

Hoja de Ejercicios 4

Variables Instrumentales y MC2E. Modelos de ecuaciones simultáneas

ECONOMETRÍA I. UC3M

1. [W 15.1] Considere un modelo simple para estimar el efecto de disponer de un ordenador personal (PC) en la nota media de la universidad (GPA) para los graduados en una universidad pública

$$GPA = \beta_0 + \beta_1 PC + u,$$

donde PC es una variable binaria que indica si se posee un ordenador.

- a) ¿Es plausible que PC esté correlacionada con u ? Justifique la respuesta.
 - b) Explique por qué PC está posiblemente relacionada con la renta de los padres. ¿Significa esto que la renta de los padres es una variable instrumental para PC ? Justifique la respuesta.
 - c) Suponga que, hace cuatro años, la universidad daba becas para comprar ordenadores a aproximadamente la mitad de los nuevos estudiantes, y que los estudiantes que recibían los ordenadores eran seleccionados al azar. Explique detalladamente como usaría dicha información para construir una variable instrumental para PC .
2. [W 15.2] Suponga que quiere estimar el efecto de la asistencia a clase sobre los resultados de los estudios. Un modelo básico es

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + u,$$

donde las variables se definen como habitualmente.

- a) Sea $dist$ la distancia desde el lugar de residencia del estudiante a las aulas. ¿Es razonable suponer que $dist$ no está correlacionado con u ?
- b) Suponiendo que $dist$ y u no están correlacionadas, ¿qué otro supuesto debe satisfacer $dist$ para que sea un instrumento válido para $atndrte$?
- c) Suponga que añadimos la interacción $priGPA \cdot atndrte$,

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA \cdot atndrte + u.$$

Si $atndrte$ está correlacionado con u , entonces, en general, también lo estará $priGPA \cdot atndrte$. ¿Cuál podría ser un buen instrumento para $priGPA \cdot atndrte$? **[Pista:** si $E[u|priGPA, ACT, dist] = 0$, es decir $priGPA, ACT, dist$ son exógenos, entonces cualquier función de $priGPA$ y $dist$ no está correlacionada con u].

3. [W 15.5] Suponga que $\sigma_x = \sigma_u$ en un modelo de regresión

$$y = \beta_0 + \beta_1 x + u,$$

con lo que la varianza poblacional del término de error es la misma que la de x . Suponga que la variable instrumental z , presenta una cierta correlación con u , $Cov(z, u) = 0,1$. Suponga también que z y x tienen una correlación algo mayor, $Cov(x, z) = 0,2$.

- a) ¿Cuál es el sesgo asintótico del estimador de VI que utiliza z como instrumento?
 b) ¿Cuánta correlación debería existir entre x y u para que el estimador MCO tenga más sesgo asintótico que el estimador anterior de VI?

4. [W 15.7] El siguiente es un modelo se quiere utilizar para medir el efecto de un programa de elección de colegio en función de los resultados de un test estandarizado de rendimiento,

$$score = \beta_0 + \beta_1 choice + \beta_2 faminc + u_1,$$

donde $score$ es el resultado del test realizado a nivel estatal, $choice$ es una variable binaria que indica si el estudiante ha cursado el último año en un colegio de su elección, y $faminc$ es la renta familiar. La VI para $choice$ es $grant$, la cantidad en dólares recibida por los estudiantes para pagar la matrícula en los colegios elegidos. La beca recibida varía en función del nivel de ingresos de la familia, que es la razón por la que controlamos por $faminc$ en la ecuación.

- a) Incluso con $faminc$ en la ecuación, ¿por qué $choice$ podría estar correlacionada con u_1 ?
 b) Si dentro de cada tramo de renta, el importe de la beca fuese asignado aleatoriamente, ¿estaría $grant$ correlacionada con u_1 ?
 c) Escriba la ecuación de la forma reducida para $choice$. ¿Qué debe cumplirse para que $grant$ esté correlacionada parcialmente con $choice$?
 d) Escriba la forma reducida para $score$ y explique su utilidad. [Pista: cómo interpretarías el coeficiente de $grant$?]
5. [W 15.8] Se quiere contrastar si las niñas que acuden a un instituto femenino tienen mejores resultados en matemáticas que las niñas que acuden a escuelas mixtas. Dada una muestra aleatoria de alumnas de distintos institutos de un estado de los EE.UU, $score$ es el resultado del examen de matemáticas estandarizado, y $girlhs$ una variable ficticia que indica si la alumna iba a una escuela femenina.

