

La estimación de modelos interdependientes con relaciones lineales y logarítmicas: los mínimos cuadrados bietápicos con duplicidad de escalas

I. INTRODUCCIÓN

Cualquier economista acostumbrado a operar con series macroeconómicas es consciente de la necesidad que en muchos casos la realidad impone de especificar modelos interdependientes. Sin embargo, su estimación no está exenta de dificultades tal como demostró Haavelmo¹ en la década de los cuarenta. En particular, la estimación de cada ecuación del modelo por Mínimos Cuadrados Ordinarios (MCO), da lugar a estimadores sesgados e inconsistentes. Y en este sentido, el método de los Mínimos Cuadrados Bietápicos (MCB), entendido como un caso especial de la técnica de las Variables Instrumentales (VI), permite obtener estimadores consistentes. Fueron H. Theil² y R. L. Basmann³ quienes por primera vez, y en forma independiente, propusieron la utilización de este método.

La ventaja de los MCB con respecto a las Variables Instrumentales radica precisamente en el hecho de que proporciona un criterio objetivo para generar instrumentos. Éstos se obtienen a partir de la regresión de las variables endógenas incluidas como explicativas en la ecuación estructural correspondiente, con respecto a la totalidad de variables predeterminadas del modelo. Esta objetividad se mantiene en tanto en cuanto: a) dispongamos de suficientes ob-

1. HAAVELMO, T.: «The Statistical Implications of a System of Simultaneous Equations», *Econometrica*, vol. 11, núm. 1, 1943.

2. THEIL, H.: «Repeated Least Squares Applied to Complete Equation Systems», *The Hague: Central Planning Bureau*, 1953.

3. BASMANN, R. L.: «A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation», *Econometrica*, 25, enero 1957.

servaciones para poder realizar la primera fase de la estimación, es decir, la regresión entre variables endógenas explicativas y predeterminadas del modelo; en otro caso, la necesidad de recurrir, por ejemplo, a la técnica de los componentes principales,⁴ aumenta la dosis de subjetivismo de la estimación por el hecho de carecer de criterios que permitan determinar *a priori* el número de componentes principales a retener, y *b*) las relaciones del modelo sean lineales.

Este último aspecto, el *b*, ha recibido poca atención por parte de los econométricos. No obstante, en algunos casos puede resultar de considerable importancia.

En la exposición que sigue, para ilustrar nuestras ideas, recurriremos a un sencillo modelo keynesiano de determinación de la renta. El simple y clásico modelo de dos ecuaciones, diseminado entre los distintos libros introductorios de economía para explicar el concepto de multiplicador, será el que nos servirá como pauta. A continuación se presenta la formulación general del estimador propuesto (que hemos optado por denominarlo Mínimos Cuadrados Bietápicos con Duplicidad de Escalas) para el caso de modelos multiecuacionales más desagregados. Finalmente, se ofrecen algunas indicaciones de posibles circunstancias en economía en las que puede resultar aconsejable recurrir al método descrito de los Mínimos Cuadrados Bietápicos con Duplicidad de Escala (MCBDE).

II. EL MODELO KEYNESIANO DE DETERMINACIÓN DE LA RENTA: VARIANCIA ASINTÓTICA DE LOS ESTIMADORES MCB Y MCBDE

Supongamos el siguiente modelo de dos ecuaciones:

$$\begin{aligned} y &= E + A & [1] \\ \ln E &= \beta \ln Y + u \end{aligned}$$

en donde:

Y = PNB

E = Gasto endógeno

A = Gasto autónomo (Exportaciones «más» gasto público corriente y de inversión)

Como es usual, postulamos que u es una perturbación aleatoria con media cero, variancia constante y covariancias nulas:

$$E(uu') = \sigma_u^2 I$$

4. KLOEK, T., y L. B. M. MENNES: «Simultaneous Equation Estimation Based on Principal Components of Predetermined Variables», *Econometrica*, vol. 28, 1960.

El problema que comporta la estimación de [1] por MCO es que la perturbación aleatoria u estará correlacionada con la variable endógena explicativa (el logaritmo neperiano del PNB). Por ello, de seguir tal procedimiento se obtiene un estimador del parámetro poblacional sesgado e inconsistente.

