ANEXO. Demostración de que la varianza residual es mayor en el modelo IIa que en el Ia

La alta correlación lineal entre x1 y x2 significa que puede establecerse un modelo lineal que las relaciona:

$x_{2i} = (r_{12}\sigma_1 / \sigma_2)\, x_{1i} + w_i = (\Sigma x_{1i}x_{2i} / \Sigma x_{2i}^2)x_{1i} + w_i$ (III)

donde σ_1 y σ_2 son las desviaciones típicas de x_1 y x_2, respectivamente, y el error wi tiene distribución normal con media 0 y varianza σ_w^2.

Sustituyendo III en Ia se obtiene:

$y_i = \beta_1 x_{1i} + \beta_2 [(\Sigma x_1 i x_{2i} / \Sigma x_{2i}^2)x_{1i} + w_i] + u_i =$

$= [\beta_1 + \beta_2 (\Sigma x_{1i}x_{2i} / \Sigma x_{2i}^2)]x_{1i} + \beta_2 w_i + u_i =$

$= \beta_1{}^* x_{1i} + v_i$ (IIa)

donde $\beta_1{}^* = \beta_1 + \beta_2 (\Sigma x_{1i}x_{2i} / \Sigma x_{2i}^2)$, por lo tanto el sesgo introducido por $\beta_1{}^*$ es $\beta_2 (\Sigma x_{1i}x_{2i} / \Sigma x_{2i}^2)$ como señalan Sáez y Barceló; y $v_i = \beta_2 w_i + u_i$ es normal de media 0 y varianza $\sigma_v^2 = \beta_2\sigma_w^2 + \sigma_u^2 \geq \sigma_u^2$. La igualdad se produce sólo en el caso $\sigma_w^2 = 0$, es decir, cuando x_2 está completamente determinada por x_1 ($r_{12} = 1$)

Respuesta

Correspondencia: Marc Sáez. Departament d'Economia.
Universitat de Girona. Campus de Montivili. 17071 Girona.

Recibido: 21 de abril de 1999
Aceptado: 21 de abril de 1999

Hemos leído con sumo interés los comentarios de Llorca[1] a nuestra carta, en la que propusimos un criterio para omitir variables superfluas en modelos de regresión[2]. Como toda réplica que se valga, estamos en parte de acuerdo y en parte en desacuerdo con los contenidos en estos comentarios.

Como bien menciona Llorca, nuestro propósito fue el de proponer un criterio *sencillo*, en el sentido de no considerar incumplimientos o violaciones de las hipótesis básicas del modelo, lo que sin duda complicaría en exceso cualquier discusión. Es bien sabido que la elevada multicolinealidad *no* representa un incumplimiento de *ninguna* de las hipótesis básicas del modelo. En este sentido a veces se ha dicho que la multicolinealidad no es un problema sino una molestia que, por desgracia, siempre se debe soportar. Llorca, plantea modelos en los que se incumple al menos una de las principales hipótesis básicas. Nos referimos en concreto a la existencia de regresores estocásticos.

Así, en su modelo Ia:

$y_i = \beta(_1 x_{1i} + \beta_2 x_{2i} + u_i$ (Ia)

la variable x_2 es estocástica, por cuanto tal y como la formula contiene un error aleatorio:

$x_{2i} = (\Sigma x_{1i}x_{2i}/\Sigma x_{2i}^2)x_{1i} + w_i$ (III)

donde el error w_i tiene una distribución normal con media 0 y varianza σ_w^2.

Pero es que además, los errores u_i y w_i podrían correlacionados. Como consecuencia, el estimador de β_1 no *sólo* estaría sesgado en (IIa), resultado de la omisión de una variable relevante, *sino también* en (Ia), a causa del incumplimiento de la citada hipótesis básica. Así, Llorca compararía estimadores sesgados por lo que, creemos, no puede deducir ninguna consecuencia relevante.

Por otra parte, y de nuevo parafraseando a Llorca, sus comentarios *contienen un error*. En la última frase del anexo señala que 'la igualdad se produce sólo en el caso $\sigma_w^2=0$, es decir, cuando x_2 está completamente determinada por x_1 ($r_{12}=1$)'. Volviendo a (III) es evidente que aún siendo $\sigma_w^2=0$, el coeficiente de correlación entre ambos regresores puede ser diferente de la unidad. De hecho cuando $\sigma_w^2=0$ (y por tanto $\sigma_v^2=\sigma_u^2$) la variable explicativa x_2 será determinista, el caso que, precisamente, comentamos nosotros.

M. Sáez y M.A. Barceló
Departament d'Economia, Universitat de Girona

Bibliografía

1. Llorca J. Omisión de variables en modelos de regresión con alta multicolinealidad. Gac Sanit 1999;13(3):243-4.

2. Sáez M y Barceló MA. Un criterio para omitir variables superfluas en modelos de regresión. Gac Sanit 1998;12:281-3.