COMUNICACIONES

NOTA SOBRE LOS MÉTODOS ESTADÍSTICOS UTILIZADOS POR LA *OCDE* EN EL ESTUDIO COMPARATIVO SOBRE LAS ESTRUCTURAS PROFESIONALES Y EDUCATIVAS Y EL NIVEL DE DESARROLLO ECONÓMICO *

ENRIQUE COSÍO PASCAL
Universidad de París

I. INTRODUCCIÓN

EN ESTOS COMENTARIOS, el trabajo de la OCDE citado como referencia [4] es considerado como una función producción en el cuadro de la teoría general de la producción.

Por consiguiente, se hará un análisis del método estadístico (regresiones simples y múltiples) utilizado por la OCDE para llegar a la conclusión de que los parámetros estimados tienen un sesgo no homogéneo hacia arriba, lo que hace muy difícil su interpretación.

II. ANTECEDENTES SOBRE LA TEORÍA DE LA PRODUCCIÓN

En la teoría de la producción [1] se considera un espacio de bienes donde todas las producciones posibles son representadas por un vector perteneciente a este espacio.

La función producción es una función numérica definida en el espacio de producción. Por ejemplo:

$$g(a_1, a_2, \ldots, a_N) \leq 0 \qquad (1)$$

donde por convención los productos tienen signo positivo y los insumos signo negativo. Evidentemente la elección de los insumos y los productos es puramente convencional.

Supongamos que se ha elegido a_1 como producto y todos los otros a_i como insumos. Entonces, de acuerdo con (1) tendremos:

$$a_1 \leq f(a_2, \ldots, a_N) \qquad (2)$$

* Este trabajo fue presentado al Seminario de la OCDE (Organización de Cooperación y Desarrollo Económicos) sobre los Métodos de Planificación de los Recursos Humanos en los Países en Vía de Desarrollo (Balance y Perspectivas), París, septiembre de 1971. El trabajo previo de la OCDE que se comenta [4], se titula, *Structures professionnelles et educatives et niveaux de développement economique,* París, OCDE, 1970. (Hay edición en inglés).

1 Véase [1], pp. 37-49 y [3], pp. 38-69.

Si el vector $(a_1, a_2, \ldots, a_N)$ es técnicamente eficaz [2] entonces podemos escribir (2) como:

$$a_1 = f(a_2, \ldots, a_N) \tag{3}$$

donde f satisface las hipótesis hechas sobre las funciones producción: [3]

H_a — Hipótesis de aditividad
H_d — Hipótesis de divisibilidad
H_r — Hipótesis sobre los rendimientos (f homogénea de primer grado)
H_c — Hipótesis de convexidad

III. El estudio de la OCDE

El estudio de la OCDE fue hecho probando diferentes relaciones:

$$L_j/L = f(n) \tag{4}$$

donde la variable dependiente es el porciento de una profesión dada j referida al conjunto de la economía o a un sector de actividad económica y n representa una de las variables explicativas seleccionadas para expresar los niveles de desarrollo económico y tecnológico.

$$n = \begin{cases} Y/L \\ \Sigma I/L \\ In \\ En/L \end{cases} \tag{5}$$

donde Y/L es la productividad bruta del trabajo, $\Sigma I/L$ la formación bruta de capital, In un indicador no monetario [4] del desarrollo económico y En/L el consumo de energía por trabajador.

Los sistemas (4) y (5) son utilizados para analizar la estructura profesional.

Para analizar las relaciones entre la profesión y los niveles de educación se utilizan:

$$L_{jk}/L_j = f(n) \tag{6}$$

$$L_{jk}/L = f(n) \tag{7}$$

donde L_{jk} representa el número de personas de la profesión j que poseen el nivel de instrucción k.

En esta perspectiva, se han utilizado además las relaciones:

$$\left.\begin{aligned} & L_{jk}/L_j = f(L_k/L) \\ \text{o bien} \quad & L_{jk}/L = f(L_k/L) \end{aligned}\right\} \tag{8}$$

$$\left.\begin{aligned} \text{y} \quad & L_{jk}/L_j = f(n, L_k/L) \\ \text{o bien} \quad & L_{jk}/L = f(n, L_k/L) \end{aligned}\right\} \tag{9}$$

[2] Véase [3], p. 39.
[3] Véase [3], pp. 45-49.
[4] Véase [4], p. 49.