En efecto, si volvemos a reescribir [1] de la siguiente forma:

$$E^* = \beta Y^* + u \quad [1]$$

indicando por el signo * que se ha tomado logaritmo de la correspondiente variable ($E^* = \ln E, Y^* = \ln Y$), puede demostrarse que se verifica:

$$\text{plim} \left(\frac{1}{n} \sum Y^* u \right) \neq 0$$

Para solventar esta dificultad, el método de los MCB, que en realidad constituye un caso especial de la técnica de las VI,⁵ obtiene un instrumento (\hat{Y}^*) a partir de la regresión de Y^* con respecto a la variable predeterminada del modelo:

$$\hat{Y}^* = \hat{\pi}_1 \cdot A \quad [2]$$

de donde se deduce el estimador:

$$\hat{\beta} = \frac{\sum E^* \cdot \hat{Y}^*}{\sum Y^* \cdot \hat{Y}^*} \quad [3]$$

$$\text{var. asint.} (\hat{\beta}) = \sigma^2 \cdot n^{-1} \cdot \frac{\text{plim} \left(\frac{1}{n} \sum \hat{Y}^{*2} \right)}{\left[\text{plim} \left(\frac{1}{n} \sum \hat{Y}^* \cdot Y^* \right) \right]^2} \quad [4]$$

Por la hipótesis de independencia entre el gasto autónomo y las perturbaciones, el instrumento \hat{Y}^* será independiente de dichas perturbaciones, con lo que queda garantizada la propiedad de consistencia.

Sin embargo, ¿se trata de un estimador eficiente? Cualquiera que sea la definición del instrumento, si establecemos:

$$Y^* = \hat{Y}^* + v$$

siendo v una nueva perturbación aleatoria que recoge la discrepancia entre el

5. Véase, por ejemplo, GOLDBERGER, A. S.: «An Instrumental Variable Interpretation of k-Class Estimation», *The Indian Economic Journal*, vol. 13, 1965.

valor de la variable endógena explicativa y el de su correspondiente instrumento, se desprende:

$$\text{plim} \left(\frac{1}{n} \sum \hat{Y}^{*2} \right) = \text{plim} \left(\frac{1}{n} \sum Y^{*2} \right) + \\ + \text{plim} \left(\frac{1}{n} \sum v^2 \right) - 2 \text{plim} \left(\frac{1}{n} \sum vY^* \right)$$

suponiendo que se verifica: $\text{plim} \left(\frac{1}{n} \sum v\hat{Y}^* \right) = 0$, se obtiene:

$$\text{plim} \left(\frac{1}{n} \sum \hat{Y}^{*2} \right) = \text{plim} \left(\frac{1}{n} \sum Y^{*2} \right) - \text{plim} \left(\frac{1}{n} \sum v^2 \right)$$

Análogamente, el denominador de la igualdad [4] puede también expresarse de la siguiente forma:

$$\left[\text{plim} \left(\frac{1}{n} \sum \hat{Y}^* Y^* \right) \right]^2 = \left[\text{plim} \left(\frac{1}{n} \sum Y^{*2} \right) - \text{plim} \left(\frac{1}{n} \sum v^2 \right) \right]^2$$

En consecuencia:

$$\text{var. asint.}(\hat{\beta}) = \sigma_u^2 \cdot n^{-1} \cdot \frac{1}{\text{plim} \left(\frac{1}{n} \sum Y^{*2} \right) - \text{plim} \left(\frac{1}{n} \sum v^2 \right)}$$

De aquí se deduce que cuanto menor sea la variancia poblacional de v

$$\sigma_v^2 = \text{plim} \left(\frac{1}{n} \sum v^2 \right)$$

mayor será la eficiencia del estimador. Dicho en otros términos, para obtener un estimador eficiente es preciso definir un instrumento muy correlacionado con la correspondiente variable endógena.

