para investigar los retornos a la educación en una ecuación salarial típica:

$$lwage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 cap + \epsilon \quad (1)$$

donde  $E(\epsilon | educ, exper, tenure, cap) = 0$

$lwage$  = logaritmo del salario mensual (en miles de dólares);  $educ$  = años de educación completos;  $exper$  = experiencia laboral (en años);  $tenure$  = antigüedad en el trabajo (en años);  $cap$  = habilidad o capacidad.

Los efectos esperados de estas cuatro variables son positivos. Sabemos que habilidad y educación están positivamente correlacionados:  $C(educ, cap) > 0$ , y que experiencia ( $exper$ ) y antigüedad ( $tenure$ ) no están correlacionados con educación ( $educ$ ) ni con habilidad ( $cap$ ). Sin embargo, la habilidad no es observable, con lo cual el modelo que podemos estimar es:

$$lwage = \gamma_0 + \gamma_1 educ + \gamma_2 exper + \gamma_3 tenure + \epsilon \quad (2)$$

Existe la posibilidad de utilizar una medida imperfecta de habilidad. En la muestra disponemos del coeficiente intelectual de los individuos ( $IQ$ ).

También poseemos información sobre las siguientes variables (no correlacionadas con  $\epsilon$ ):  $meduc$  (educación de la madre) y  $KWW$  (puntuación en un test de conocimiento general).

- (a) Aunque nos interesaría estimar el modelo (1), dado que la habilidad no es observable, estimaremos via MCO el modelo (2). ¿Qué esperaríamos de la estimación del retorno a la educación en ese caso? Estime el modelo de la ecuación (2) e interprete los coeficientes.
  - (b) Asuma que  $IQ = cap$ , de modo que el coeficiente intelectual ( $IQ$ ) es exactamente igual a la habilidad y estime la ecuación (1). Compute el efecto de un año adicional de educación sobre los salarios.
  - (c) Evalúe si  $IQ$  es una variable endógena. Explique su respuesta.
  - (d) Evalúe si  $meduc$  y  $KWW$  son instrumentos válidos para  $IQ$ . Explique su respuesta.
  - (e) Decida cuál método es más adecuado en este caso, MCO o MC2E. Compute el retorno a la educación con el método más adecuado.
  - (f) Realice las dos etapas de la estimación de MC2E y compare los errores estándar de la segunda etapa con los obtenidos haciendo la estimación automática en `gretl`.
3. (Basado en Stock y Watson, pág. 480) ¿Cuál es el efecto de la fertilidad en la oferta de trabajo? ¿Cae la oferta de trabajo de una mujer cuando tiene un hijo más? En este ejercicio estimaremos este efecto usando datos de mujeres casadas del Censo de 1980 de EEUU. El archivo `fertility.gdt` contiene información de mujeres casadas de entre 21 y 35 años de edad con dos o más niños. La variable `weeksm1` representa las semanas trabajadas por cada mujer en 1979, `morekids` es una variable artificial que toma el valor 1 si la mujer tiene más de dos hijos.

- i) Regrese *weeksm1* sobre la variable *morekids* usando MCO. En promedio, ¿Trabajan menos las mujeres con más de dos hijos que las mujeres con dos hijos? ¿Cuánto menos?
- ii) Explique por qué la regresión anterior parece inapropiada para estimar el efecto causal de la fertilidad (*morekids*) en la oferta de trabajo (*weeksm1*).
- iii) La base de datos contiene la variable *samesex*, que vale 1 si los dos primeros hijos son del mismo sexo (niño-niño o niña-niña) y 0 en caso contrario. ¿Las parejas con dos hijos del mismo sexo tienen mayor probabilidad de tener un tercero? ¿Es una diferencia grande? ¿Es estadísticamente significativa?
- iv) Explique por qué *samesex* es un instrumento válido para *morekids*.
- v) Estime la regresión de *weeksm1* sobre *morekids* usando *samesex* como instrumento. ¿Cuán grande es el efecto de la fertilidad en la oferta de trabajo?
- vi) ¿Cambian los resultados al incluir las variables *black*, *hispan* y *othrace* en la regresión de oferta de trabajo (tratando a estas variables como exógenas)? Explique.