- a) ¿Qué otros factores tendría en cuenta en la ecuación?
 b) Escriba una ecuación relacionando $score$ con $girlhs$ y los factores adicionales que haya indicado en (a).
 c) Suponga que el apoyo y la motivación de los padres son factores no medibles, incluidos por tanto en el término de error en (b). ¿Estarán dichos factores correlacionados con $girlhs$? Justifique su respuesta.
 d) Discuta las condiciones para que una variable que indique el número de escuelas femeninas en un radio de 20 millas del hogar de la alumna sea un instrumento válido para $girlhs$.

6. [W 15.13] Los datos FERTIL2 incluyen, para mujeres de Botswana en 1988, información sobre el número de hijos ($children$), los años de educación ($educ$), la edad (age) y variables sobre el status económico.

- a) Estime por MCO el modelo

$$children = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 age^2 + u$$

e interprete los resultados. En particular, para una edad (age) dada, ¿cuál es el efecto estimado de un año adicional de educación sobre la fertilidad? Si 100 mujeres reciben un año adicional de educación, ¿cuántos niños menos se espera que tenga?

- b) *frsthalf* es una variable binaria igual a 1 si la mujer nació durante los primeros seis meses del año. Suponiendo que *frsthalf* no está correlacionada con el término de error de (a), muestre que es un instrumento apropiado para *educ*.
- c) Estime el modelo (a) usando *frsthalf* como instrumento para *educ*. Compare el efecto estimado de *educ* con el del correspondiente estimador MCO del apartado (a) y contraste la posible endogeneidad de *educ*.
- d) Añada las variables ficticias *electric*, *tv* y *bicycle* al modelo y suponga que son exógenas. Estime la ecuación por MCO y por MC2E y compare los coeficientes estimados de *educ*. Interprete el coeficiente de *tv* y explique por qué la posesión de una televisión tiene un efecto negativo sobre la fertilidad.

7. [W 15.14] Utilice los datos de CARD para este ejercicio.

- a) Considere la ecuación

$$\begin{aligned} \log(wage) = & \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 \\ & + \beta_4 black + \beta_5 smsa + \beta_6 south + u, \end{aligned}$$

Para que el estimador de VI sea consistente, el instrumento para *educ*, *nearc4* (que es una variable binaria que indica si el individuo se crió cerca de una escuela para niños de 4 años) no puede estar correlacionada con *u*. ¿Podría estar *nearc4* correlacionada con algún componente del término de error, como habilidad no observada? Justifique su respuesta.

- b) Para una submuestra de hombres, se dispone de los resultados de un test *IQ*. Regrese *IQ* sobre *nearc4* para comprobar si los resultados medios del *IQ* varían con el hecho de que el hombre se criase cerca de una escuela para niños de 4 años. ¿Cuál es su conclusión?
- c) Regrese ahora *IQ* sobre *nearc4*, *smsa66* y las variables regionales *reg662*, ..., *reg669*. ¿Sigue existiendo relación entre *IQ* y *nearc4* una vez que se ha tenido en cuenta el efecto de las variables regionales? ¿Cómo afecta su respuesta a las conclusiones del apartado (b)?
- d) A partir de los apartados (a) y (b), ¿qué puede concluirse sobre la importancia de controlar por *smsa66* y las variables regionales de 1996 en la ecuación para $\log(wage)$?
8. [W 16.1] Considere un sistema de dos ecuaciones en la forma de “oferta y demanda”, es decir, con la misma variable y_1 (típicamente, cantidad) apareciendo en el lado izquierdo de ambas ecuaciones:

$$\begin{aligned} y_1 &= \alpha_1 y_2 + \beta_1 z_1 + u_1 \\ y_1 &= \alpha_2 y_2 + \beta_2 z_2 + u_2. \end{aligned}$$

- a) Si $\alpha_1 = 0$ ó $\alpha_2 = 0$, explique por qué existe una forma reducida para y_1 . (Una forma reducida expresa y_1 como una función lineal de las variables exógenas y de los errores estructurales). Si $\alpha_1 \neq 0$ y $\alpha_2 = 0$, escriba la forma reducida para y_2 .
- b) Si $\alpha_1 \neq 0$, $\alpha_2 \neq 0$ y $\alpha_1 \neq \alpha_2$, escriba la forma reducida para y_1 . ¿Existe en este caso una forma reducida para y_2 ?
- c) ¿Es posible que se cumpla la condición $\alpha_1 \neq \alpha_2$ en los ejemplos de oferta-demanda? Justifique su respuesta.