¿Satisface este requisito el estimador MCB usual aplicado a [1]? En general, la respuesta es negativa. Así, diferenciando [2]:

$$\frac{d\hat{Y}^*}{\hat{Y}^*} = \hat{\Pi}_1 \cdot dA$$

Es decir, estimar [2] equivale a considerar que el «porcentaje de variación del PNB depende de la variación absoluta del gasto autónomo». Dado que en economía se opera con series crecientes en el tiempo, carece de sentido presuponer que la misma variación absoluta del gasto autónomo dará lugar al mismo porcentaje de variación del producto a principio que a final del período. A *grosso modo* si, por ejemplo, al principio del período diez mil

millones más de gasto autónomo justifican un crecimiento del producto de un 6 por ciento, a final del período, cuando el nivel de PNB se haya duplicado, la misma variación absoluta del gasto autónomo sólo conllevará una elevación aproximada del producto de un 3 por ciento. Por ello, sin poner en duda la propiedad de consistencia de la estimación de β partiendo de un instrumento generado por [2], deseamos resaltar la falta de sentido económico de esta forma de proceder. Este aspecto, en definitiva, se reflejará en una baja correlación entre \tilde{Y}^* e Y^* , con lo que pagaremos un alto precio por la consistencia: la elevada variancia del coeficiente $\tilde{\beta}$ estimado.

Una segunda alternativa podría consistir en formar el instrumento a partir de la regresión entre el logaritmo neperiano del PNB y el logaritmo neperiano del gasto autónomo. Pero tal posibilidad no siempre resultará factible. Si estuviésemos considerando un modelo econométrico más desagregado, parte de sus variables predeterminadas podrían adoptar valores positivos y/o negativos, y el logaritmo de un número negativo carece de sentido.

Una tercera posibilidad, que siempre será realizable y que nosotros proponemos, comporta calcular la regresión entre los valores absolutos del PNB y los valores absolutos del gasto autónomo:

$$\tilde{Y} = \tilde{\Pi}_2 \cdot A \quad [5]$$

y, posteriormente, obtener el instrumento a utilizar en la estimación de [1] extrayendo logaritmos neperianos de \tilde{Y} :

$$\tilde{Y}^* = \ln \tilde{Y} \quad [6]$$

Casi siempre que las variables endógenas explicativas sean crecientes en el tiempo, el instrumento definido por [6] estará más correlacionado con Y^* que el definido por [2]. Por ello, al efectuar un cambio de escala entre la primera y la segunda fase de la estimación (de aquí el nombre de Mínimos Cuadrados Bietápicos con Duplicidad de Escalas), se mejorará la eficiencia. Adicionalmente, dado que el nuevo instrumento sólo es función de la variable predeterminada del modelo, sigue verificándose la propiedad de consistencia para el estimador, que vendrá definido por:

$$\tilde{\beta} = \frac{\sum E^* \cdot \tilde{Y}^*}{\sum Y^* \cdot \tilde{Y}^*} \quad [7]$$

$$\text{var. asint. } (\beta) = \sigma_u^2 \cdot n^{-1} \cdot \frac{\text{plim} \left(\frac{1}{n} \sum \tilde{Y}^{*2} \right)}{\left[\text{plim} \left(\frac{1}{n} \sum \tilde{Y}^* \cdot Y^* \right) \right]^2} \quad [8]$$

La ventaja de [7] sobre [3] depende crucialmente del hecho de que la variable endógena explicativa sea creciente en el tiempo. A este respecto, a título ilustrativo, partiendo de datos relativos a la economía española para el período 1954-1972 (véase cuadro 1), se ha procedido a estimar [4] y [8]

CUADRO 1

*Evolución del PNB del gasto autónomo y del gasto endógeno
en la economía española*

(Datos en miles de millones de pesetas)

Años	Y = PNB a precios de mercado en ptas. corrientes	A = Gasto autónomo (Exportaciones "más" gasto público)	E = Gasto endógeno (E = Y - A)
1954	337	56	281
1955	376	64	312
1956	432	74	358
1957	506	88	418
1958	582	96	486
1959	603	116	487
1960	620	143	477
1961	707	160	546
1962	817	186	631
1963	964	215	749
1964	1.088	256	832
1965	1.287	294	993
1966	1.477	354	1.123
1967	1.632	398	1.234
1968	1.805	470	1.335
1969	2.011	545	1.466
1970	2.252	657	1.595
1971	2.538	791	1.747
1972	2.960	921	2.039

FUENTE: Contabilidad Nacional de España, años 1954 a 1964 y años 1964 a 1972.