Los sistemas (8) y (9) tienen por objeto resolver lo que en la publicación de la OCDE es llamado "el problema de identificación".[5]

Para analizar la estructura educativa, la OCDE utiliza las relaciones:

$$L_k/L = f(n) \tag{10}$$

En primer término, se presenta un enfoque desde el punto de vista de la teoría de la producción para compararlo con las funciones utilizadas en el estudio de la OCDE.

Para hacer esto, se tomará la función producción (3) y los sistemas (4) y (5). Las observaciones que se harán son aplicables fácilmente a las ecuaciones (6), (7) y (10). Las diferencias de interpretación que resultan surgen del cuadro de esta exposición, pero no representan un obstáculo para la generalización del razonamiento que será aplicado.

Una función producción donde se han considerado diferentes categorías de mano de obra puede expresarse:

$$Y = f(K, L_1, L_2, \ldots, L_n) \tag{11}$$

donde K es el capital fijo.

Como f satisface H_r, se puede escribir:

$$Y/L = f(K/L, L_1/L, \ldots, L_n/L) \tag{12}$$

Cambiando nuestra definición de producto se puede pasar a:

$$L_1/L = g(Y/L, K/L, L_2/L, \ldots, L_n/L) \tag{13}$$

Para la OCDE la energía *per capita* consumida y el índice no monetario son también factores de la producción y sustituyen el capital fijo por la formación bruta de capital, por lo que:

$$L_1/L = N(Y/L, K/L, L_2/L, \ldots, L_n/L, En/L, In) \tag{14}$$

Para situarse en el contexto del estudio de la OCDE es necesario hacer la hipótesis siguiente:

H_p — Una vez fijada la proporción L_j/L, las otras proporciones L_h/L para $j \neq h$ son constantes.

Aplicando H_p a (14) se puede escribir:

$$L_j/L = N_j(Y/L, K/L, En/L, I_n) \tag{15}$$

Entonces el método de la OCDE consiste en descomponer la ecuación (15) en cuatro modelos para las regresiones simples y en dos para las múltiples. Para poder hacer esto, es necesario que las variables explicativas en (15) cumplan ciertas condiciones de independencia; de otra manera se introduce un sesgo en la estimación de los coeficientes de regresión.

Las regresiones se han hecho con un ajuste lineal logarítmico. Con un modelo de este tipo, (15) se transforma en:

$$\text{Log } L_j/L = \alpha \log Y/L + \beta \log K/L + \gamma \log En/L + \delta \log In + \xi + \eta \tag{16}$$

[5] Véase [4], p. 44.

donde α, β, γ, δ y ξ son los coeficientes de estimación y η un error aleatorio que será eliminado por la regresión.

La descomposición de (16) ajustada por la OCDE es:

$$\log L_j/L = \alpha^\circ \log Y/L + \xi_1 + \eta_1 \qquad (17)$$

$$\log L_j/L = \beta^\circ \log K/L + \xi_2 + \eta_2 \qquad (18)$$

$$\log L_j/L = \gamma^\circ \log En/L + \xi_3 + \eta_3 \qquad (19)$$

$$\log L_j/L = \delta^\circ \log In + \xi_4 + \eta_4 \qquad (20)$$

$$\log L_j/L + \tilde{\alpha} \log Y/L + \tilde{\beta} \log K/L + \xi_5 + \eta_5 \qquad (21)$$

en donde los ξ_i son los parámetros de ajuste y los η_i son los errores aleatorios que serán iguales a cero después de haber hecho la regresión.

Para facilitar la exposición utilizaremos expresiones matriciales para escribir:

$$l = \log L_j/L \quad \text{y} \quad y = [\log Y/L \quad 1] \qquad \alpha_1 = \begin{bmatrix} \alpha^\circ \\ \xi_1 \end{bmatrix}$$

donde l es un vector columna (m $\times$ 1), y una matriz (m $\times$ 2) y α_1 un vector columna (2 $\times$ 1), suponiendo que el número de observaciones es igual a m.

Luego (17) se escribirá:

$$l = y\alpha_1 + \eta_1^\circ \qquad (22)$$

donde η_1° es un vector columna (m $\times$ 1).

De igual manera, para (16), se establece:

$$\alpha_1' = \begin{bmatrix} \alpha \\ \xi \end{bmatrix} \quad \beta_1 = \begin{bmatrix} \beta \\ \gamma \\ \delta \end{bmatrix} \quad k = [\log K/L \quad \log En/L \quad \log In]$$

de donde (16) puede escribirse:

$$l = y\alpha_1' + k\beta_1 + \eta^\circ \qquad (23)$$

donde α_1' es un vector columna (2 $\times$ 1), k es una matriz (m $\times$ 3), β_1 un vector columna (3 $\times$1) y η° un vector columna (m $\times$ 1).