9. [W 16.2] La variable *corn* denota el consumo per cápita de maíz en un municipio, *price* su precio en dicho municipio, *income* la renta per cápita del municipio y *rainfall* la lluvia caída durante la última estación de crecimiento del maíz. El siguiente sistema de ecuaciones simultáneas impone la condición de equilibrio de que la oferta iguala la demanda,

$$\begin{aligned} \text{corn} &= \alpha_1 \text{price} + \beta_1 \text{income} + u_1 \\ \text{corn} &= \alpha_2 \text{price} + \beta_2 \text{rainfall} + u_2. \end{aligned}$$

¿Cuál es la ecuación de oferta y cuál es la ecuación de demanda? Explique su respuesta.

10. [W. 16.4] Se supone que los ingresos anuales y el consumo de alcohol se relacionan de acuerdo con el siguiente sistema de ecuaciones simultáneas:

$$\begin{aligned} \log(\text{earnings}) &= \beta_0 + \beta_1 \text{alcohol} + \beta_2 \text{educ} + u_1, \\ \text{alcohol} &= \gamma_0 + \gamma_1 \log(\text{earnings}) + \gamma_2 \text{educ} + \gamma_3 \log(\text{price}) + u_2, \end{aligned}$$

donde *price* es un índice local de los precios del alcohol, que incluye impuestos locales y estatales. Suponiendo que *educ* y *price* son exógenos, y que $\beta_1, \beta_2, \gamma_1, \gamma_2$ y γ_3 son todos diferentes de cero, ¿qué ecuación está identificada? ¿Cómo la estimaría?

11. [W. 16.9] Utilice los datos de SMOKE para este ejercicio.

- a) Se considera el siguiente modelo para estimar los efectos de fumar sobre la renta anual (posiblemente a través de días de trabajo perdidos por enfermedad, o por efectos sobre la productividad):

$$\log(\text{income}) = \beta_0 + \beta_1 \text{cigs} + \beta_2 \text{educ} + \beta_3 \text{age} + \beta_4 \text{age}^2 + u_1$$

donde *cigs* es el número diario de cigarrillos consumidos. ¿Cuál es la interpretación de β_1 ?

- b) Para reflejar el hecho de que el consumo de cigarrillos puede determinarse conjuntamente con la renta, se considera la ecuación de demanda de cigarrillos

$$\begin{aligned} \text{cigs} &= \gamma_0 + \gamma_1 \log(\text{income}) + \gamma_2 \text{educ} + \gamma_3 \text{age} + \gamma_4 \text{age}^2 \\ &+ \gamma_5 \log(\text{cigpric}) + \gamma_6 \text{restaurn} + u_2 \end{aligned}$$

donde *cigpric* es el precio de un paquete de cigarrillos (en centavos), y *restaurn* es una variable binaria igual a uno si la persona vive en un estado con restricciones al consumo de tabaco en los restaurantes. Suponiendo que estas dos últimas variables son exógenas al individuo, ¿qué signos esperarías para γ_5 y γ_6 ?

- c) ¿Bajo qué supuestos estaría identificada la ecuación de renta del apartado (a)?
d) Estime la ecuación de renta por MCO y discuta el valor del estimador de β_1 .
e) Estime la forma reducida para *cigs*, regresando *cigs* sobre todas las variables exógenas. ¿Son $\log(\text{cigpric})$ y *restaurn* significativas en la forma reducida?
f) Estime ahora la ecuación de renta por MC2E. Compare el estimador de β_1 obtenido con el de MCO.
g) ¿Cree que los precios de los cigarrillos y las restricciones a fumar en los restaurantes son realmente exógenas en la ecuación de renta?

12. [W. 16.10] Utilice los datos de MROZ y considere el ejemplo de oferta de trabajo de mujeres que trabajan.

- a) Estime la ecuación de oferta de trabajo

$$\begin{aligned}\log(hours) = & \alpha_1 \log(wage) + \beta_{10} + \beta_{11}educ + \beta_{12}age \\ & + \beta_{13}kidslt6 + \beta_{14}nwifeinc + u_1\end{aligned}$$

por MC2E usando $exper$ y $exper^2$ como instrumentos para $\log(wage)$, y compare el resultado con el obtenido con $hours$ como variable dependiente.

- b) En la ecuación de oferta de trabajo del apartado (a), es posible que $educ$ sea endógena, dado que la habilidad está omitida. Utilice $motheduc$ y $fatheduc$ como instrumentos para $educ$. Ahora hay dos variables explicativas endógenas en esta ecuación.
- c) Contraste las restricciones de sobre-identificación en la estimación MC2E del apartado (b). ¿Pasan los instrumentos el contraste?