sustituyendo los límites probabilísticos que en ellas aparecen por los respectivos valores muestrales. El resultado obtenido ha sido el siguiente:

a) *Variancia asintótica estimada del estimador MCB*

$$\text{var. } \hat{\beta} = \sigma_u^2 \cdot n^{-1} \cdot (0,0016035)$$

b) *Variancia asintótica estimada del estimador MCBDE*

$$\text{var. } \tilde{\beta} = \sigma_u^2 \cdot n^{-1} \cdot (0,0011003)$$

de donde se deduce una reducción porcentual en la variancia asintótica del estimador del orden de un 30 por ciento. Es decir:

$$\frac{\text{var. \hat{\beta}} - \text{var. \tilde{\beta}}}{\text{var. \hat{\beta}}} \times 100 = 31,38\%$$

La magnitud de este dato pone de manifiesto la importancia que en algunos casos puede tener efectuar un cambio de escala entre las dos fases de la estimación.

III. FORMULACIÓN GENERAL DEL ESTIMADOR

Dada la primera ecuación de un modelo interdependiente de G ecuaciones:

$$y_1^* = Y_1^* \beta_1 + X_1^* \gamma_1 + u_1 \quad [9]$$

en donde y_1^* es un vector columna de n observaciones de los logaritmos de y_1 , Y_1^* es una matriz $n \times g$ compuesta por los logaritmos de las observaciones de las variables endógenas incluidas como explicativas en la primera ecuación, β_1 es un vector columna de g coeficientes estructurales, X_1^* es una matriz $n \times k$ compuesta por los logaritmos de las observaciones de las variables predeterminadas incluidas como explicativas (en realidad, es indiferente que tales variables predeterminadas aparezcan bajo la forma logarítmica o lineal, y hemos supuesto la primera por razones de homogeneidad con las variables endógenas), y u_1 es un vector columna de n perturbaciones aleatorias. Es decir,

$$y_1^* = \begin{bmatrix} \ln y_{11} \\ \ln y_{12} \\ \cdot \\ \cdot \\ \ln y_{1n} \end{bmatrix}$$

$$Y_1^* = \begin{bmatrix} \ln y_{21} & \dots & \ln y_{g+1,1} \\ \ln y_{22} & \dots & \ln y_{g+1,2} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \ln y_{2n} & \dots & \ln y_{g+1,n} \end{bmatrix}$$

etcétera.

El instrumento para estimar [9] por el método de los MCBDE se define a partir de:

$$Z = \begin{bmatrix} \tilde{Y}_1^* & X_1^* \end{bmatrix}$$

siendo:

$$\tilde{Y}_1 = X(X'X)^{-1}X'Y_1$$

en donde X es una matriz $n \times K$ de observaciones de la totalidad de variables predeterminadas del modelo.

Posteriormente, extrayendo logaritmos de los elementos de esta igualdad:

$$\tilde{Y}_1^* = \ln[\tilde{Y}_1] = \ln[X(X'X)^{-1}X'Y_1]$$

En consecuencia, el estimador MCBDE se obtiene calculando:

$$\begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\gamma}_1 \end{bmatrix} = \begin{bmatrix} \ln \left[Y_1' X (X'X)^{-1} X' \right] Y_1^* & \ln \left[Y_1' X (X'X)^{-1} X' \right] X_1^* \\ X_1^{*'} Y_1^* & X_1^{*'} X_1^* \end{bmatrix}^{-1}$$

$$\begin{bmatrix} \ln \left[Y_1' X (X'X)^{-1} X' \right] y_1^* \\ X_1^{*'} y_1^* \end{bmatrix}$$

$$\text{var. } \tilde{\text{ásint.}} \begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\gamma}_1 \end{bmatrix} = \sigma_u^2 \begin{bmatrix} \ln \left[Y_1' X (X'X)^{-1} X' \right] Y_1^* \ln \left[Y_1' X (X'X)^{-1} X' \right] X_1^* \\ X_1^{*'} Y_1^* & X_1^{*'} X_1^* \end{bmatrix}^{-1}$$