Ahora se trata de saber si la estimación de α_1 dada por el modelo (22) es igual o equivalente a la estimación α_i dada por el modelo (23).

El ajuste por mínimos cuadrados del modelo (23) produce:

$$l = y\, a_1' + k\, b_1 \qquad (24)$$

donde a_1' y b_1 son las estimaciones de α_1' y β_1 respectivamente.

Por otra parte, el ajuste por mínimos cuadrados de (22) produce:

$$l = y\,a_1 \tag{25}$$

donde
$$a_1 = ({}^t y\, y)^{-1}\, {}^t y l \tag{26}$$

donde a_1 es la estimación de α_1 y ${}^t y$ es la matriz transpuesta de y.
Sustituyendo (24) en (26) tenemos:

$$a_1 = ({}^t y\, y)^{-1}\; {}^t y\,(y\, a_1' + k\, b_1)$$

Entonces:
$$a_1 = a_1' + ({}^t y\, y)^{-1}\, {}^t y\, k\, b_1 \tag{27}$$

de donde, para que $a_1 = a_1'$, es necesario que:

$$({}^t y\, y)^{-1}\, {}^t y\, k\, b_1 = 0$$

pero sabemos que: $({}^t y\, y)^{-1} \neq 0$ y $b_1 \neq 0$

y entonces:
$$a_1 = a_1' \;\leftrightarrow\; {}^t y k = 0$$

o expresado en otra forma:

$$\begin{bmatrix} a^\circ \\ \\ \\ e_1 \end{bmatrix} = \begin{bmatrix} a \\ \\ \\ e \end{bmatrix} \leftrightarrow 0 =$$

$$\begin{bmatrix} \Sigma(\log Y/L)\,(\log K/L) & \Sigma(\log K/L)\,(\log En/L) & \Sigma(\log Y/L)\,(\log In) \\ \\ \\ \Sigma \log K/L & \Sigma \log En/L & \Sigma \log In \end{bmatrix}$$

donde a°, a, e_1 y e son las estimaciones de α°, α, ξ_1 y ξ.

Pero se sabe que:

$$\Sigma \log K/L \neq 0$$
$$\Sigma \log En/L \neq 0$$
$$\Sigma \log In \neq 0$$

entonces, en general, $e_1 \neq e$, lo cual se puede comprobar en el estudio de la OCDE.[6]

Por otra parte, se puede suponer que:

$$\Sigma(\log\,(Y/L)\,(\log K/L) \neq 0$$
$$\Sigma(\log Y/L)\,(\log En/L) \neq 0$$
$$\Sigma(\log Y/L)\,(\log In) \neq 0$$

[6] Véase [4], pp. 54, 77, 82, 83, 85 y 95.

por lo que la producción *per capita* (y lo mismo el ingreso) no es independiente del capital *per capita* ni de la energía *per capita* consumida ni del índice no monetario de desarrollo.

Encontramos entonces que en general:

$$a^{\circ} \neq a \quad \text{y} \quad e_1 \neq e$$

de donde podemos deducir que los modelos (17) a (21) proporcionarán valores sesgados para los parámetros estimados de manera que el valor de las elasticidades y de los parámetros de escala ajustados reflejan también el efecto de las variables omitidas sobre la variable explicada (L_j/L).

Sería interesante determinar la naturaleza del sesgo, es decir, ver si:

$$e_1 > e \quad \text{o} \quad e_1 < e$$

$$a^{\circ} > a \quad \text{o} \quad a^{\circ} < a$$

Esto no es un problema difícil, por lo menos para los coeficientes de regresión (elasticidades).

Estableciendo la hipótesis de independencia de los errores, la matriz de varianza-covarianza es una matriz diagonal tal que:

$$({}^t y\, y)^{-1} > 0$$

Por otra parte, por lo menos el primer renglón de la matriz ${}^t yk$ es positiva, dado que las variables que figuran en él varían en el mismo sentido.

Luego el signo de la expresión:

$$({}^t y\, y)^{-1}\ {}^t yk\, b_1$$

depende del signo de b_1, es decir, del signo de las elasticidades de las variables omitidas.

