Violación de Supuestos

Oscar Gálvez-Soriano

University of Houston
Department of Economics

October 2021

1. Violación del supuesto de ortogonalidad
    - Variables omitidas
    - Error de medición
    - Simultaneidad

Variables omitidas

Considere que el modelo verdadero es:

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \gamma q + v$$

con

$$E\left(v | x_1, x_2, ..., x_k, q\right) = 0$$

Supongamos ahora que estimamos el modelo omitiendo q. Normalmente porque q no se puede medir. El modelo estimado es:

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

donde ahora el error tiene dos términos:

$$u = \gamma q + v$$

Variables omitidas

Sabemos que el término de error v tiene esperanza condicional igual a cero, pero ¿qué podemos decir sobre u?

Supongamos que ahora nuestro modelo es:

$$y = \alpha + \beta x + \gamma q + v$$

con

$$E\left(v|x, q\right) = 0$$

Pero en realidad estimamos el siguiente modelo:

$$y = \alpha + \beta x + u$$

$$plim\ \hat{\beta} = \frac{cov(x, y)}{var(x)}$$

Variables omitidas

Desarrollando el numerador,

$$cov\left(x,y\right)=cov\left(x,\alpha+\beta x+\gamma q+v\right)$$

$$cov\left(x,y\right)=cov(x,\alpha)+cov(x,\beta x)+cov(x,\gamma q)+cov(x,v)$$

$$cov\left(x,y\right)=0+\beta cov(x,x)+\gamma cov(x,q)+0=\beta var(x)+\gamma cov(x,q)$$

$$\Rightarrow\;plim\;\hat{\beta}=\beta+\gamma\frac{cov(x,q)}{var(x)}$$

Variables omitidas

Si $cov(x,q) \neq 0$ o si $\gamma \neq 0$, podemos decir algo acerca de la relación entre $\hat{\beta}$ y β.

- Si tanto la $cov(x,q)$ como γ tienen el mismo signo, entonces el sesgo asintótico será positivo. Es decir, $\hat{\beta}$ estará sesgado hacia arriba.
- Si la $cov(x,q)$ y γ tienen diferentes signos, entonces el sesgo asintótico será negativo. Es decir, $\hat{\beta}$ estará sesgado hacia abajo.

Hagamos un ejemplo en STATA...

Error de medición

Algunas razones para el error de medición:

- Sesgos de información auto-reportada.
- Sesgos de memoria.
- Errores codificando los datos.
- Diferentes interpretaciones de la misma pregunta

Error de medición

Supongamos que el modelo verdadero es:

$$y^* = \alpha + \beta x^* + u$$

pero no observamos y^* o x^* mientras que en su lugar, observamos
$y = y^* + v$
$x = x^* + w$
donde v y w son independientes entre ellos y de u.

Error de medición

Tenemos dos posibles casos:

1. Si nuestro problema es que y^* está medido con error y este error es independiente de x^*, entonces usar MCO está bien dado que los supuestos del modelo aún se cumplen (pero los errores estándar serán más grandes).
2. Si es x^* la que está medida con error (error de medición clásico), tendremos un sesgo de atenuación.

Simultaneidad

Supongamos que el modelo verdadero es:

$$y = \alpha_1 + \beta x + \varepsilon_1$$

$$x = \alpha_2 + \delta y + \gamma z + \varepsilon_2$$

Entonces, el estimador de MCO, $plim\ \hat{\beta} = \frac{cov(x,y)}{var(x)}$, es:

$$cov(x, y) = cov(x, \alpha_1 + \beta x + \varepsilon_1)$$

$$cov(x, y) = cov(x, \alpha_1) + cov(x, \beta x) + cov(x, \varepsilon_1)$$

Simultaneidad

$$cov(x, y) = 0 + \beta cov(x, x) + cov(x, \varepsilon_1) = \beta var(x) + cov(x, \varepsilon_1)$$

$$\Rightarrow \; plim \; \hat{\beta} = \beta + \frac{cov(x, \varepsilon_1)}{var(x)}$$

Pero noten que el segundo término NO desaparece por el problema de endogeneidad.