$$\begin{bmatrix} \ln \left[Y_1' X (X'X)^{-1} X' \right] \ln \left[X (X'X)^{-1} X' Y_1 \right] \ln \left[Y_1' X (X'X)^{-1} X' \right] X_1^* \\ X_1^{*'} \ln \left[X (X'X)^{-1} X' Y_1 \right] & X_1^{*'} X_1^* \end{bmatrix}$$

$$\begin{bmatrix} Y_1^{*'} \ln \left[X (X'X)^{-1} X' Y_1 \right] & Y_1^{*'} X_1^* \\ X_1^{*'} \ln \left[X (X'X)^{-1} X' Y_1 \right] & X_1^{*'} X_1^* \end{bmatrix}^{-1}$$

$$\tilde{\sigma}_u^2 = \frac{1}{n - g - k} (y_1^* - Y_1^* \tilde{\beta}_1 - X_1^* \tilde{\gamma}_1)' \cdot (y_1^* - Y_1^* \tilde{\beta}_1 - X_1^* \tilde{\gamma}_1)$$

Como puede comprobarse, el estimador propuesto comporta una mayor complejidad de cálculo que el de los MCB, y en general, la mejora en la eficiencia asintótica con respecto a este último dependerá de que las variables endógenas explicativas que componen Y_1 sean crecientes en el tiempo. Adicionalmente, como es obvio, el modelo debe contener relaciones lineales (ya sean de definición o de otro tipo) además de las logarítmicas para que esté justificado el cómputo de MCBDE.

En el supuesto de que nos enfrentásemos conjuntamente ante un problema de falta de grados de libertad ($K \geq n$) y de relaciones lineales y logarítmicas, siguiendo la sugerencia de Kloeck y Mennes,⁶ podría definirse la siguiente matriz de instrumentos:

$$Z = \left\{ \ln \left[P(P'P)^{-1}P'Y_1 \right] \quad X_1^* \right\}$$

en donde P es una matriz $n \times b$ ($b < n$) de componentes principales de la totalidad de variables predeterminadas del modelo estandarizado.⁷ El estimador que se obtendría sería igual al precedente, salvo que la matriz:

$$P(P'P)^{-1}P'Y_1$$

sustituiría a la matriz:

$$X(X'X)^{-1}X'Y_1$$

Esta utilización de los componentes principales para formar instrumentos en lugar de «proxies», se aparta algo de la propuesta de Kloeck y Mennes. No obstante, en nuestro caso constituye una prolongación natural de la interpretación del estimador MCB como aplicación particular de las VI. Además, entre otros, puede encontrarse un enfoque similar en los libros de Klein⁸ y de Huang.⁹

6. KLOECK, T., y L. B. M. MENNES: *op. cit.*

7. Una variable estandarizada X_i^* de X_i se obtendrá a partir de:

$$X_i^* = \frac{X_i - \bar{X}}{\left[\frac{\sum (X_i - \bar{X})^2}{n-1} \right]^{1/2} \cdot (n-1)^{1/2}}$$

siendo \bar{X} el valor medio de X_i . La media de las variables estandarizadas es cero y la suma de sus cuadrados unitaria, con lo que se eliminan los efectos que podrían tener las distintas escalas variables predeterminadas (un índice frente a una magnitud absoluta) sobre los componentes principales obtenidos.

8. KLEIN, L. R.: *A Textbook of Econometrics*, Prentice Hall, Nueva Jersey, 1974 (segunda edición).

9. HUANG, D. S.: *Regression and Econometric Methods*, Wiley and Sons, Nueva York, 1970.

IV. APLICACIONES EN ECONOMÍA

La justificación del método propuesto la hemos deducido de un ejemplo económico. En realidad, las posibilidades de aplicación son múltiples. Citemos a título meramente ilustrativo algunas situaciones en que podría estar justificado el empleo de los MCBDE.

