

Trabajo Práctico 6

Variables omitidas y errores de medición

Contenidos: Omisión de variables relevantes y errores de medición.

1. Omisión de variables relevantes

En el Trabajo Práctico N° 3 se presentaron, estimaron y discutieron las ecuaciones de ingresos o ecuaciones de Mincer. Aquí volvemos a este tema con el objetivo de estudiar las consecuencias de no disponer de cierta información que se considera relevante.

a) Considere que los salarios se determinan de la siguiente forma:

$$\log w = \text{edu} \beta + \text{habilidad} \gamma + Z \varphi + \mu \quad (1)$$

donde $\log w$ (vector $N \times 1$) es el logaritmo del salario mensual, edu son los años de educación formal (vector $N \times 1$) y Z es una matriz con otras variables explicativas ($N \times q$) incluyendo la constante (primer columna de unos). Se supone que:

- El vector $N \times 1$ de términos aleatorios μ satisface los supuestos clásicos
- La variable *habilidad* no está correlacionada con ningún otro regresor del modelo una vez que se controla por *edu*. Es decir, puede escribirse $\text{habilidad} = \delta_1 + \delta_2 \text{edu} + r$ donde r no está correlacionado con ninguna de las variables incluidas en el vector Z .

Debido a la dificultad, o aún imposibilidad, de medir todas las dimensiones que determinan la *habilidad* de un individuo, es usual considerar que la *habilidad* no es observable para el analista o que es “inobservable”.

Suponga que este modelo se estima por MCO omitiendo la variable *habilidad*. Sabemos que el retorno a la educación estimado $\hat{\beta}$ será sesgado para β . ¿Cuál es el sesgo? ¿De qué depende la magnitud y el signo del sesgo?

b) Suponga ahora que entre las variables incluidas en el vector Z están la experiencia laboral, la antigüedad en la empresa, el estado civil, la región de residencia y la raza. De esta manera el modelo (1) toma la siguiente forma:

$$\log w = \varphi_1 + \varphi_2 \exp + \varphi_3 \text{ant} + \varphi_4 \text{casado} + \varphi_5 \text{sur} + \varphi_6 \text{urbano} + \varphi_7 \text{negro} + \beta \text{edu} + \gamma \text{habilidad} + \mu$$

donde las variables se definen como:

<i>exp</i>	Experiencia en el mercado laboral (en años)
<i>ant</i>	Antigüedad en la firma (en años)
<i>casado</i>	=1 si casado, =0 si no es casado
<i>sur</i>	=1 para la región sur, =0 en caso contrario
<i>urbano</i>	=1 si vive en zona urbana, =0 en caso contrario
<i>negro</i>	=1 si raza negra, =0 en caso contrario
<i>habilidad</i>	No observable

Se supone que se satisfacen los supuestos enunciados en el inciso anterior.

Si bien la *habilidad* es inobservable, es factible de ser aproximada por algunas medidas de inteligencia, como por ejemplo el coeficiente intelectual (*CI*). En este sentido decimos que el *CI* es una *proxy* de la variable *habilidad*.

b.1) La base de datos *omisión de habilidad.dta* contiene información sobre todas estas variables para 935 trabajadores varones¹. A partir de estos datos, estimar por MCO el modelo que omita *habilidad*.

b.2) Estimar el modelo que reemplaza *habilidad* por la proxy *CI*.

b.3) Compare los resultados de los dos incisos previos.

b.4) Supusimos que a variable *habilidad* no está correlacionada con ningún otro regresor del modelo una vez que se controla por *edu*. ¿Piensa que la variable proxy *CI* también satisface este supuesto? Argumente su respuesta usando nuevamente los resultados obtenidos en los incisos (b.1) y (b.2).

¹ Estos datos fueron los utilizados en el trabajo de Blackburn y Neumark (1992). "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials", *Quarterly Journal of Economics* 107, 1421-1436.

Parte II

2. Problema empírico: Variables omitidas y error de medición en una ecuación de Mincer

El objetivo de este ejercicio consiste en introducir el análisis de variables omitidas a nivel empírico, es decir, a un ejemplo concreto.

A partir del trabajo de Blackburn and Neumark (1995)², responda las siguientes preguntas:

- a) ¿Cuál es el modelo que intentan estimar los autores?
- b) ¿Qué método de estimación proponen?
- c) ¿Cuál es el problema que encuentran al estimar este modelo? Justifique.
- d) ¿En qué sentido la variable habilidad es relevante para este modelo?
- e) ¿Los coeficientes que salen de la regresión de la ecuación de Mincer sin la inclusión de la habilidad, son confiables? Justifique
- f) ¿Qué coeficiente sesga la no inclusión de la habilidad en la regresión?
- g) ¿Qué solución proponen los autores al problema de variables omitidas?
- h) ¿Dónde se encuentra el error de medición en la ecuación que desean estimar los autores?
- i) ¿Qué problema genera la presencia de este error de medición?
- j) ¿Cuál es la conclusión del paper?
- k) En vistas a lo visto a lo largo de la guía y en el trabajo, ¿Cuales son los caminos a seguir para un investigador que sospecha que su modelo que va a regresar está omitiendo una variable relevante?

² Blackburn and Neumark (1995), "Are OLS estimates of the return to schooling biased downward? Another look". *Review of Economics and Statistics*, vol.77, no.2 (May), pp. 217-230.

Notas computacionales

Para crear una nueva variable igual al logaritmo de la variable original.

```
generate lx = log(x)
```

Para realizar un test t de que un coeficiente es 0

```
test x
```

Para realizar un test t de que un coeficiente es igual a otro

```
test x=y
```

Para realizar un test t de que un coeficiente es igual a un número, por ejemplo 5

```
test x=5
```