Pero del hecho de que el coeficiente de correlación sea positivo entre L_j/L y n, resulta que $b_1 > 0$,[7] de donde se puede concluir que existe una sobrestimación de las elasticidades calculadas por la OCDE.

Después de (27) se puede concluir que cuando la variable omitida para explicar la proporción L_j/L es más significativa, el sesgo de la elasticidad estimada con el modelo de la OCDE es más grande.

Pero es importante señalar que el sesgo puede no ser homogéneo para cada categoría profesional, dado que hay algunas como las de los ingenieros, las de los químicos, etc. (grupos 0-0, 0-1, 0-X, 0-02)[8] cuya proporción debe estar explicada con amplitud por el capital *per capita* en la mano de obra total.

Esto es fácilmente comprobable en el cuadro II-1 del libro de la OCDE, en donde se observa que las elasticidades para estos mismos grupos son sensiblemente superiores a las otras, sobre todo para la productividad del trabajo, que es explicada ampliamente por el capital *per capita.*

[7] Por ejemplo, en el cuadro II-1 del libro de la OCDE [4], p. 54, se observa que todas las elasticidades son positivas, por lo que de acuerdo con (27) deben estar sobrestimadas, dado que $a^{\circ} > a$.

[8] Véase [4], p. 49.

Ahora es necesario considerar que la estimación del modelo (16) presenta un gran problema: la correlación que existe entre las variables estudiadas.

Los esfuerzos realizados por la OCDE para la estimación de este modelo (21) y de otros donde han sido utilizadas regresiones múltiples, han sido infructuosos.[9] Debería advertirse a los lectores de las consecuencias y los riesgos de la descomposición del modelo, que reduce la varianza de los estimadores y que permite matizar su significado económico.

IV. Conclusiones

Por desgracia, hasta ahora los métodos estadísticos no van más allá de medir la pérdida de precisión que causa la eliminación de una de las variables colineales.

Si la relación que liga las variables explicativas es estable, entonces la omisión de una de las variables colineales no es particularmente grave, pero si la relación no es estable, entonces las estimaciones de los parámetros tienen el riesgo de ser muy imprecisas.[10]

Tomando el ejemplo del estudio de la OCDE (§ 2.3), la colinearidad encontrada en las regresiones múltiples doble-logarítmicas indica que existe una relación lineal doble-logarítmica entre las diferentes variables explicativas que no es estable, ya que los parámetros cambian en función de la calidad de la mano de obra, la calidad del capital y la composición del producto global en cada país. Esta relación logarítmica lineal es fijada por las relaciones tecnológicas de cada país, evidentemente diferentes.[11]

Por lo tanto, aun a sabiendas de que el sesgo será incontrolable, estamos obligados a desagregar la función (16) y proceder al ajuste.

De este modo, la utilización del estudio de la OCDE por los planificadores, aun con la precaución recomendada por los autores, es difícil.

Podemos entonces concluir que es muy peligrosa la aplicación de este estudio a un país particular para probar la sensibilidad de las proyecciones de mano de obra, si por lo menos no se han tomado las precauciones necesarias para intentar cuantificar el sesgo de las elasticidades calculadas, asunto bastante delicado también.

No obstante, es necesario aclarar que el trabajo de la OCDE no es más que el inicio del análisis sistemático de los datos disponibles sobre mano de obra, y que la calidad de estos datos varía grandemente de un país a otro.

En estas condiciones, se considera que la OCDE actuó juiciosamente al aplicar un modelo simple, ya que un refinamiento matemático o estadístico no conduce a resultados aceptables si se debe trabajar con datos poco confiables.

Por lo tanto, es difícil sostener un juicio definitivo sobre este sistema de análisis hasta que se haya aplicado de nuevo a cifras de encuestas futuras y censos y pueda ser confrontado como prueba de coherencia a la evolución observada en cada país o conjunto de países.

[9] Véase [4], cap. VI.
[10] Véase [2], pp. 232-237 y 331-334.
[11] Mark Blaugy [4], pp. 297-301.

BIBLIOGRAFÍA

[1] G. Debreu, *Theory of Value, an Axiomatic Analysis of Economic Equilibrium,* Nueva York, John Wiley & Sons, 1965.

[2] E. Malinvaud, *Méthodes statistiques de l'économétrie,* París, Dunod, 1969.

[3] E. Malinvaud, *Leçons de théorie microéconomique,* París, Dunod, 1969.

[4] OCDE, *Structures professionnelles et educatives et niveaux de développement économique,* París, 1970.