Supongamos un modelo macroeconómico lineal salvo en el caso de las funciones de ingresos impositivos, que viniesen definidas en términos logarítmicos. Ésta es una circunstancia muy frecuente ya que los coeficientes estimados, en el caso de las especificaciones logarítmicas, indican elasticidades. Suponer la constancia de una elasticidad impositiva es una hipótesis habitual y razonable. Si definimos, por ejemplo, el logaritmo neperiano de los impuestos directos personales en función del logaritmo neperiano de la renta antes de impuestos, el coeficiente estimado será una elasticidad impositiva. Por su valor (próximo o alejado de la unidad) fácilmente podremos deducir el significado económico de la estimación. Sin embargo, para efectuar la estimación por MCB, tendrá más sentido relacionar los valores absolutos de la renta disponible antes de impuestos con los valores absolutos de las variables predeterminadas del modelo, y posteriormente, deducir el instrumento extrayendo logaritmos de los valores proyectados, que generar el instrumento a partir de la regresión del logaritmo neperiano de la renta disponible antes de impuestos con respecto a los valores absolutos de las demás variables predeterminadas del modelo.

En el ejemplo citado, la conveniencia de aplicar MCBDE se fundamenta en el hecho de que la serie de renta disponible posiblemente será creciente en el tiempo. De seguir la alternativa usual de los MCB —logaritmo neperiano de la renta disponible antes de impuestos en función de las variables predeterminadas del modelo—, estaremos implícitamente postulando que a principios de período un aumento de X unidades monetarias de una variable exógena del modelo —las exportaciones, por ejemplo—, ejerce el mismo efecto sobre el «crecimiento porcentual» de la renta disponible que a final del período, cuando el valor de esta variable puede ser cuatro o más veces superior a su valor inicial, pues normalmente la renta disponible debería medirse a precios corrientes. En consecuencia, la correlación entre instrumento y variable endógena será baja, y por ende, la eficiencia del estimador. Utilizando dos escalas —proyección de los valores absolutos de la renta disponible antes de impuestos, y *a posteriori*, extracción de los correspondientes logaritmos— lograremos un instrumento más correlacionado con el regresor, lo que revertirá en una disminución de la variancia asintótica del estimador.

¿Y por qué no establecer directamente la regresión entre el logaritmo neperiano de la renta disponible antes de impuestos y el logaritmo neperiano de las variables predeterminadas del modelo? Como ya se indicó con anterioridad, tal alternativa no será generalmente factible, pues en casi todo

modelo econométrico parte de sus variables predeterminadas pueden adoptar valores positivos, cero o negativos, y el logaritmo de cero o de un número negativo carece de sentido. Piénsese por ejemplo en la inversión privada neta retardada, variable predeterminada que generalmente interviene como explicativa en la función de inversión, o en la inversión de determinados sectores considerados en forma exógena en el modelo. En consecuencia, si bien en algunos casos la formación de instrumentos a partir de la extracción de logaritmos de todas las variables predeterminadas puede estar justificada —nosotros hemos hecho uso de tal posibilidad en la estimación de un modelo econométrico sobre precios y salarios relativo a la economía española—,¹⁰ tal forma de proceder no siempre resultará válida.

Otro ejemplo nos vendría dado por la introducción en el modelo de cualquier variante de una función de producción del tipo Cobb-Douglas. Para determinar el empleo, es frecuente especificar:

$$\ln N = -\beta_0 + \beta_1 \ln X - \beta_2 \ln K - \beta_3 T$$

siendo N el empleo, X el nivel de output industrial, K el stock de capital, y T el transcurso del tiempo. Centrándonos en el output X , dado que generalmente será una variable endógena en el modelo, tiene también más sentido económico relacionar sus valores absolutos con los valores absolutos de las variables predeterminadas, y posteriormente extraer logaritmos de los valores proyectados, que establecer directamente la regresión entre el logaritmo neperiano de X y la totalidad de variables predeterminadas.

No pretendemos que el método propuesto tenga validez general, pero pueden idearse multitud de ejemplos en economía en los que tiene mayor significado económico y resulta más aconsejable que la aplicación directa de los MCB, pues no hay que olvidar que la consistencia es una propiedad asintótica, y en ocasiones, de poco consuelo le sirva al investigador obligado a trabajar con pocas observaciones. En tales circunstancias, la reducida variancia del estimador puede constituir una razón de gran peso en el momento de decidir la adopción de uno u otro método de estimación.

*Facultad de Ciencias Económicas
Universidad de Barcelona*

10. RAYMOND, J. L.: «Precios y Salarios: Estabilidad versus Crecimiento en España», *Confederación Española de las Cajas de Ahorro* (en curso de publicación). Puede encontrarse un resumen en *Hacienda Pública Española*, núm. 32, 1975.